IUT de Paris Descartes

Exercices sur les *Chaînes de Markov.*

- **Exercice 1 :**

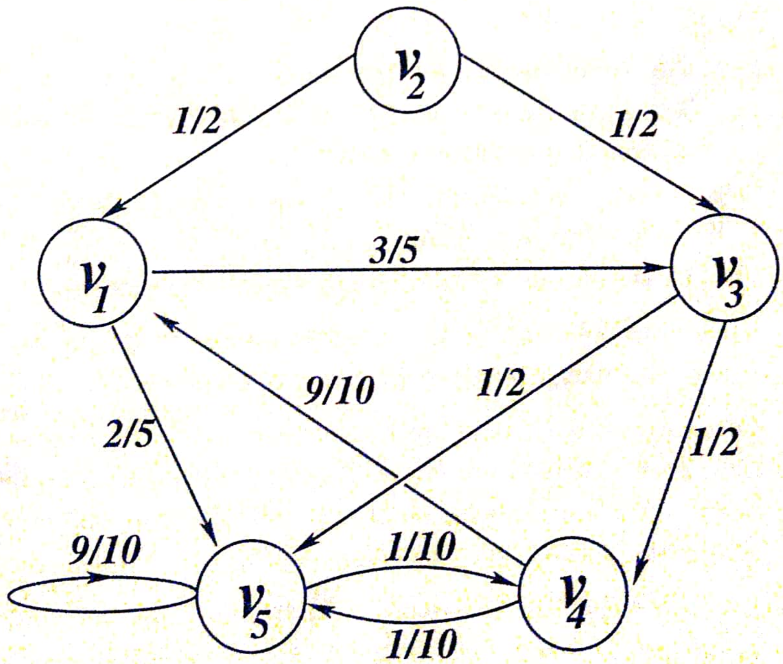
  On considère deux chaînes de Markov $(X_k)_{k\geq 0}$ et $(X'_k)_{k\geq 0}$ sur l'espace d'états $E = \{1, 2, \ldots, 5\}$ ayant pour matrices de transition respectives les matrices P, P' données par:

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{4}{5} & \frac{1}{5} \\ \frac{1}{6} & 0 & 0 & 0 & \frac{5}{6} \end{pmatrix}, \quad P' = \begin{pmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{3}{4} & 0 & 0 & 0 & \frac{1}{4} \\ \frac{1}{5} & \frac{4}{5} & 0 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{5}{6} & 0 & 0 \end{pmatrix}$$

  1. Identifier les graphes de transition associés à chacune de ces chaînes.
  2. En supposant que $\mathbb{P}\{X_0 = 1\} = 1$, identifier $Loi(X_0)$ puis $Loi(X_1)$ et $Loi(X_2)$.
  3. En supposant que $\mathbb{P}\{X'_0 = 1\} = \frac{1}{2} = \mathbb{P}\{X'_0 = 5\}$, identifier $Loi(X'_0)$ puis $Loi(X'_1)$ et $Loi(X'_2)$.

- **Exercice 2 :**

  On considère une chaîne de Markov $(X_k)_{k\geq 1}$ prenant ses valeurs dans un espace d'états $\mathcal{V} = \{v_1, v_2, v_3, v_4, v_5\}$ et ayant le graphe de transition suivant:



  1. Quelle est la matrice de transition P associée à cette chaîne?
  2. En résolvant l'équation $\pi \times P = \pi$, spécifier la (ou les) mesure(s) invariante(s) associée(s) à cette chaîne.

- **Exercice 3 :** On mesure à intervalles réguliers (toutes les heures) la taille d'une population de bactéries; l'évolution de cette taille est modélisée comme une chaîne de Markov $(X_1, X_2, \ldots, X_k, \ldots)$ prenant ses valeurs dans l'espace d'états $E = \{e_0, e_1, e_2, e_3\}$, où l'état e_0 correspond à une population en exctinction (0 individus), les états e_1 et e_2 à une population d'environ 1 (resp. 2) Million(s) d'individus, et l'état e_3 à une population d'au moins 3 Millions d'individus.

  1. En supposant que les transitions de la chaîne sont données par

  $$p_{0,0} = 1,\ p_{1,0} = 0,9,\ p_{1,2} = p_{2,1} = 0,1,\ p_{2,3} = 0,9,\ p_{3,2} = 0,05,\ p_{3,3} = 0,95,$$

  tous les autres coefficients de transition étant nuls, tracer le graphe des transitions de la chaîne.
  Quelle est (ou quelles sont) la (les) mesure(s) invariante(s) associée(s) à cette chaîne?

  2. Mêmes questions avec des transitions données par

  $$p_{0,0} = 1,\ p_{1,0} = 0,95,\ p_{1,2} = p_{2,1} = 0,05,\ p_{2,3} = 0,95,\ p_{3,3} = 1.$$

- **Exercice 4 :** Un joueur dispose de 6 boules réparties au sein de deux urnes A et B.

  A- On suppose qu'à chaque seconde, le joueur choisit l'une des 6 boules entièrement au hasard et la change d'urne.

  1. Modéliser ce processus au moyen d'une chaîne de Markov $(X_n)_{n\geq 0}$ à valeurs dans un espace E de sept états. (Graphe des transitions et matrice stochastique SVP).
  2. Utiliser la *réversibilité* de cette chaîne pour trouver sa mesure invariante $\boldsymbol{\pi_X}$. (*Rappel* : $\binom{6}{0} + \binom{6}{1} + \ldots + \binom{6}{6} = 2^6$).
  3. Cette chaîne est-elle irréductible? Apériodique?

  B- Comment généraliser les considérations de la question précédente au cas où les urnes renferment en tout n boules ?

  C- On suppose à présent qu'à chaque seconde, le joueur choisit l'urne A avec probabilité $\frac{1}{4}$ ou l'urne B avec probabilité $\frac{3}{4}$, puis qu'il prend une boule au hasard dans l'urne choisie pour la changer d'urne (pas de changement si l'urne choisie est vide).

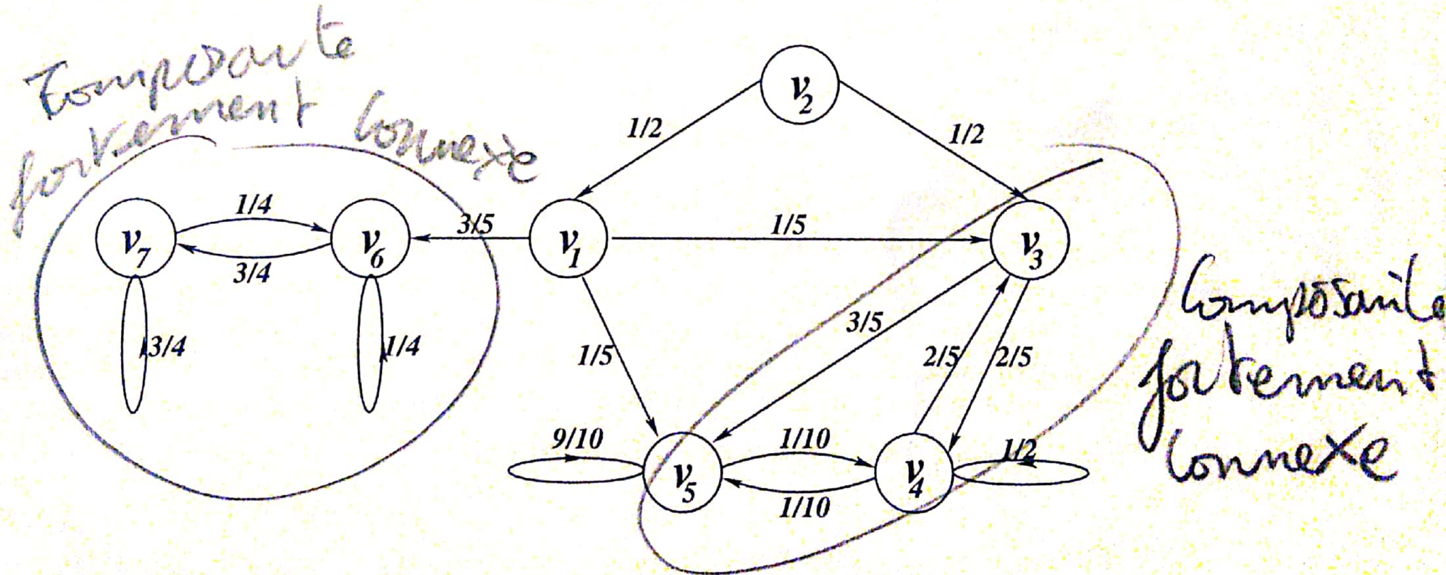
  1. Modéliser ce nouveau processus au moyen d'une nouvelle chaîne $(Y_n)_{n\geq 0}$ à valeurs dans l'espace E. (Graphe et matrice SVP).
  2. Trouver la mesure invariante $\boldsymbol{\pi_Y}$ de la chaîne $(Y_n)_{n\geq 0}$. (*Aide* : $1 + 3 + 3^2 + \ldots + 3^6 = \frac{3^7-1}{3-1} = 2187$).
  3. Cette chaîne est-elle irréductible? Apériodique? Qu'en concluez-vous?

- **Exercice 5 :**
Un joueur s'amuse à déplacer une boule blanche et une boule noire au sein de deux urnes A et B au fil du temps. A chaque instant, il lance une pièce équilibrée : si cette pièce produit "Pile", il changera la boule blanche d'urne avec prob. $p = 0,9$, tandis qu'avec prob. $1 - p = 0,1$ rien ne changera ; en revanche, si la pièce produit "Face", il changera la boule noire d'urne avec prob. $p' = 0,25$, tandis qu'avec prob. $1 - p' = 0,75$ rien ne changera[1].

   A- Soit Y_n le nombre de boules figurant dans l'urne A au temps n. Montrer que la suite $(Y_n)_{n\geq 0}$ *n'est pas une chaîne de Markov* !

   B- Soit X_n la variable comptant le nombre de boules blanches figurant dans l'urne A à l'instant n et X'_n la variable comptant le nombre de boules noires figurant dans l'urne A à l'instant n. On définit $Z_n = (X_n; X'_n)$. Vérifier que $(Z_n)_{n\geq 0}$ constitue une chaîne de Markov, dont on décrira l'espace d'états et les transitions (graphe, matrice).

- **Exercice 6 : Etude d'une chaîne à 7 états** On considère une chaîne de Markov $(X_k)_{k\geq 1}$ prenant ses valeurs dans un espace d'états $\mathcal{V} = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ et ayant le graphe de transition suivant:



1. Quelle est la matrice de transition P associée à cette chaîne ?
2. Donner la classification des états de cette chaîne (une classe T d'états transients, deux classes C_1, C_2 d'états récurrents).
3. On fait démarrer la chaîne de l'état v_1. Quelle est la loi de la chaîne après un pas $(Loi(X_1|X_0 = v_1))$? Après deux pas ?
4. Cette chaîne est-elle irréductible ? Combien a-t-elle de mesures invariantes ?

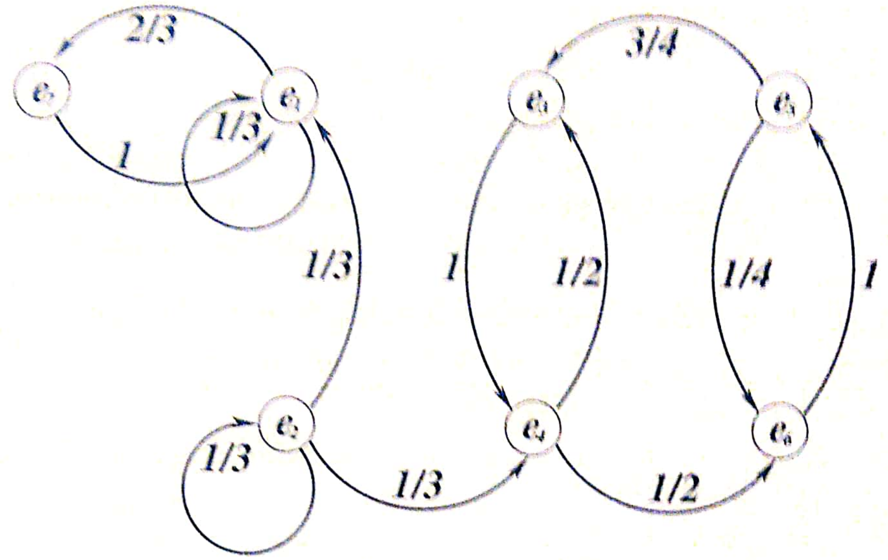
[1](Notre joueur utilise donc une pièce équilibrée mais aussi la fonction *"rand"* d'*Excel*).

5. Spécifier les mesures invariantes $\pi[1], \pi[2]$ satisfaisant

$$\pi[1](C_1) = 1 = \pi[2](C_2)$$

Ces deux classes récurrentes C_1 et C_2 sont-elles *apériodiques* ?

- **Exercice 7 :**
Reprendre les questions de l'Ex. précédent dans le cas de la chaîne $(Y_k)_{k \geq 1}$ ayant le graphe de transitions ci-dessous :



- **Exercice 8 : application aux mélanges de cartes.**
Un soir d'hiver, au Casino de Bouville, le croupier *Jean-Paul S.* décide de tromper son ennui en analysant des chaînes de Markov qui représentent sa manière de mélanger un jeu de 32 cartes. A l'instant initial de son expérience de mélange, le jeu de cartes est parfaitement ordonné suivant les conventions habituelles (Trèfle, Carreau, Coeur, Pique, l'ordre des cartes allant crescendo du 8 à l'As au sein d'une même couleur).

1. Supposons qu'à de nombreuses reprises, JPS prélève une seule carte au hasard au sein du paquet pour la glisser à nouveau entièrement au hasard au sein du paquet.
Modéliser ce processus de mélange par une chaîne de Markov (on commencera par se donner un espace d'états E approprié, puis on spécifiera les probabilités de transition correspondant à ce processus de mélange).
2. Cette chaîne est-elle irréductible ? Apériodique ?
3. Quelle pourrait bien être sa mesure d'équilibre ?
4. D'une manière générale, si notre croupier choisit une méthode de mélange modélisable au moyen d'une chaîne de Markov irréductible et apériodique, quelle devrait être la mesure d'équilibre correspondante ? Comment mesurer la vitesse de convergence de cette chaîne vers son équilibre ?

- **Exercice 9 : une application tennistique**
On rappelle qu'au tennis, un set est constitué d'une suite de jeux remportés par l'un ou l'autre des deux joueurs, et que le set lui-même est empoché par le premier des deux deux joueurs parvenant à remporter 6 ou 7 jeux et deux jeux de plus que son adversaire ; si toutefois les deux joueurs parviennent au score de 6-6 sans que le set se soit terminé au préalable, ils doivent disputer un dernier jeu appelé "jeu décisif" (*"tie break"* en anglais).
Deux joueurs (J_1 et J_2) s'apprêtent à disputer un set, et on part du principe que chaque jeu - y compris l'éventuel *tie break* - pourra être remporté avec probabilité $p = 0,55$ par J_1 (et avec prob. $1 - p = 0,45$ par J_2).

1. Soit $(X_n)_{n \geq 0}$ le processus comptant le nombre de jeux remportés par J_1 au fil du temps. Quel est l'espace d'états E de ce processus ?

2. Le processus $(X_n)_{n \geq 0}$ constitue-t-il une chaîne de Markov (homogène) sur E ?

3. Soit $(Y_n)_{n \geq 0}$ le processus comptant le nombre de jeux remportés par J_2 au fil du temps, puis $(Z_n) = ((X_n; Y_n))$. Quel est l'espace d'états $\tilde{E}$ de $(Z_n)_{n \geq 0}$?

4. Montrer que $(Z_n)_{n \geq 0}$ constitue une chaîne de Markov (homogène) sur $\tilde{E}$ en spécifiant son graphe de transitions et la matrice stochastique correspondante P.

5. On souhaite à présent prendre en compte une certaine asymétrie en modélisant comme suit le déroulement du set : un jeu autre que l'éventuel *tie break* sera remporté par J_1 avec prob. $0,75$ s'il est lui-même au service, tandis que s'il reçoit, sa probabilité de remporter le jeu s'élève à $0,35$; quant à l'éventuel *tie break*, on considère toujours qu'il sera remporté par J_1 avec prob. $0,55$.
Comment modéliser l'évolution du score dans ce set par une chaîne de Markov ?
Comment calculeriez-vous la prob. qu'a J_1 de remporter ce set ?
Et si le match est disputé au meilleur des trois sets, les scores des sets successifs évoluant indépendamment et suivant ces mêmes modalités, comment calculeriez-vous la prob. qu'a J_1 de sortir vainqueur de ce match ?

- **Exercice 10 : modélisation simplifiée d'un protocole de type *ALOHA***
  Deux utilisateurs tentent de partager équitablement un même canal de communication. A chaque nouvel instant, chacun des deux utilisateurs cherche avec probabilité $p = 0,1$ à envoyer un paquet sur ce canal, tandis qu'avec probabilité $1 - p = 0,9$ il ne cherche pas à envoyer de nouveau paquet.
  Cependant le canal en question peut faire passer tout au plus *un seul paquet* à chaque instant ; dès lors que plusieurs paquets se présentent simultanément à l'entrée du canal, il y a conflit, et ce conflit est résolu équitablement de la façon suivante : les utilisateurs ayant tenté de faire passer un paquet le reprennent, et durant les instants suivants ils tenteront avec probabilité $\nu = 0,2$ de faire passer ce même paquet, cessant par ailleurs de tenter de faire passer encore un autre paquet.
  Chacun des utilisateurs peut donc être *encombré* d'un "ancien paquet" ayant donné lieu à un conflit, ou encore *libre* de tout "ancien paquet". Dans le 1er cas, il tentera avec prob. ν de faire passer un paquet sur le canal à l'instant suivant, dans le 2ème cas cette probabilité devient p.

  1. Soit X_n la variable prenant la valeur 0 si le 1er utilisateur est *"libre"* à l'instant n et la valeur 1 s'il est *encombré*.
     Peut-on considérer que le processus (X_n) constitue à lui seul une chaîne de Markov ?
  2. Soit Y_n la variable prenant la valeur 0 si le 2ème utilisateur est *"libre"* à l'instant n et la valeur 1 s'il est *encombré*.
     Vérifier que le couple $(X_n; Y_n)$ constitue une chaîne de Markov à 4 états en spécifiant le graphe de transitions et la matrice de transitions correspondants.
  3. Quelle est la mesure de probabilité invariante $\pi^{(inv.)}$ associée à cette chaîne ? Pourquoi sait-on à l'avance que cette chaîne admet *une et une seule mesure invariante* ??
  4. Peut-on affirmer que *Loi* $(X_n; Y_n)$ converge "rapidement" vers $\pi^{(inv.)}$?
  5. Quelle est la proportion de temps où le 1er utilisateur est encombré ?
  6. Le 1er utilisateur doit maintenant transmettre plus de paquets à long terme, son paramètre p passant de la valeur $0,1$ à $p_1 = 0,15$, tandis que son paramètre ν ne change pas ($\nu_1 = \nu = 0,2$).
     Par souci d'équité, le 2ème utilisateur décide alors d'utiliser une autre valeur pour le paramètre ν, passant de ν à $\nu_2 > \nu$ - son paramètre p demeurant par ailleurs inchangé ($p_2 = p = 0,1$).
     Dans ces conditions, comment devrait-il régler la valeur de ν_2 pour faire en sorte que son *"temps d'encombrement"* soit ni plus ni moins important que celui du 1er utilisateur ?
  7. Reprendre les questions précédentes dans un contexte à trois utilisateurs (donnant lieu à l'étude d'une chaîne $(X_n; Y_n; Z_n)$ à 8 états).

IUT de Paris Descartes, Dpt INFO

Exercices de Statistique (S3)
TCL et Intervalles de Confiance

A. **Votes des électeurs indécis :**

Les candidats BO et MR s'affrontent pour obtenir les faveurs d'un corps électoral comportant un million de votants.

Le jour de l'élection, sur tous ces votants, 1000 seulement ont pris une décision claire, celle de voter pour BO. Quant aux 999'000 autres, ils sont totalement indécis et choisissent de remplir leur devoir civique en tirant à Pile ou Face dans l'isoloir pour voter ensuite en faveur de BO ou de MR.

Quelle est la loi de la variable X comptant le nombre de bulletins favorables à MR ?

Comment estimer la probabilité de voir BO remporter ces élections ?

B. **TCL appliqué aux variables de Poisson :**

Un Professeur de Psychologie du Rêve sait que le nombre d'étudiants s'inscrivant son cours de Master I suit une loi de Poisson de paramètre lambda=120. La salle qui lui est allouée comporte 135 places seulement, et si le nombre d'inscrits dépasse cette capacité, le cours devra être dédoublé.

Quelle est la probabilité d'un tel dédoublement ?

Rappel utile : une somme de variables de Poisson *indépendantes* est encore poissonienne ...

C. **Fluctuations d'une cote de popularité :**

Dans un grand pays démocratique, un quotidien publie chaque mois la "cote" du chef du gouvernement à partir des résultats d'un sondage effectué sur un échantillon représentatif de $n = 1000$ personnes.

Les sondages de janvier et février ont donné respectivement les cotes de 38% et 36%.

Dans ces conditions, quels sont les intervalles de confiance I_J et I_F contenant les véritables cotes p_J puis p_F correspondant à ces deux mois, au niveau de confiance 95% ? Peut-on raisonnablement affirmer que cette cote de popularité a chuté de deux points entre janvier et février ??

D. Sondages et Audimat :

On interroge 400 téléspectateurs choisis au hasard. Parmi eux, 152 individus déclarent avoir regardé la nouvelle émission *"Perdu en Translation"* diffusée le mardi soir, les 248 autres déclarent ne pas avoir regardé cette nouvelle et passionnante émission.

1. En donnant successivement à $(1-\alpha)$ les valeurs 90%, 95%, 98% puis 99%, estimer par un intervalle de confiance au niveau $(1-\alpha)$ le pourcentage de téléspectateurs ayant regardé cette émission.
2. Quelle aurait dû être la taille n de l'échantillon de téléspectateurs interrogés si le but avait été de déterminer ce pourcentage à 2% près, pour chacun des niveaux de confiance envisagés dans la question précédente ?
3. Tracer deux courbes correspondant aux bornes des intervalles de confiance au niveau 95% en fonction de n (on pourra donner à n les valeurs 100, 150, 200, 250, ..., 1000).

E. Etre ou ne pas être ... élu :

A la veille d'une importante consultation électorale, on effectue un sondage sur un échantillon de $n = 100$ votants. Cet échantillon fournit 55 intentions de vote favorables au candidat A.

1. Donner un intervalle de confiance au niveau $0,99$ pour le pourcentage d'électeurs favorables au candidat A.
2. Quelle doit être la taille n de l'échantillon pour que la largeur de l'intervalle de confiance au niveau 99% soit au plus de 12% ?
3. Quelle devrait être la taille de l'échantillon fournissant 55% d'intentions de vote favorables au candidat A pour que ce candidat soit sûr au niveau de risque 1% d'être élu ??

F. Intervalles de confiance et fiabilité des sondages :

On considère un échantillon de 1000 personnes auprès desquelles on effectue un sondage avant un référendum : 450 répondent *oui*.

1) Quelle valeur proposer pour estimer la probabilité p, probabilité qu'un électeur vote *oui* le jour du scrutin ?

2) Calculer un intervalle de confiance, centré sur cette estimation, pour le paramètre p avec une confiance de 0,95 puis 0,98.

3) Proposer un intervalle de la forme $[0; \dots]$ dans lequel se trouve p avec une confiance de 0,95.

4) Quelle confiance accordez-vous à l'affirmation: "le *non* va l'emporter"?

5) Combien faut-il interroger de personnes pour obtenir une fourchette pour p de largeur $2 \times 0,01$?

G. **Atmosphère, atmosphère :**

On suppose que le taux de prsence d'un gaz nocif dans l'atmosphère, mesuré en $\mu g/m^3$, suit une loi normale d'espérance τ inconnue et d'écart-type σ *lui aussi inconnu.* dix mesures de ce taux, ce qui permet d'obtenir des valeurs $x_1, x_2, \ldots, x_{10}$ telles que $\sum_{i=1}^{10} x_i = 504 g/m^3$ et $\sum_{i=1}^{10} x_i = 26'425\,(g/m^3)^2$.

On effectue tout d'abord une centaine de mesures, obtenant ainsi des valeurs $x_1, x_2, \ldots, x_{100}$ telles que $\sum_{i=1}^{100} x_i = 5065 g/m^3$ et $\sum_{i=1}^{100} (x_i)^2 = 268'742\,(g/m^3)^2$; une seconde équipe effectue effectue 400 mesures, obtenant des valeurs $x'_1, x'_2, \ldots, x'_{400}$ telles que $\sum_{i=1}^{400} x'_i = 20208 g/m^3$ et $\sum_{i=1}^{100} (x'_i)^2 = 1'068'200\,(g/m^3)^2$.

Dans ces conditions, quels sont les intervalles de confiance au niveau 90% obtenus par chacune des deux équipes pour le paramètre τ ?

H. **Tests portant sur une proportion :**

1. Dans la population française, le pourcentage d'individus dont le sang est de rhésus négatif s'élève à 15%.
   Dans un échantillon représentatif de 200 Basques français, on observe que 44 individus sont de rhésus négatif.
   Au vu de ces résultats, peut-on affirmer au risque $\alpha = 5\%$ que les Basques diffèrent du reste de la France pour ce qui est de la caractéristique "rhésus sanguin" ?

2. On considère une population où le pourcentage d'individus présentant des rides s'élève à 25%, et l'on demande à 200 individus choisis au hasard d'utiliser une crème anti-rides. Quelque temps plus tard, on constate que 40 personnes parmi ces 200 individus présentent des rides.
   Au risque $\alpha = 5\%$, peut-on se permettre de rejeter une hypothèse H_0 d'efficacité de ce traitement anti-rides ?

I. **Conformité d'un dosage de substance active**

Les spécifications d'un certain médicament indiquent que chaque comprimé doit contenir $2,5\,g$ de substance active.

Cent comprimés sont choisis au hasard dans une certaine production puis analysés ; il s'avère que ces comprimés contiennent en moyenne $2,6\,g$ de substance active, avec un écart-type empirique s valant $0,4\,g$.

Au risque $\alpha = 5\%$, peut-on considérer que cette production respecte les spécifications prescrites ?

5.4 Exercices d'Application

I Intervalles de confiance et fiabilité des sondages :

On considère un échantillon de 1000 personnes auprès desquelles on effectue un sondage avant un référendum : 450 répondent *oui*.

1) Quelle valeur proposer pour estimer la probabilité p, probabilité qu'un électeur vote *oui* le jour du scrutin ?

2) Calculer un intervalle de confiance, centré sur cette estimation, pour le paramètre p avec une confiance de 0,95 puis 0,98.

3) Proposer un intervalle de la forme $[0; \dots]$ dans lequel se trouve p avec une confiance de 0,95.

4) Quelle confiance accordez-vous à l'affirmation: "le *non* va l'emporter"?

5) Combien faut-il interroger de personnes pour obtenir une fourchette pour p de largeur $2 \times 0,01$?

II Arrondis et TCL :

Dans un supermarché, aprés avoir pesé une marchandise, la balance électronique délivre un ticket où le prix est arrondi au centime près.

1) Pour une pesée, on note X la variable aléatoire donnant, en centimes, la différence entre le prix indiqué sur l'étiquette et le prix exact. Expliquer pourquoi on peut modéliser la loi de X par une loi uniforme sur l'intervalle $[-0,5; +0,5]$. Calculer l'espérance et la variance de X.

2) Quelle est la perte moyenne du supermarché pour 10^4 unités pesées? Quelle est la probabilité que le supermarché gagne plus de 20 centimes pour 10^4 unités? perde plus de 50 euros pour 10^6 unités?

III Stratégies d'examen :

Un QCM comporte 100 questions. Deux réponses sont proposées pour chaque question, l'une d'entre elles est juste, l'autre est fausse. Un étudiant décide de répondre à chaque question en choisissant au hasard sa réponse.

1) Calculer la probabilité qu'il ait au moins 50 réponses justes.

2) Compléter les phrases " Avec une probabilité de 95%, le nombre de réponses justes est supérieur à ..." et "Avec 5% de chance, le nombre de réponses justes est supérieur à ..."

3) Le professeur hésite entre deux modes de notation. Le premier consiste à noter +1 les réponses justes, 0 sinon ; le second consiste à noter +1 les réponses justes, 0 les non-réponses et -1 les réponses fausses. On appelle Y et Z les notes (aléatoires) qu'obtiendra l'étudiant avec le premier mode et avec le second mode de calcul. Quelles notes moyennes l'étudiant peut-il espérer? A-t-il intérêt à changer de stratégie en choisissant entre "non-réponse", "réponse 1" ou "réponse 2" avec les probabilités respectives q, p et p ?

IV TCL et Mesures Physiques :

En poste à l'Observatoire de Göttingen, l'astronome *Carl-Friedrich G.* décide de mesurer la distance séparant son lieu de travail d'une étoile lointaine.
Connaissant fort bien le matériel dont il dispose et prenant en compte les aléas liés à chaque mesure (perturbations atmosphériques, etc...), il part du principe que chaque mesure M ne lui fournit qu'une valeur aléatoire d'espérance d (vraie valeur de cette distance, en années-lumière) et d'écart-type 2 années lumière.
Dans ces conditions, il décide de prendre un certain nombre de mesures *indépendantes* $M_1, M_2, \ldots, M_n$, puis d'adopter leur moyenne

$$\overline{M}_n = \frac{M_1 + M_2 + \ldots + M_n}{n}$$

comme évaluation de cette distance.
Combien de mesures indépendantes notre astronome doit-il effectuer pour être raisonnablement sûr (sûr à 95%) de ne pas se tromper au-delà d'une demi-année lumière dans son évaluation?

V TCL et garanties :

Un fabriquant de clefs USB sait qu'une proportion de 0,05 de sa production est défectueuse.
1) Il garantit à un client qui achète 10000 pièces de la rembourser si plus de m clefs sont défectueuses. Comment le fabriquant doit-il choisir m pour n'avoir pas plus d'une "chance" sur cent d'avoir à rembourser son client?
2) Le service des fraudes vient faire un contrôle, prélève 1000 clefs et note S le nombre de clefs défectueuses.
a) Donner un intervalle de la forme $[50 - a, 50 + a]$ dans lequel S prend ses valeurs avec une probabilité supérieure à 0,95.
b) Idem avec un intervalle de la forme $[0, x]$, puis $[y, \infty[$. Traduire les résultats par des phrases.

VI Surcharge :

Un avion long courrier peut transporter 100 passagers et leurs bagages. L'avion pèse 120 tonnes sans passagers ni bagages, mais équipage compris et le plein de carburant effectué. Les consignes de sécurité interdisent au commandant de bord de décoller si le poids de l'appareil chargé dépasse 129 tonnes. Les 100 places ont été réservées. Le poids d'un voyageur est une v.a. X d'espérance mathématique 70 kg et d' écart-type 10 kg. Le poids des bagages d'un voyageur est une v.a. Y d'espérance mathématique 20 kg et d'écart-type 10 kg. Toutes ces variables aléatoires sont mutuellement indépendantes, leurs lois de probabilité sont inconnues.
1) Soit C la v.a. égale au poids de l'avion lorsque les 100 voyageurs et leurs bagages seront chargés.

a- Donner l'expression en kilos de C en fonction des différentes v.a. décrites dans l'énoncé.
b- Calculer l'espérance et la variance de C et donner la loi de C.
c- Quelle est la probabilité p que le commandant refuse d'embarquer une partie des bagages afin de respecter les consignes de sécurité ?
2) L'avion doit effectuer 800 vols dans des conditions identiques à celles décrites ci-dessus. On fait toutes les hypothèses d'indépendance nécessaires. Soit N le nombre de vols où le commandant refuse d'embarquer une partie des bagages afin de respecter les consignes de sécurité.
a- Quelle est la loi de probabilité de N ? Quelle approximation peut-on en faire?
b- Quelle est la probabilité qu'au moins une fois parmi les 800 vols, le commandant refuse d'embarquer une partie des bagages ?

VII Sondages et Audimat :

On interroge 400 téléspectateurs choisis au hasard. Parmi eux, 152 individus déclarent avoir regardé la nouvelle émission *"Perdu en Translation"* diffusée le mardi soir, les 248 autres déclarent ne pas avoir regardé cette nouvelle et passionnante émission.

1. En donnant successivement à $(1-\alpha)$ les valeurs 90%, 95%, 98% puis 99%, estimer par un intervalle de confiance au niveau $(1-\alpha)$ le pourcentage de téléspectateurs ayant regardé cette émission.
2. Quelle aurait dû être la taille n de l'échantillon de téléspectateurs interrogés si le but avait été de déterminer ce pourcentage à 2% près, pour chacun des niveaux de confiance envisagés dans la question précédente ?
3. Tracer deux courbes correspondant aux bornes des intervalles de confiance au niveau 95% en fonction de n (on pourra donner à n les valeurs 100, 150, 200, 250, ..., 1000).

VIII Etre ou ne pas être ... élu :

A la veille d'une importante consultation électorale, on effectue un sondage sur un échantillon de $n = 100$ votants. Cet échantillon fournit 55 intentions de vote favorables au candidat A.

1. Donner un intervalle de confiance au niveau $0,99$ pour le pourcentage d'électeurs favorables au candidat A.
2. Quelle doit être la taille n de l'échantillon pour que la largeur de l'intervalle de confiance au niveau 99% soit au plus de 12% ?
3. Quelle devrait être la taille de l'échantillon fournissant 55% d'intentions de vote favorables au candidat A pour que ce candidat soit sûr au niveau de risque 1% d'être élu ??

IX Le TCL est universel ⊞ :
Choisir une loi parmi celles mentionnées dans l'exercice 13 et enregistrer dans la variable m la valeur de l'espérance et dans la variable s l'écart-type de cette loi.

1. Pour $n = 2$ puis $n = 10$ puis $n = 30$ puis $n = 100$, simuler n réalisations de cette loi (désignées par $X_1, \ldots, X_n$) et calculer la moyenne $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ de ces réalisations.

2. Reprendre 1000 fois la question 1) en stockant les 1000 valeurs des moyennes $\bar{X}_n$.

3. Centrer et réduire ces 1000 valeurs en effectuant le calcul de $(\bar{X}_n - m)/(s/\sqrt{n})$.

4. Regrouper les 1000 valeurs de la question précédentes en 10 classes centrées en 0 et de largeur 1, et tracer l'histogramme correspondant.

5. Comparer avec les histogrammes des autres étudiants du groupe. Que constatez-vous? (*effectuer les comparaisons pour les 4 valeurs proposées de n*). Décrire précisément la conséquence du TCL qui vient ainsi d'être illustrée.

X Planche de Galton ⊞ :

- Qu'est-ce qu'une planche de Galton (*Galton Board*, en anglais) ? [3]
- En quoi est-ce qu'une telle planche fournit une illustration du TCL ?
- Comment vous y prendriez-vous pour *simuler* une telle expérience ?

XI Calcul d'une valeur approchée de π ⊞ :
On se propose de calculer une valeur approchée de π par une méthode probabiliste, appelée méthode de Monte-Carlo. Il s'agit en fait d'une illustration de la loi des grands nombres et du théorème central limite.
Méthode : On lance un grand nombre de points dans le carré $[-1, 1] \times [-1, 1]$ et on compte la proportion de points qui se trouvent à l'intérieur du disque de centre $(0, 0)$ et de rayon 1. D'après la loi des grands nombres, cette proportion fournit une valeur approchée de la probabilité de "tomber" à l'intérieur du disque. On en déduit ainsi (cf question 1) ci-dessous) une valeur approchée de π, on dit une “estimation de π”. De plus, le théorème central limite permet de prédire, en fonction du nombre de points lancés, avec une probabilité de 95% l'écart maximal entre la (vraie) valeur de π et la valeur approchée obtenue.

1. Tracer sur une feuille de brouillon “à main levée” le carré $[-1, 1] \times [-1, 1]$ et le disque de centre $(0, 0)$ et de rayon 1.

[3] A titre exceptionnel, on pourra prendre le temps de se renseigner directement sur Internet durant le TP !

Combien vaut l'aire du carré? l'aire du disque?
Quelle est la probabilité qu'un point lancé au hasard et de façon uniforme dans le carré "tombe" à l'intérieur du disque?
Comment traduire qu'un point de coordonnées (x, y) est situé à l'intérieur du disque?

2. Pour une valeur de n à fixer ($n > 35$), simuler le lancer de n points dans le carré (*il suffit de choisir séparément (indépendamment) les abscisses et les ordonnées au hasard dans* $[-1, 1]$).
Représenter ces n points dans un graphique.
Calculer la proportion de points tombés à l'intérieur du disque et stocker l'estimation de π ainsi obtenue.
3. Renouveler l'estimation de la question précédente jusqu'à obtenir N estimations de π (prendre par exemple $N = 1000$), autrement dit N valeurs approchées.
En éliminant les estimations "trop grandes" et "trop petites", dans quel intervalle se situent 95% des estimations?
4. En utilisant le TCL, calculer de façon théorique l'intervalle de fluctuation centré sur π dans lequel doivent se trouver 95% des estimations. Comparer avec la question précédente.

XII Estimation de Taux de Fertilité :

On considère un échantillon de 169 brebis de la race "Ile de France". Ces brebis ont été mises en lutte, on a obtenu 108 brebis pleines (c'est à dire fécondées).
A l'issue de cette expérience, on cherche à estimer le taux de fertilité t de la race "Ile de France".

QCM1 : Soit $\hat{t}$ l'estimateur de t obtenu par la méthode du maximum de vraisemblance. Peut-on affirmer que

A $\hat{t}$ est un estimateur sans biais de t ?

B $\hat{t}$ est un estimateur biaisé de t ?

C $\hat{t} = 69,2\%$?

D $\hat{t} = 63,9\%$?

E $\hat{t} = 59,2\%$?

QCM2 : Un intervalle de confiance au niveau 95% pour ce taux t est donné par

A $I = [0,636; 0,651]$?

B $I = [0,586; 0,651]$?

C $I = [0,566; 0,721]$?

D $I = [0,546; 0,731]$?

E $I = [0,586; 0,601]$?

QCM3 : Peut-on affirmer que

A Un intervalle de confiance au niveau 99% pour t sera plus large que celui obtenu au niveau 95% ?

B Un intervalle de confiance au niveau 90% pour t sera plus large que celui obtenu au niveau 95% ?

C Un intervalle de confiance au niveau 99% pour t sera plus étroit que celui obtenu au niveau 95% ?

D Avec 200 brebis plutôt que 169, même si l'estimée ponctuelle de t devait différer de celle obtenue précédemment, on serait en mesure de fournir un intervalle de confiance plus étroit pour t au niveau 95% ?

E Avec 500 brebis plutôt que 169, si l'estimée ponctuelle de t est proche de celle obtenue précédemment, on sera en mesure de fournir un intervalle de confiance plus étroit pour t au niveau 95% ?

X Taux d'échec d'un vaccin :

Des études antérieures ont révélé que le taux d'échec τ associé à une certaine vaccination est situé entre 10% et 15%.

On prépare une expérience dans le but de déterminer à 1% près la proportion de sujets non-immunisés par ce vaccin, en acceptant un coefficient de risque $\alpha = 0,05$. n personnes seront donc vaccinées, et l'on relèvera par la suite qui parmi ces n personnes n'est pas immunisé à l'issue de la vaccination.

QCM4 : Peut-on affirmer que

A Le nombre de sujets non-immunisés à l'issue de cette nouvelle campagne suit une loi de Poisson ?

B Le nombre de sujets non-immunisés à l'issue de cette nouvelle campagne suit une loi binomiale ?

C Le nombre de sujets non-immunisés à l'issue de cette nouvelle campagne suit approximativement une loi de Poisson ?

D Les variables "nombre de sujets non-immunisés" et "nombre de sujets immunisés" sont indépendantes ?

E Les variables "nombre de sujets non-immunisés" et "nombre de sujets immunisés" sont négativement corrélées ?

QCM5 : Pour parvenir à ses fins (déterminer à 1% près et au niveau de confiance 95% la proportion de sujets non-immunisés), quel est le nombre minimal n_0 de sujets que l'on doit vacciner ?

A $n_0 = 100$

B $n_0 = 1000$

C $n_0 = 3900$

D $n_0 = 4900$

E $n_0 = 5400$

XIII Encore des sondages :

A la veille d'une consultation électorale majeure, on a interrogé une centaine d'électeurs constituant un échantillon représentatif. 58 d'entre eux ont déclaré avoir l'intention de voter pour le candidat *Toulemonde*.

QCM6 : Quelles sont les propositions correctes parmi celles ci-dessous ?

A $I = [0,453; 0,707]$ constitue un intervalle de confiance au niveau 95% pour la cote de M. Toulemonde.

B $I = [0,483; 0,677]$ constitue un intervalle de confiance au niveau 95% pour la cote de M. Toulemonde.

C $I = [0,453; 0,707]$ constitue un intervalle de confiance au niveau 99% pour la cote de M. Toulemonde.

D Au niveau de confiance 99%, on peut affirmer que M. Toulemonde va remporter ces élections.

E Au niveau de confiance 99%, on n'est pas en mesure de garantir que M. Toulemonde va remporter ces élections.

QCM7 : Pour une même fréquence observée d'électeurs favorables à M. Toulemonde, quelle devrait être la taille minimale n_0 de l'échantillon permettant d'affirmer au niveau de confiance 95% que M. Toulemonde sera élu ?

A $n_0 = 100$

B $n_0 = 103$

C $n_0 = 147$

D $n_0 = 200$

E $n_0 = 10000$

XIV Estimations de valeurs moyennes au sein de deux populations :

Dans une région d'Europe, l'étude de la masse du cerveau mesurée en grammes chez des sujets âgés de 20 à 49 ans a conduit aux résultats suivants :

Hommes

Val. approx.	*1170*	*1220*	*1270*	*1320*	*1370*	*1420*	*1470*	**Total**
Effectifs	*5*	*36*	*45*	*50*	*61*	*49*	*19*	*265*

IUT DE PARIS

Femmes								
Val. approx.	*1070*	*1120*	*1170*	*1220*	*1270*	*1320*	*1370*	Total
Effectifs	*12*	*22*	*45*	*54*	*52*	*20*	*10*	*215*

QCM8 : Quel est l'intervalle de confiance au niveau 99% pour la valeur moyenne de cette masse au sein de la population masculine ?

A $I = [1326; 1345]$

B $I = [1323; 1349]$

C $I = [1328; 1344]$

D $I = [1322; 1350]$

E $I = [1320; 1351]$

QCM9 : Quel est l'intervalle de confiance au niveau 99% pour la valeur moyenne de cette masse au sein de la population féminine ?

A $I' = [1203; 1230]$

B $I' = [1209; 1233]$

C $I' = [1209; 1229]$

D $I' = [1206; 1233]$

E $I' = [1206; 1330]$

QCM10 : Quelles sont les propositions correctes parmi celles ci-dessous ?

A Pour l'ensemble de la population, une estimation ponctuelle de cette masse moyenne s'obtient en effectuant la demi-somme $\frac{1}{2}(\hat{m}_h + \hat{m}_f)$.

B Pour l'ensemble de la population, une estimation ponctuelle de cette masse moyenne s'obtient en effectuant une somme $\frac{265}{480}\hat{m}_h + \frac{215}{480}\hat{m}_f$.

C Pour l'ensemble de la population, l'intervalle de confiance au niveau 99% pour cette valeur moyenne est moins large que pour la seule population masculine ?

D Pour l'ensemble de la population, l'intervalle de confiance au niveau 99% pour cette valeur moyenne est plus large que pour la seule population masculine ?

E Pour l'ensemble de la population, l'intervalle de confiance au niveau 99% pour cette valeur moyenne est de même largeur que pour la seule population masculine ?

XV Estimation de la fréquentation d'un hôpital :

On considère un hôpital comportant cent salles de consultation. Chacune de ces salles accueille quotidiennement un nombre de patients qui est modélisé par une variable de Poisson de paramètre $\lambda = 10$, et l'on suppose que ces variables de Poisson sont indépendantes. Soit

S le nombre total de patients qui viennent pour une consultation à l'hôpital un jour donné, puis $M = S/100$ le nombre moyen de consultations par salle ce jour-là.
QCM11 : Quelles sont les propositions correctes parmi celles ci-dessous ?

A $\mathbb{E}(M) = 0,1$

B $\mathbb{E}(M) = 10$

C $\mathbb{E}(S) = 100$

D $\mathbb{E}(S) = 1000$

E $\mathrm{Var}(S) = 10000$

QCM12 : Quelle est la probabilité de voir S dépasser la valeur 1050 un jour donné (à 10^{-2} près) ?

A $0,06$

B $0,11$

C $0,26$

D $0,48$

E $0,96$

XVI Estimation d'un paramètre poissonien :

On suppose que le nombre d'accidents survenant dans une certaine ville un jour donné suit une loi de Poisson de paramètre λ inconnu. Pendant un an, on a relevé quotidiennement le nombre X d'accidents survenus dans cette ville durant la journée, le tableau ci-dessous résume les résultats de cette enquête :

Nombre d'accidents dans la journée

Val. observ.	***0***	***1***	***2***	***3***	***Total***
Nombre de j.	***200***	***100***	***55***	***10***	***365***

On part du principe que les variables $X_1, X_2, \ldots, X_{365}$ associées à chaque journée sont indépendantes, et l'on note M la variable aléatoire de moyenne annuelle des nombres d'accidents quotidiens.
QCM13 : Peut-on affirmer que

A M suit exactement une loi de Poisson ?

B M suit approximativement une loi binomiale ?

C M suit approximativement une loi normale ?

D $\mathbb{E}(M) = \lambda$?

E $\mathrm{Var}(M) = \lambda$?

QCM14 : On note s^2 la variance empirique obtenue pour ces variables de Poisson $X_1, X_2, \ldots, X_{365}$ et $\hat{x}$ leur moyenne empirique. Peut-on affirmer que

A $\hat{x} = 0,66$?

B $\hat{x} = 91,3$?

C $s^2 = 0,69$?

D $s^2 = 6,85$?

E $s^2 = 68,5$?

QCM15 : On choisit de noter $I_{1-\alpha}$ un intervalle de confiance de niveau $(1-\alpha)$ pour le paramètre λ. En arrondissant les bornes à 10^{-2} près, peut-on affirmer que

A $I_{0,95} = [0,57; 0,75]$?

B $I_{0,95} = [0,64; 0,68]$?

C $I_{0,95} = [90,5; 92,1]$?

D $I_{0,90} = [0,59; 0,73]$?

E $I_{0,90} = [90,8; 91,8]$?

XVII Mesures et précision :

Un appareil dose la concentration d'une substance sans biais, mais avec une erreur de mesure suivant une loi normale de moyenne nulle et de variance $100\,(mg/L)^2$. Si on effectue plusieurs mesures d'une concentration sur un même prélèvement, on suppose que toutes les erreurs de mesures sont indépendantes. En répétant les mesures puis en choisissant leur moyenne comme résultat, on souhaite obtenir une précision (demi-longueur d'int. de confiance de niveau 95%) inférieure ou égale à $5\,mg/L$.

QCM16 : Combien faut-il effectuer de mesures pour obtenir la précision désirée ?

A 5

B 10

C 12

D 16

E 20

QCM17 : Avec l'appareil de la question précédente, le coût d'une mesure est de 100 Euros. On se demande s'il est rentable d'acheter un nouvel appareil pour doser la concentration de la substance. Un nouvel appareil produit des mesures sans biais et comportant une erreur de variance $40\,(mg/L)^2$ seulement, mais le coût de chaque mesure avec ce nouvel appareil s'élève à 200 Euros. Ici encore, l'objectif est d'obtenir une précision inférieure ou égale à $5\,mg/L$, quitte à combiner de nombreuses mesures. Peut-on affirmer que

A Pour atteindre ce but, il faut effectuer 5 mesures avec le nouvel appareil ?

B Pour atteindre ce but, il faut effectuer 7 mesures avec le nouvel appareil ?

C Pour atteindre ce but, il faut effectuer 9 mesures avec le nouvel appareil ?

D Il vaut mieux garder l'ancien appareil de dosage.

E Il vaut mieux acheter le nouvel appareil de dosage.

XVIII Encore des sondages :

Une élection oppose deux candidats, et tout porte à croire qu'elle sera très serrée : on part du principe que tous les électeurs vont voter pour l'un ou l'autre candidat, et que les scores de chacun d'eux seront proches de 50%. On souhaite effectuer un sondage à la sortie des bureaux de vote.

QCM18 : Combien d'électeurs faut-il interroger pour atteindre une précision (demi-longueur d'intervalle de confiance au niveau 95%) de $0,02$?

A 996

B 1010

C 1563

D 2401

E 2542

QCM19 : Si l'on choisit d'interroger 1000 électeurs, quelle sera la précision obtenue pour les proportions de voix remportées par chaque candidat ?

A $0,01$

B $0,03$

C $0,05$

D $0,07$

E $0,10$

6.4 Exercices d'Application

I Echantillonnage et Tests d'Efficacité :

1. Au niveau national, en 2014, les candidats aux trois types de baccalauréat se sont répartis de la façon suivante: général 47% ; technologique 20% ; professionnel 33%.
   Sur un échantillon de 250 candidats, on a observé la répartition d'effectifs suivante : général 131 ; technologique 37 ; professionnel 82.
   Cet échantillon est-il représentatif de la population nationale? (*Vous répondrez en effectuant un test au niveau de risque 5% puis 10%.*)

2. Les tableaux suivants proviennent d'une enquête effectuée auprès de jeunes ayant subi les épreuves du baccalauréat général ou technologique.

Bac général

	Présents	Admis
Garçons	150	132
Filles	186	173

Bac technologique

	Présents	Admis
Garçons	70	62
Filles	72	67

(a) Transformer le tableau concernant le bac général pour qu'il se présente sous la forme d'un tableau croisé des variables "Sexe" et "Réussite", prenant les valeurs respectives $\{G, F\}$ et $\{Admis, Pas\ admis\}$.

(b) Effectuer un test au niveau de risque 5% pour décider si, au vu de l'échantillon, la phrase suivante est fondée: "Au bac général, les filles réussissent mieux que les garçons."
(*Indic: on trouve une distance du khi-deux égale 2.49*)

(c) En ce qui concerne le bac technologique, le même test conduit à une valeur de khi-deux égale à 0.86 (on ne demande pas de faire ce calcul). Pour quel type de bac, général ou technologique, la réussite est-elle plus conditionnée au sexe? (*justifiez votre réponse*)

(d) Les chiffres des tableaux correspondent **en milliers** aux résultats nationaux du bac 2014. Que répondez-vous la question posée en **b)** en tenant compte de toute la population?
e)* Utiliser les deux tableaux de l'enquête pour tester l'indépendance entre le sexe et le choix de la filière technologique ou générale.

II Dé électronique ⊞ :

On considère un dé électronique à 6 faces, numérotées de 1 à 6. On a "lancé" le dé 1000 fois et obtenu les résultats suivants

face	1	2	3	4	5	6
effectif	155	170	155	152	168	200

1. A votre avis, ce dé est-il équilibré? *Vous répondrez en effectuant un test d'ajustement au niveau de risque de votre choix.*
2. Auriez-vous fourni pareille réponse si les mêmes proportions avaient été observées sur 10000 lancers ?
3. Tester l'hypothèse selon laquelle le dé est pipé en faveur du 6 avec probabilité $\frac{1}{6}+\frac{1}{10}$, les autres faces étant équiprobables entre elles.
4. ⊞ Simuler 1000 lancers de 2 dés à 6 faces équilibrés. Afficher les résultats obtenus pour la somme des deux dés dans un tableau d'effectifs. Effectuer un test d'ajustement avec la loi théorique attendue.

III Tests de générateurs aléatoires ⊞ :

Pour générer des nombres aléatoires, qu'ils soient entiers ou réels, `Scilab` dispose de fonctions prédéfinies qui produisent des lois précises. Voici quelques exemples que vous trouverez, ainsi que d'autres, dans la rubrique `grand` (prononcez "g-rand") de l'aide.
- lois discrètes pour générer des nombres entiers : loi Binomiale, loi Uniforme dans $\{1, 2, \ldots, n\}$, loi de Poisson,
- lois continues pour générer des nombres réels : loi Uniforme dans $[a, b]$, loi Normale, loi Exponentielle.

1. Choisir une loi discrète, et les paramètres nécessaires, parmi celles proposées ci-dessus et simuler un échantillon de taille 1000 de cette loi.
   Calculer la moyenne de l'échantillon et comparer avec la moyenne attendue.
   Tracer un histogramme de la répartition de l'échantillon.
2. Construire une fonction `testkhi2` qui permet d'effectuer un test d'ajustement. Cette fonction prendra comme variables d'entrée `(obs,theo,niv)` et fournira en variables de sortie `[dist,concl]`, où
   - `obs,theo` désignent deux vecteurs lignes de même longueur contenant les effectifs observés et les effectifs théoriques dans un tableau d'effectifs,
   - `niv` désigne le niveau du test à fixer entre 0 et 1,
   - `dist` désigne la distance du khi-deux entre le tableau observé et le tableau théorique,
   - `concl` désigne la conclusion du test d'ajustement sous la forme `rejet` ou `pas rejet`.
3. Utiliser la fonction de la question b) pour effectuer un test d'ajustement sur les données simulées à la question a).
4. Reprendre les questions a) et c) avec une loi uniforme dans $[0, 1]$, puis avec une loi uniforme dans $[a, b]$ (à vous de choisir a et b), puis (*) avec une autre loi continue.

IV Virus :

On désire tester la polyvalence d'un traitement sur une maladie dont le virus présente plusieurs souches à la suite de mutations génétiques. On effectue donc des tests sur des patients volontaires atteints par différentes souches, ce qui donne les résultats suivants :

Souche	1	2	3	4	5	6
Nb. malades	575	988	2240	209	210	287
Nb. guérisons	48	89	180	17	18	13

Tester au niveau $\alpha = 5\%$ l'hypothèse : "la souche n'intervient pas dans les chances de guérison". *Attention : effectuer un test d'indépendance entre deux variables clairement identifiées et adapter le tableau en conséquence!*

V Echantillonnage et Tests d'Efficacité :

On sait qu'une maladie atteint 10% des jeunes ovins d'une région donnée. Un chercheur a expérimenté un traitement sur un échantillon de n agneaux. Il a recensé 5% d'agneaux malades sur cet échantillon.
Quelle est la valeur minimale de l'entier n (taille de l'échantillon d'agneaux traités) permettant de conclure, au risque $\alpha = 5\%$, à une certaine efficacité du traitement appliqué ?

VI To smoke ... or not to smoke :

Après avoir suivi pendant vingt ans un groupe de 200 sujets, on a observé les résultats suivants en comptabilisant les cancers apparus dans chacun des deux groupes :

	Non–fumeurs	*Fumeurs*
Apparition d'un cancer	*20*	*40*
Pas de cancer	*180*	*160*

Utiliser deux méthodes de test différentes pour décider, au risque $\alpha = 5\%$, si les différences observées entre fumeurs et non-fumeurs sont significatives ou non.

VII Réussite au Baccalauréat :

Le taux de réussite au Baccalauréat d'une certaine série, une année donnée, est de 67%. Pour les deux questions qui suivent, on effectuera des tests au niveau de risque $\alpha = 5\%$.

1. Dans le centre d'examen A, il y a eu 216 reçus pour 300 candidats présents. Les résultats de ce centre sont-ils conformes aux résultats nationaux ?
2. Dans un centre d'examen B situé dans la même ville, il y a eu 128 reçus pour 200 candidats. Les résultats des centres A et B sont-ils significativement différents ?

VIII Botanique Mendélienne :

On a effectué le croisement de balsamines blanches avec des balsamines pourpres. En première génération, les fleurs sont toutes pourpres. On obtient en deuxième génération quatre catégories, avec les effectifs suivants :

Couleurs	*Pourpre*	*Rose*	*Blanc lavande*	*Blanc*	Total
Effectifs obs	***1790***	***547***	***548***	***213***	3098
Effectifs théoriques	$\frac{9}{16} \times 3098$	$\frac{3}{16} \times 3098$	$\frac{3}{16} \times 3098$	$\frac{3098}{16}$	3098

Au risque $\alpha = 5\%$, peut-on accepter l'hypothèse de répartition mendélienne $(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16})$?

IX Groupes sanguins et apparitions d'une maladie :

On cherche à savoir si la fréquence d'apparition d'une maladie est liée au groupe sanguin. Sur 200 malades observés, on a dénombré 104 personnes du groupe O, 76 du groupe A, 18 du groupe B et 2 du groupe AB.

On admet que dans la population générale, la répartition entre les groupes est la suivante : 47% d'individus dans le groupe O, 43% dans le groupe A, 7% dans le groupe B et 3% dans le groupe AB. Que peut-on en conclure ?

X Hypothèse d'une répartition gaussienne :

Lors d'une étude biologique portant sur une certaine espèce de mollusque, on a mesuré le taux de protéines x en mg de 36 individus de cette espèce, obtenant les résultats suivants :

x	*]0 ; 1,5]*	*]1,5 ; 3]*	*]3 ; 4,5]*	*]4,5 ; 6]*	*]6 ; 7,5]*	*]7,5 ; 9]*	*]9;10,5]*
Nombre d'individus	***8***	***7***	***4***	***9***	***2***	***3***	***3***

1. Donner des estimations non-biaisées de la moyenne et de l'écart-type associés à la variable x au sein de cette population.
2. Peut-on admettre l'hypothèse selon laquelle ce taux est distribué de façon gaussienne au sein de la population considérée ?

XI Test Poissonien dans une situation réelle Des plaignants[6] ont poursuivi en justice le Ministère israélien de la Santé suite à une campagne de vaccination menée sur des enfants et ayant entraîné des dommages fonctionnels irréversibles pour certains d'entre eux. Ce vaccin était en fait connu pour entraîner de tels dommages en de très rares circonstances ;

[6] cf. *Murray Atkin*, "Evidence and the Posterior Bayes Factor", 17 Math. Scientist 15 (1992)

des études antérieures menées dans d'autres pays ont montré que ce risque était d'un cas sur 310'000 vaccinations. Les plaignants avaient été informés de ce risque et l'avaient accepté. Les doses ayant provoqué les dommages objet de la plainte provenaient d'un lot ayant servi à vacciner un groupe de 300'533 enfants ; dans ce groupe, quatre cas de dommages ont été diagnostiqués.

1. On modélise l'événement *"Le vaccin provoque des dommages fonctionnels irréversibles sur l'enfant i"* par une variable aléatoire de Bernoulli X_i de paramètre p. Calculer la valeur p_0 correspondant aux résultats des études antérieures.
2. Montrer que l'on peut modéliser la loi du nombre N de cas de dommages dans le groupe par une loi de Poisson de paramètre λ. Calculer la valeur λ_0 attendue sous l'hypothèse d'une conformité du vaccin aux études antérieures.
3. L'hypothèse (H_0) : "$p = p_0$" correspond au risque que les plaignants avaient accepté d'encourir, l'hypothèse alternative étant (H_1) : "$p > p_0$".
   Construire, à partir de la variable N, un test de niveau α pour l'hypothèse (H_0). Accepte-t-on (H_0) pour un niveau de risque $\alpha = 5\%$?

XII Locations et saisons :

Des appartements à la montagne peuvent être loués à la semaine. Dans la comparaison des taux d'occupation de ces appartements pour un mois d'hiver (Janvier) et pour un mois d'été (Juillet), on dispose de deux échantillons, l'un de 300 observations instantanées en Janvier, l'autre de 200 observations instantanées en Juillet :

	Janvier	*Juillet*
Occupation	*240*	*150*
Inoccupation	*60*	*50*

Au risque $\alpha = 5\%$, convient-il de considérer que les taux d'occupation de ces appartements sont identiques en Janvier et en Juillet ?

XIII Estimation de Taux de Fertilité :

On considère un échantillon de 169 brebis de la race "Ile de France". Ces brebis ont été mises en lutte, on a obtenu 108 brebis pleines (c'est à dire fécondées).
A l'issue de cette expérience, on cherche à estimer le taux de fertilité t de la race "Ile de France".
QCM1 : Soit $\hat{t}$ l'estimateur de t obtenu par la méthode du maximum de vraisemblance. Peut-on affirmer que

A $\hat{t}$ est un estimateur sans biais de t ?

B $\hat{t}$ est un estimateur biaisé de t ?

C $\hat{t} = 69,2\%$?

D $\hat{t} = 63,9\%$?

E $\hat{t} = 59,2\%$?

QCM2 : Un intervalle de confiance au niveau 95% pour ce taux t est donné par

A $I = [0,636; 0,651]$?

B $I = [0,586; 0,651]$?

C $I = [0,566; 0,721]$?

D $I = [0,546; 0,731]$?

E $I = [0,586; 0,601]$?

QCM3 : Peut-on affirmer que

A Un intervalle de confiance au niveau 99% pour t sera plus large que celui obtenu au niveau 95% ?

B Un intervalle de confiance au niveau 90% pour t sera plus large que celui obtenu au niveau 95% ?

C Un intervalle de confiance au niveau 99% pour t sera plus étroit que celui obtenu au niveau 95% ?

D Avec 200 brebis plutôt que 169, même si l'estimée ponctuelle de t devait différer de celle obtenue précédemment, on serait en mesure de fournir un intervalle de confiance plus étroit pour t au niveau 95% ?

E Avec 500 brebis plutôt que 169, si l'estimée ponctuelle de t est proche de celle obtenue précédemment, on sera en mesure de fournir un intervalle de confiance plus étroit pour t au niveau 95% ?

XIV Taux d'échec d'un vaccin :

Des études antérieures ont révélé que le taux d'échec τ associé à une certaine vaccination est situé entre 10% et 15%.
On prépare une expérience dans le but de déterminer à 1% près la proportion de sujets non-immunisés par ce vaccin, en acceptant un coefficient de risque $\alpha = 0,05$. n personnes seront donc vaccinées, et l'on relèvera par la suite qui parmi ces n personnes n'est pas immunisé à l'issue de la vaccination.

QCM4 : Peut-on affirmer que

A Le nombre de sujets non-immunisés à l'issue de cette nouvelle campagne suit une loi de Poisson ?

B Le nombre de sujets non-immunisés à l'issue de cette nouvelle campagne suit une loi binomiale ?

C Le nombre de sujets non-immunisés à l'issue de cette nouvelle campagne suit approximativement une loi de Poisson ?

D Les variables "nombre de sujets non-immunisés" et "nombre de sujets immunisés" sont indépendantes ?

E Les variables "nombre de sujets non-immunisés" et "nombre de sujets immunisés" sont négativement corrélées ?

QCM5 : Pour parvenir à ses fins (déterminer à 1% près et au niveau de confiance 95% la proportion de sujets non-immunisés), quel est le nombre minimal n_0 de sujets que l'on doit vacciner ?

A $n_0 = 100$

B $n_0 = 1000$

C $n_0 = 3900$

D $n_0 = 4900$

E $n_0 = 5400$

XV Encore des sondages :

A la veille d'une consultation électorale majeure, on a interrogé une centaine d'électeurs constituant un échantillon représentatif. 58 d'entre eux ont déclaré avoir l'intention de voter pour le candidat *Toulemonde.*

QCM6 : Quelles sont les propositions correctes parmi celles ci-dessous ?

A $I = [0,453; 0,707]$ constitue un intervalle de confiance au niveau 95% pour la cote de M. Toulemonde.

B $I = [0,483; 0,677]$ constitue un intervalle de confiance au niveau 95% pour la cote de M. Toulemonde.

C $I = [0,453; 0,707]$ constitue un intervalle de confiance au niveau 99% pour la cote de M. Toulemonde.

D Au niveau de confiance 99%, on peut affirmer que M. Toulemonde va remporter ces élections.

E Au niveau de confiance 99%, on n'est pas en mesure de garantir que M. Toulemonde va remporter ces élections.

QCM7 : Pour une même fréquence observée d'électeurs favorables à M. Toulemonde, quelle devrait être la taille minimale n_0 de l'échantillon permettant d'affirmer au niveau de confiance 95% que M. Toulemonde sera élu ?

A $n_0 = 100$

B $n_0 = 103$

C $n_0 = 147$

D $n_0 = 200$

E $n_0 = 10000$

XVI Estimations de valeurs moyennes au sein de deux populations :

Dans une région d'Europe, l'étude de la masse du cerveau mesurée en grammes chez des sujets âgés de 20 à 49 ans a conduit aux résultats suivants :

Hommes

Val. approx.	*1170*	*1220*	*1270*	*1320*	*1370*	*1420*	*1470*	**Total**
Effectifs	*5*	*36*	*45*	*50*	*61*	*49*	*19*	*265*

Femmes

Val. approx.	*1070*	*1120*	*1170*	*1220*	*1270*	*1320*	*1370*	**Total**
Effectifs	*12*	*22*	*45*	*54*	*52*	*20*	*10*	*215*

QCM8 : Quel est l'intervalle de confiance au niveau 99% pour la valeur moyenne de cette masse au sein de la population masculine ?

A $I = [1326; 1345]$

B $I = [1323; 1349]$

C $I = [1328; 1344]$

D $I = [1322; 1350]$

E $I = [1320; 1351]$

QCM9 : Quel est l'intervalle de confiance au niveau 99% pour la valeur moyenne de cette masse au sein de la population féminine ?

A $I' = [1203; 1230]$

B $I' = [1209; 1233]$

C $I' = [1209; 1229]$

D $I' = [1206; 1233]$

E $I' = [1206; 1330]$

QCM10 : Quelles sont les propositions correctes parmi celles ci-dessous ?

A Pour l'ensemble de la population, une estimation ponctuelle de cette masse moyenne s'obtient en effectuant la demi-somme $\frac{1}{2}(\hat{m}_h + \hat{m}_f)$.

B Pour l'ensemble de la population, une estimation ponctuelle de cette masse moyenne s'obtient en effectuant une somme $\frac{265}{480}\hat{m}_h + \frac{215}{480}\hat{m}_f$.

C Pour l'ensemble de la population, l'intervalle de confiance au niveau 99% pour cette valeur moyenne est moins large que pour la seule population masculine ?

D Pour l'ensemble de la population, l'intervalle de confiance au niveau 99% pour cette valeur moyenne est plus large que pour la seule population masculine ?

E Pour l'ensemble de la population, l'intervalle de confiance au niveau 99% pour cette valeur moyenne est de même largeur que pour la seule population masculine ?

XVII Estimation de la fréquentation d'un hôpital :

On considère un hôpital comportant cent salles de consultation. Chacune de ces salles accueille quotidiennement un nombre de patients qui est modélisé par une variable de Poisson de paramètre $\lambda = 10$, et l'on suppose que ces variables de Poisson sont indépendantes. Soit S le nombre total de patients qui viennent pour une consultation à l'hôpital un jour donné, puis $M = S/100$ le nombre moyen de consultations par salle ce jour-là.

QCM11 : Quelles sont les propositions correctes parmi celles ci-dessous ?

A $\mathbb{E}(M) = 0,1$

B $\mathbb{E}(M) = 10$

C $\mathbb{E}(S) = 100$

D $\mathbb{E}(S) = 1000$

E $\mathrm{Var}(S) = 10000$

QCM12 : Quelle est la probabilité de voir S dépasser la valeur 1050 un jour donné (à 10^{-2} près) ?

A $0,06$

B $0,11$

C $0,26$

D $0,48$

E $0,96$

XVIII Estimation d'un paramètre poissonien :

On suppose que le nombre d'accidents survenant dans une certaine ville un jour donné suit une loi de Poisson de paramètre λ inconnu. Pendant un an, on a relevé quotidiennement le nombre X d'accidents survenus dans cette ville durant la journée, le tableau ci-dessous résume les résultats de cette enquête :

Nombre d'accidents dans la journée

Val. observ.	*0*	*1*	2	3	*Total*
Nombre de j.	*200*	*100*	55	*10*	*365*

On part du principe que les variables $X_1, X_2, \ldots, X_{365}$ associées à chaque journée sont indépendantes, et l'on note M la variable aléatoire de moyenne annuelle des nombres d'accidents quotidiens.

QCM13 : Peut-on affirmer que

A M suit exactement une loi de Poisson ?

B M suit approximativement une loi binomiale ?

C M suit approximativement une loi normale ?

D $\mathbb{E}(M) = \lambda$?

E $\mathrm{Var}(M) = \frac{\lambda}{365}$?

QCM14 : On note s^2 la variance empirique obtenue pour ces variables de Poisson $X_1, X_2, \ldots, X_{365}$ et $\hat{x}$ leur moyenne empirique. Peut-on affirmer que

A $\hat{x} = 0,66$?

B $\hat{x} = 91,3$?

C $s^2 = 0,69$?

D $s^2 = 6,85$?

E $s^2 = 68,5$?

QCM15 : On choisit de noter $I_{1-\alpha}$ un intervalle de confiance de niveau $(1-\alpha)$ pour le paramètre λ. En arrondissant les bornes à 10^{-2} près, peut-on affirmer que

A $I_{0,95} = [0,57; 0,75]$?

B $I_{0,95} = [0,64; 0,68]$?

C $I_{0,95} = [90,5; 92,1]$?

D $I_{0,90} = [0,59; 0,73]$?

E $I_{0,90} = [90,8; 91,8]$?

XIX Mesures et précision :

Un appareil dose la concentration d'une substance sans biais, mais avec une erreur de mesure suivant une loi normale de moyenne nulle et de variance $100\,(mg/L)^2$. Si on effectue plusieurs mesures d'une concentration sur un même prélèvement, on suppose que toutes les erreurs de mesures sont indépendantes. En répétant les mesures puis en choisissant leur moyenne comme résultat, on souhaite obtenir une précision (demi-longueur d'int. de confiance de niveau 95%) inférieure ou égale à $5\,mg/L$.

QCM16 : Combien faut-il effectuer de mesures pour obtenir la précision désirée ?

A 5

B 10

C 12

D 16

E 20

QCM17 : Avec l'appareil de la question précédente, le coût d'une mesure est de 100 Euros. On se demande s'il est rentable d'acheter un nouvel appareil pour doser la concentration de la substance. Un nouvel appareil produit des mesures sans biais et comportant une erreur de variance $40\,(mg/L)^2$ seulement, mais le coût de chaque mesure avec ce nouvel appareil s'élève à 200 Euros. Ici encore, l'objectif est d'obtenir une précision inférieure ou égale à $5\,mg/L$, quitte à combiner de nombreuses mesures. Peut-on affirmer que

A Pour atteindre ce but, il faut effectuer 5 mesures avec le nouvel appareil ?

B Pour atteindre ce but, il faut effectuer 7 mesures avec le nouvel appareil ?

C Pour atteindre ce but, il faut effectuer 9 mesures avec le nouvel appareil ?

D Il vaut mieux garder l'ancien appareil de dosage.

E Il vaut mieux acheter le nouvel appareil de dosage.

XX Encore des sondages :

Une élection oppose deux candidats, et tout porte à croire qu'elle sera très serrée : on part du principe que tous les électeurs vont voter pour l'un ou l'autre candidat, et que les scores de chacun d'eux seront proches de 50%. On souhaite effectuer un sondage à la sortie des bureaux de vote.

QCM18 : Combien d'électeurs faut-il interroger pour atteindre une précision (demi-longueur d'intervalle de confiance au niveau 95%) de $0,02$?

A 996

B 1010

C 1563

D 2401

E 2542

QCM19 : Si l'on choisit d'interroger 1000 électeurs, quelle sera la précision obtenue pour les proportions de voix remportées par chaque candidat ?

A $0,01$

B $0,03$

C $0,05$

D $0,07$

E $0,10$

Sem 1 Proba / Stats TD 1

Exercice F

1) $\frac{450}{1000} = 0,45 \rightarrow \hat{p} = 0,45$

2) $X_1 \dots X_{1000}$ des variables aléatoires de Bernoulli i.i.d de paramètre p.

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad (n = 1000)$$

$$E(\overline{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$
$$= \frac{1}{n}\sum_{i=1}^{n} p \quad \text{car } \sum_{i=1}^{n} p = (\underbrace{p+p \dots +p}_{n \text{ fois}})$$
$$= \frac{\cancel{n}p}{\cancel{n}}$$
$$= p$$

$$Var(\overline{X}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$
$$= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \quad \text{car } X_i \text{ iid}$$
$$= \frac{1}{n^2}\sum_{i=1}^{n} p(1-p) = \frac{p(1-p)}{n}$$

D'après le TCL $\frac{\overline{X}_n - E(X_n)}{\sqrt{Var(X_n)}} \simeq N(0,1)$

$$\frac{Y_n - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

On cherche $\varepsilon > 0$ tq :

$p \in [p - \varepsilon ; p + \varepsilon]$ à 95%

$\varepsilon = Q_{\alpha/2} \times \sqrt{\frac{p(1-p)}{n}}$

$\varepsilon = 1,96 \times \sqrt{\frac{0,45(1-0,45)}{1000}}$ $p \simeq \hat{p}$

$\varepsilon = 0,03$

On cherche $\varepsilon > 0$ tq :

$p \in [p - \varepsilon ; p + \varepsilon]$ à 98%

$\varepsilon = Q_{\alpha/2} \times \sqrt{\frac{p(1-p)}{n}}$

$\alpha = 1 - 0,98 = 0,02$

$P(X < Q_{\alpha/2}) = 1 - \alpha/2$

$= 1 - 0,01$

$= 0,99$

$Q_{\alpha/2} = 2,3$

$\varepsilon = 2,3 \times \sqrt{\frac{p(1-p)}{n}}$

$\varepsilon = 0,037$

3) $p \in [0 ; p + \varepsilon]$ à 95%

Sem 1 Proba / Stat TD 2

$$P(p \in [0, \bar{X}_n + \varepsilon]) = 0,95$$

$$P(0 \leq p \leq \bar{X}_n + \varepsilon) = 0,95$$

$$P(p \leq \bar{X}_n + \varepsilon) = 0,95$$

$$P(p - \bar{X}_n \leq \varepsilon) = 0,95$$

$$P(\bar{X}_n - p \geq -\varepsilon) = 0,95$$

$$P\left(\frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \geq \frac{-\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$P\left(N \leq \frac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$\frac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}} = 1,645$$

$$\varepsilon = 1,645 \times \sqrt{\frac{p(1-p)}{n}}$$

$$= 0,026$$

Exo F 4) $\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_i$

$X_i \sim$ Bernoulli iid

$P(p\,[\hat{p}, \overline{X}_n + 0{,}05])$

C'est proba que la vrai proportion de oui soit compris entre $\hat{p}$ et 0,5

$P(p \leqslant \overline{X_n} + 0{,}05)$

$P(-0{,}05 \leqslant \overline{X_n} - p)$

$P\left(\frac{-0{,}05}{\sqrt{\frac{p(1-p)}{n}}} \leqslant \frac{\overline{X_n} - p}{\sqrt{\frac{p(1-p)}{n}}}\right)$

$P\left(N \geqslant \frac{-0{,}05\sqrt{1000}}{\sqrt{0{,}45 \times 0{,}55}}\right)$

$= P(N \geqslant -3{,}18)$

$= P(N \leqslant 3{,}18)$
$= 0{,}9993$

5) On cherche n tq :
$P[\overline{X}_n - 0{,}01 ; \overline{X_n} + 0{,}01] = 0{,}95$

car cet intervalle est de taille 0,2

$P(p - 0{,}01 \leq \overline{X}_n \leq p + 0{,}01) = 0{,}95$

$P\left(\frac{-0{,}01\sqrt{n}}{\sqrt{0{,}45 \times 0{,}55}} \leq \overline{X}_n \leq \frac{0{,}01\sqrt{n}}{\sqrt{0{,}45 \times 0{,}55}}\right)$

$1{,}96 = \frac{0{,}01\sqrt{n}}{\sqrt{0{,}45 \times 0{,}55}}$

$n = \left(\frac{1{,}96}{0{,}01}\right)^2 \times 0{,}45 \times 0{,}55$

$\simeq 9500$

Exo C) Soit $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ avec X_i v.a. de Bernoulli i.i.d de param p inconnu

$E(\overline{X}) = p$

$Var(\overline{X}) = \frac{p(1-p)}{n}$

D'après le TCL :

$N = \frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$

on cherche ε tq :

$P(p \in [\overline{X}_n - \varepsilon, \overline{X}_n + \varepsilon]) = 0{,}95$

$\Rightarrow P\left(\frac{-\varepsilon\sqrt{1000}}{\sqrt{p(1-p)}} \leq N \leq \frac{\varepsilon\sqrt{1000}}{\sqrt{p(1-p)}}\right) = 0{,}95$

$$1{,}96 = \frac{\varepsilon\sqrt{1000}}{\sqrt{p(1-p)}}$$

$$\varepsilon = \frac{1{,}96 \times \sqrt{p(1-p)}}{\sqrt{1000}}$$

$$\varepsilon \simeq \frac{1{,}96\sqrt{0{,}38(1-0{,}38)}}{\sqrt{1000}}$$

$$\varepsilon \simeq 0{,}03$$

$$I_f = 0{,}38 \pm 0{,}3$$
$$= [0{,}35 \,,\, 0{,}41]$$

de même :

$$\varepsilon \simeq \frac{1{,}96\sqrt{0{,}36(1-0{,}36)}}{\sqrt{1000}}$$

$$I_f = 0{,}36 \pm 0{,}3$$
$$= [0{,}33 \,,\, 0{,}39]$$

On peut conclure que la variation peut découler des ~~fluct~~ fluctuations naturelles du sondages (les intervalles se chevauchent) donc c'est pas suffisant pour dire que l'on a perte d'audience Tphs

Exercice E

1) Soit $X_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ avec X_i v.a. de Bernoulli i.i.d et $n = 100$

$E(X) = p$
$V(X) = \frac{p(1-p)}{n}$

d'après le TCL

$$\frac{X_n - p}{p(1-p)} \sim N(0,1)$$

On cherche ε tq :

$$P(p-\varepsilon \leqslant \overline{X}_n \leqslant p+\varepsilon) = 0,99$$

$$P\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \leqslant \frac{\overline{X}_n - p}{\sqrt{p(1-p)}} \leqslant \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) = 0,99$$

$$2P(N \leqslant \varepsilon) - 1 = 0,99$$

$$2P(N \leqslant \varepsilon) = 1,99$$

$$P(N \leqslant \varepsilon) = 0,995$$

$$2,58 = \frac{\varepsilon\sqrt{100}}{\sqrt{p(1-p)}} \Leftrightarrow \varepsilon = \frac{2,58\sqrt{p(1-p)}}{\sqrt{100}}$$

$$\varepsilon \simeq 0,128$$

$$I = [0,422 , 0,678]$$

Sem 3 Proba/Stat TD 1

Exercice VIII p139 (conformité)

H_0 : "l'observation est la réalisation suivant la loi $\left(\frac{9}{16}; \frac{3}{16}; \frac{3}{16}; \frac{1}{16}\right)$"

H_1 : "l'observation ne suit pas la loi"

$$\underbrace{P(H_0 \mid Obs)}_{\text{postériori}} = \frac{\overbrace{P(Obs \mid H_0)}^{\text{vraisemblance}} \times \underbrace{P(H_0)}_{\text{à priori}}}{P(Obs)}$$

Dans ce cours, on ne prend en compte que la vraisemblance : proba des obs sachant que Ho est supposé vrai

On suppose Ho, on calc la proba des obs sachant Ho. Si $P(Obs \mid H_0)$ est plus faible le seuil p, on rejette Ho au risque p.

On suppose Ho (tableau) avec $p = 1\%$
On calcule la distance du χ^2 entre les effectifs observés et théorique.

$$D_{\chi^2} = \sum_{obs} \frac{(obs - théo \times tot)^2}{théo \times tot} = \frac{\left(1790 - \frac{9}{16} \times 3098\right)^2}{\frac{9}{16} \times 3098}$$

$$+ \frac{\left(547 - \frac{3}{16} \times 3098\right)^2}{\frac{3}{16} \times 3098} + \frac{\left(548 - \frac{3}{16} \times 3098\right)^2}{\frac{3}{16} \times 3098}$$

$$+ \frac{\left(213 - 3098/16\right)^2}{3098/16}$$

On suppose H_0, si H_0 est vrai alors D est la réalisation d'une va du K^2

test de conformité

Dans D est la réalisation d'une v.a du K degré de liberté

On D = 7,06

On calcule $P(\chi^2 \geq 1) = 0,05$

$P(\chi^2 \geq 1) = 1 - P(\chi^2 \leq S)$

$P(\chi^2 \leq S) = 0,95$

La distance dobs $\leq S$ on ne peut pas rejeter H_0 au risque 5%

Nous venons de faire un test de conformité d'observation à une loi de probabilité

<u>Ex I p137</u> (Indépendance)

1) H_0 : "L'observation est représentative de la population nationale"

H_1 : "L'observation n'est pas représentative de la population nationale"

On suppose H_0 avec s = 5%

On calcule la distance du χ^2 entre effectifs observé et théoriques

Bac	Général	Techno	Pro	Total
Obs	131	37	62	250
Théo	0,47x250 = 117,5	0,2 250 = 50	0,33x250 = 82,5	250

Sem 3 Proba/Stat TD 2

$$D = \frac{(131-147,5)^2}{147,5} + \frac{(37-50)^2}{50} + \frac{(82-82,5)^2}{82,5}$$

degré de liberté : nb de colonnes − 1

$$= 1,55 + 3,38 + 0,003$$
$$= 4,963$$

$$P(X^2 \geq s_1) = 0,05$$
$$s_1 = 5,991$$

Donc $D \leq s_1$
On ne rejette pas H_0 au risque 5%

$$P(X^2 \geq s_1 = 0,10)$$
$$s_2 = 4,605$$

Donc on rejette H_0 au risque 10%

2) a)

H_0 "X et Y indépendants"

X : "Admis ou pas"
X : "sexe"

Tableau obs

	Admis	Pas Admis	Total
Garçon	132	18	150
Filles	173	13	186
	305	31	336

$X : \Omega \longrightarrow \{garçon ; fille\}$
$X : \Omega \longrightarrow \{Admis, pas admis\}$

On suppose H_0 vraie

$P(X = \text{garçon et } Y = \text{admis})$

$= P(X = \text{garçon}) \times P(Y = \text{admis})$

$= \frac{150}{336} \times \frac{305}{336}$

Tableau Théo

	Admis	Non admis	Total
Garçon	136,16	13,84	150
Filles	168,86	17,16	186
Totale	305	31	336

b) On calcule la différence de χ^2 D :

$$D = \frac{(132 - 136,16 \times 336)^2}{136,16} + \frac{(18 - 13,84)^2}{13,84} + \frac{(173 - 168,84)^2}{168,84} + \frac{(13 - 17,13)^2}{17,16}$$

$= 2,41$

Sem 3

Proba / Stat

TD 3

Si H_0 vrai alors D est la réalisation d'une v.a du χ^2 à $(2-1)\times(2-1) = 1$

(nb Categ V.a 1 - 1)
(nb Categ V.a 2 - 1)

$$P(\chi^2 \geq s) = 0,05$$

$$s = 3,841$$

On admet H_1 car $D < S$ au risque 5%

c) On trouve qqch d'inférieur

(plus la D du χ^2 est grande plus la proba que H_1 soit admise est grande)

Donc le bac général a plus de chance d'être conditionné par le sexe que le bac techno

$$D_{tech} < D_{général}$$

→ proba que H_1 soit vrai est plus grande pour le bac général que le bac techno

Exercice 2

1)

$$\begin{pmatrix} 0 & 0 & \frac{3}{5} & 0 & \frac{2}{5} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{9}{10} & 0 & 0 & 0 & \frac{1}{10} \\ 0 & 0 & 0 & \frac{1}{10} & \frac{9}{10} \end{pmatrix}$$

2) On appelle mesure invariante π (vecteur ligne) tq

$\pi = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$

si $\pi \times P = \pi$
et $\pi_1 + \dots + \pi_5 = 1$

Si $X_0 \sim \mu = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$

Loi de départ de la chaîne de Markov

alors la loi$(X_n) = \mu \times P^n$
Donc si $X_0 \sim \pi$, la loi de $X_n = \pi$

$[\pi_1, \pi_2, \pi_3, \pi_4, \pi_5] P = [\pi_1, \pi_2, \pi_3, \pi_4, \pi_5]$

$\frac{1}{4}\pi_2 + \frac{9}{10}\pi_4 = \pi_1$

$\pi_2 = 0$

$\frac{3}{5}\pi_1 + \frac{1}{2}\pi_2 = \pi_3$

$\frac{2}{5}\pi_1 + \frac{1}{2}\pi_3 + \frac{1}{10}\pi_4 + \frac{2}{10}\pi_5 = \pi_5$

$\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 = 1$

$$\begin{cases} \pi_1 = \frac{9}{10}\pi_4 \\ \pi_2 = 0 \\ \pi_3 = \frac{3}{5}\pi_1 \\ \pi_4 = \frac{3}{10}\pi_1 + \frac{1}{10}\pi_5 \\ \pi_5 = \frac{2}{5}\pi_1 + \frac{3\pi_1}{10} + \frac{1}{10}\left(\frac{3}{10}\pi_1 + \frac{1}{10}\pi_5\right) \end{cases}$$

$$\begin{cases} \pi_4 = \frac{10}{9}\pi_1 \\ \pi_2 = 0 \\ \pi_3 = \frac{3}{5}\pi_1 \end{cases}$$

$\pi_1 = \frac{45}{487}$ $\pi_2 = 0$ $\pi_3 = \frac{27}{487}$

$\pi_4 = \frac{50}{487}$ $\pi_5 = \frac{365}{487}$

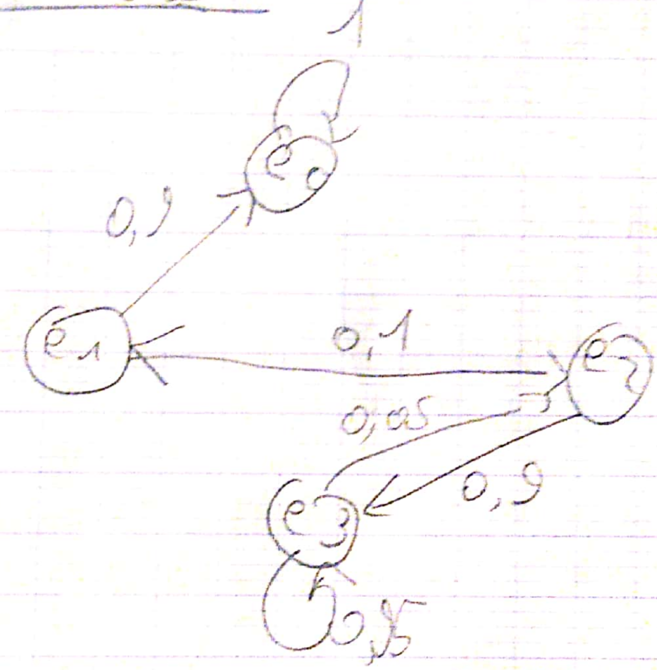
car il y a unique solution (pivot de Gauss)

Jeudi Proba / Stat TD 2

$\pi = \left(\frac{45}{487}, 0, \frac{77}{487}, \frac{50}{487}, \frac{365}{487}\right)$

Si la loi de X_0 est π alors loi$(X_n) = \pi \ \forall n$
car Loi$(X_n) = \pi \times P^n = \pi$

Exercice 3



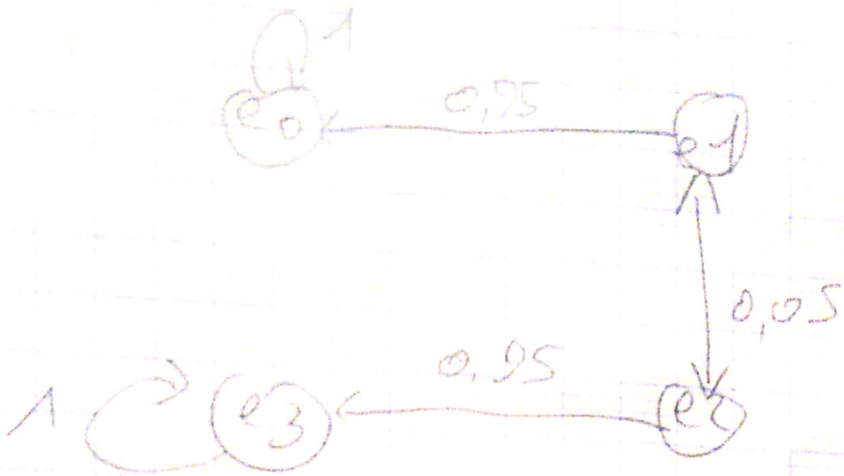
$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0{,}9 & 0 & 0{,}1 & 0 \\ 0 & 0{,}1 & 0 & 0{,}9 \\ 0 & 0 & 0{,}05 & 0{,}95 \end{pmatrix}$$

$\underbrace{(\pi_0, \pi_1, \pi_2, \pi_3)P}_{\pi} = (\pi_0, \pi_1, \pi_2, \pi_3)$

$$\Leftrightarrow \begin{cases} \pi_0 = \pi_0 + 0{,}9\pi_1 \\ \pi_1 = 0{,}1\pi_2 \\ \pi_2 = 0{,}1\pi_1 + 0{,}05\pi_3 \\ \pi_3 = 0{,}9\pi_2 + 0{,}95\pi_3 \end{cases} \Leftrightarrow \begin{cases} \pi_1 = 0 \\ \pi_2 = 0 = \pi_3 \\ \pi_0 = 1 \end{cases}$$

$(0,0,0,1)$ ← Unique mesure invariante

8)



$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0{,}95 & 0 & 0{,}05 & 0 \\ 0 & 0{,}05 & 0 & 0{,}95 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\begin{cases} \Pi_0 = \Pi_0 + 0{,}95\,\Pi_1 \\ \Pi_1 = 0{,}05\,\Pi_2 \\ \Pi_2 = 0{,}05\,\Pi_1 \\ \Pi_3 = 0{,}95\,\Pi_2 + \Pi_3 \end{cases}$$

$$\Rightarrow \begin{cases} \Pi_0 = \Pi_0 \\ \Pi_1 = 0{,}05 \times 0{,}05\,\Pi_1 = 0 \\ \Pi_2 = 0 \\ \Pi_3 = \Pi_3 \end{cases}$$

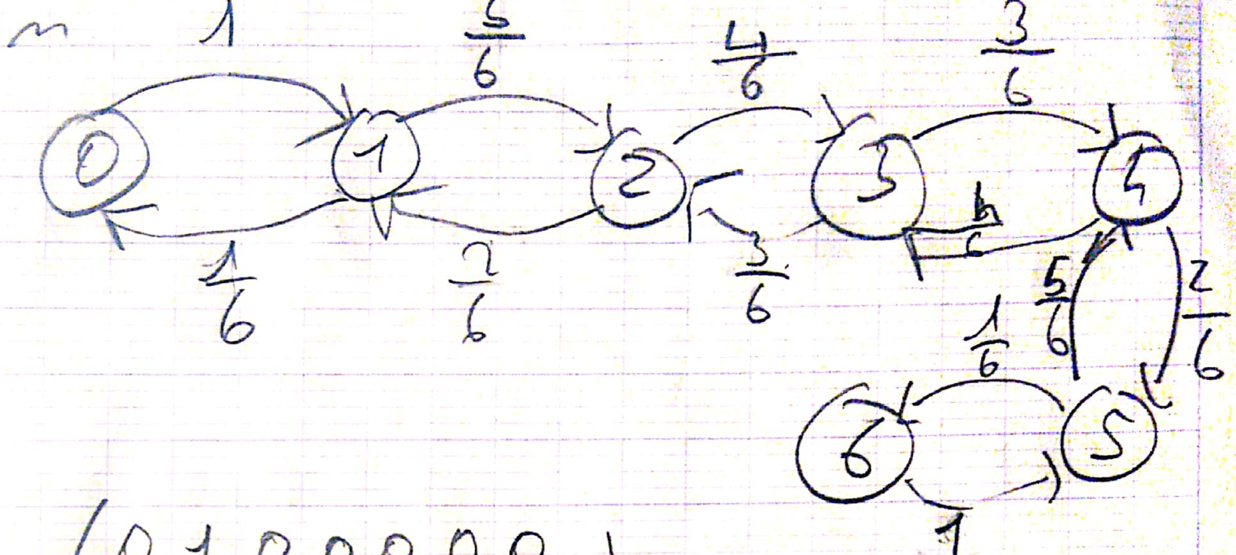
$$\Pi_0 + \Pi_1 + \Pi_2 + \Pi_3 = 1$$

$$\Leftrightarrow \begin{cases} \Pi_1 = \Pi_2 = 0 \\ \Pi_0 + \Pi_3 = 1 \end{cases}$$ ⇒ l'ensemble des mesures invariantes est de la forme $\Pi = (\alpha, 0, 0, 1-\alpha)$ avec $\alpha \in [0,1]$

9

Exercice 1 (Polya chaîne Markov)

A 1. X_n la va comptant le nb de boules dans l'urne A à l'instant n



$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & \frac{5}{6} & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{6} & 0 & \frac{4}{6} & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{6} & 0 & \frac{3}{6} & 0 & 0 \\ 0 & 0 & 0 & \frac{4}{6} & 0 & \frac{2}{6} & 0 \\ 0 & 0 & 0 & 0 & \frac{5}{6} & 0 & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Par définition

$$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

2. X chaîne de Markov réversible si il existe un vecteur de proba $\pi = (\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6)$

$$\left(\sum_{i=0}^{6} \pi_i = 1\right) \text{ tq } \pi_i . p_{ij} = \pi_j . p_{ji}$$

On dit que $(X_n)_{n \in \mathbb{N}}$ est réversible car si $X_0 \sim \pi$
alors $Loi(X_0, X_1, \dots, X_n) = Loi(X_n, \dots, X_0)$

<u>Remarque</u> : une mesure réversible est une mesure invariante (la réciproque est fausse)

<u>en effet</u> :

démo

$$\pi \times P = \sum_{i=0}^{6} \pi_i \, p_{ij}$$
$$= \sum_{i=0}^{6} \pi_j \, p_{ji} \quad (\text{car } \pi \text{ réversible})$$
$$= \pi_j \underbrace{\sum_{i=0}^{6} P_{ji}}_{=1}$$
$$= \pi_j$$

On a bien $\pi P = \pi$ donc π est une mesure invariante.

<u>À Retenir</u> : c'est plus simple de calculer une mesure réversible qu'une mesure invariante
Si la chaine n'est pas réversible, il peut quand même exister une mesure invariante (il en existe tjrs une dans le cas d'un espace d'état fini).

<u>Exo suite</u> : On cherche π une mesure réversible :

$$\sum_{j=0}^{6} \pi_0 \, p_{0j} = \sum_{j=0}^{6} \pi_j \, P_{j0}$$

Sem 6 Proba / Stats TD 2

$\Leftrightarrow \pi_0 = \pi_1 \, p_{10}$

$$\begin{cases} \pi_0 = \dfrac{\pi_1}{6} \\ \dfrac{5}{6}\pi_1 = \dfrac{1}{3}\pi_2 \\ \dfrac{2}{3}\pi_2 = \dfrac{1}{2}\pi_3 \\ \dfrac{1}{2}\pi_3 = \dfrac{2}{3}\pi_4 \\ \dfrac{1}{3}\pi_4 = \dfrac{5}{6}\pi_5 \\ \dfrac{1}{6}\pi_5 = \pi_6 \end{cases}$$

$$\begin{cases} \pi_1 = 6\pi_0 \\ \pi_2 = 15\pi_0 \\ \pi_3 = 20\pi_0 \\ \pi_4 = 15\pi_0 \\ \pi_5 = 6\pi_0 \\ \pi_6 = \pi_0 \end{cases}$$

Comme $\sum_{i=0}^{6} \pi_i = 1$

$\Rightarrow \pi_0 (1 + 6 + 15 + 20 + 15 + 6 + 1) = 1$

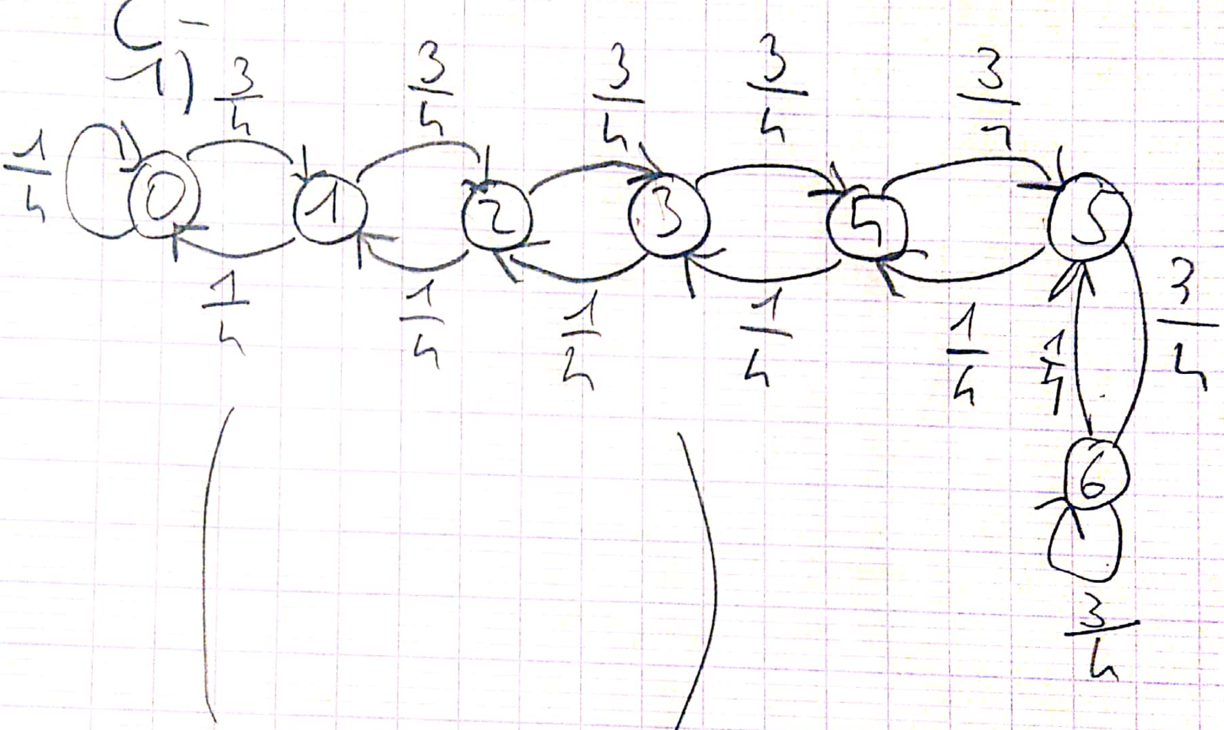
$\pi_0 = \dfrac{1}{2^6}$

3) C'est fortement connexe (le graphe)
→ irréductible

Mais la chaîne n'est pas apériodique
elle est de période 2.

B - sauté

C -
1)



2) sauté

3) Le graphe est fortement connexe
donc la chaîne est irréductible

Il y a une boucle sur un état
donc elle est apériodique (période
1).
D'après le THM de Döblin car Y
loi $(Y_n) \xrightarrow[n \to +\infty]{} \Pi$
est irréductible et apériodique.
Elle converge exponentiellement v.6.

Sem 6 Proba/Stat

Exercice 6 (poly Markov)

1)
$$\begin{pmatrix} 0 & 0 & \frac{1}{5} & 0 & \frac{1}{5} & \frac{3}{5} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{5} & \frac{3}{5} & 0 & 0 \\ 0 & 0 & \frac{2}{5} & \frac{1}{2} & \frac{1}{10} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{10} & \frac{9}{10} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

2) $T = \{v_1, v_2\}$

$C_1 = \{v_3, v_4, v_5\}$

$C_2 = \{v_6, v_7\}$

3) $\text{Loi}(X_1 \mid X_0 = v_1) = (0, 0, \frac{1}{5}, 0, \frac{1}{5}, \frac{3}{5}, 0$

$\text{Loi}(X_2 \mid X_0 = v_1) = (0, 0, \frac{1}{5}, 0, \frac{1}{5}, \frac{3}{5}, 0) \times$

$= (0, 0, 0, \frac{12}{50} \ldots$

à finir

4. Elle n'est pas irréductible car elle a des états transients
Elle a une infinité de mesures invariantes qui sont les combi convexe

Sem 7 Proba 1

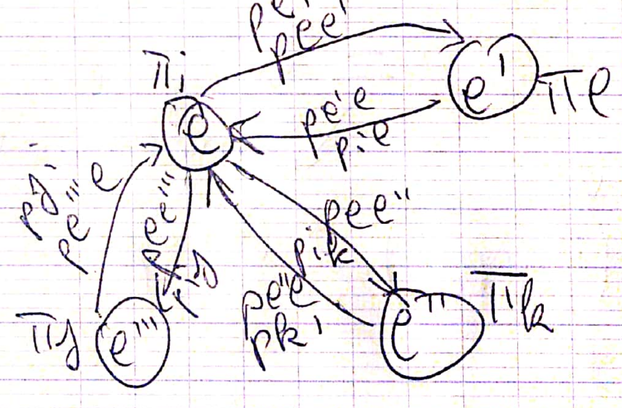
Mesures invariantes réversible

Def : Mesure "mieux" qu'invariante

Plusieurs Mesures invariantes : plusieurs classes récurentes. (une infinité).

Autant d'extremales que de classes récurentes.

Mesure ~~réversible~~ invariante définies à partir des ~~mesures invariantes réversible~~
Global Balance solution.



Reçu par e en 1 pas

$$\sum_{j\in E} \pi_j \times p_{ji}$$

Ssi π mesure invariant $\Leftrightarrow$ = Donné par e en 1 pas

$$\sum_{j\in E} \pi_j \times p_{ij} = \pi_i \underbrace{\left(\sum_{j\in E} p_{i,j}\right)}_{=1} = \pi_i$$

$\Rightarrow \sum_{j\in E} \pi_j \times p_{ji} = \pi_i \qquad \forall i \in E$

$\hookrightarrow \pi \times P = \pi$

ou $\pi_1 \times (p_{1,\cdot}) + \pi_2 \times (p_{2,\cdot}) + \dots + \pi_m \times (p_{m,\cdot}) = \pi$

Mesure Réversible :
(Detailled balance equation)
Quand on trouve une mesure réversible, c'est forcément une mesure réversible.

Reçu par e_i de e_j	Reçu par e_j de e_i
$\pi_j \times p_{ji}$	$\pi_i \times p_{ij}$

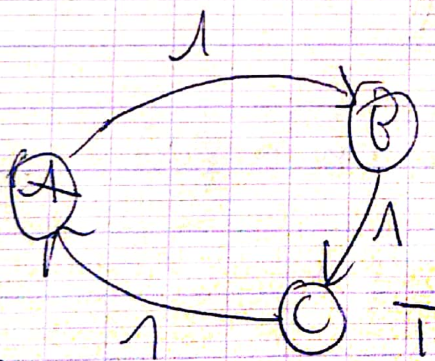
$= \forall\, i,j \in E$

$\hookrightarrow$ mesure réversible

Démo de revers $\rightarrow$ inv

(DBE) $\sum_{j\in E} (\pi_j \times p_{ji}) = \sum_{j\in E} (\pi_i \times p$

Exemple:



$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\pi^{(inv)} = \left(\frac{1}{3} ; \frac{1}{3} ; \frac{1}{3}\right)$$

$\pi^{(inv)}$ n'est pas réversible car quand A donne $\frac{1}{3}$ à B, B lui donne 0.