TD 1

1. Génotype

Le groupe sanguin d'un individu se caractérise par son génotype, réunion de 2 gènes transmis par les parents. Il existe 4 génotyps : AB, AO, BO, OO. Sachant que chacun des parents peut transmettre avec une même probabilité chacun des gènes qu'il possède, quelles seront les lois de probabilité des variables aléatoires X_1, X_2, X_3 et X_4 égales aux nombres d'enfants de chacun des quatre groupes dans une famille de cinq enfants dont les parents sont de génotype AO et BO ? Quelle est la probabilité de n'avoir dans une telle famille aucun enfant de même génotype que ses parents ?

2. Couple

En supposant l'équiprobabilité des sexes à la naissance, quel nombre minimum d'enfants un couple de parents doit-il prévoir pour s'assurer 90% de chances d'avoir au moins un garçon et une fille ?

3. Une urne

Première expérience

Une urne contient 3 boules numérotées 1, 2, 3. On tire successivement et sans remise deux boules et on associe à cette expérience le couple (X, Y) où X est le numéro de la première boule tirée de l'urne et Y est le numéro de la deuxième boule tirée de l'urne.

1) Donner la loi du couple (X, Y), les lois marginales, conditionnelles, les espérances mathématiques.

2) Quelle est la matrice de variance-covariance de ce couple ?

Deuxième expérience

Maintenant on ajoute une boule numérotée 1 dans l'urne. Le reste de l'expérience est le même que la première expérience.

3) Donner la loi du couple et les lois marginales de X et Y. Les variables aléatoires X et Y sont elles indépendantes ?

4) Calculer les espérances et les variances de X et Y.

5) Calculer la covariance de X et Y.

On propose un jeu à partir du tirage. Le joueur gagne 5 euros $\times$ la somme des deux nombres obtenus. C'est à dire que le gain de jeu est $5 \times (X + Y)$ euros.

6) Notons Z la somme des deux nombres obtenus. Donner la loi de Z.

7) En déduire la loi du gain de jeu et calculer son espérance et sa variance.

4. Examen

Le programme d'une épreuve d'examen comporte 40 sujets. Trois d'entre eux sont tirés au sort et proposés au choix à chaque candidat. Quelle est la probabilité pour un candidat n'ayant étudié que le quart des sujets :

– d'avoir préparé les trois sujets proposés ?
– d'avoir préparé deux des sujets proposés ?
– de n'en avoir préparé aucun ?
– d'en avoir préparé au moins un ?

5. Couple de variables aléatoires discrètes

Supposons que $X_1, X_2, \dots$ sont des variables aléatoires i.i.d. qui ne prennent que deux valeurs 0 et 1 avec $\mathbb{P}(X_1 = 0) = p, 0 < p < 1$. Soit N une variable aléatoire discrète dans l'ensemble $\{0, 1, 2\}$ avec $\mathbb{P}(N = 0) = \mathbb{P}(N = 1) = \mathbb{P}(N = 2)$. La variable N et la suite $(X_i)_{i\in\mathbb{N}}$ sont indépendantes. On considère la variable aléatoire S définie par

$$S = \sum_{i=1}^{N} X_i$$

où $S = 0$ si $N = 0$.

1) Donner la loi jointe du couple (N, S) et la loi marginale de S.

2) Déterminer les lois conditionnelles $\mathbb{P}(S \mid N = 1)$ et $\mathbb{P}(S \mid N = 2)$. Donner les noms et les paramètres des lois obtenues.

3) Calculer les espérances $\mathbb{E}(S \mid N = 0)$, $\mathbb{E}(S \mid N = 1)$ et $\mathbb{E}(S \mid N = 2)$. En déduire l'espérance $\mathbb{E}(S)$.

4) Supposons que N suit la loi uniforme sur l'ensemble $\{1, \dots, n\}$. Qu'en déduisez vous pour $\mathbb{E}(S)$?

5) Supposons que N suit la loi $Pois(\lambda)$. Calculer l'espérance $\mathbb{E}(S)$.

6. Couple d'entiers

Soit (X, Y) un couple de variables aléatoires dans $\mathbb{N}$. On pose

$$p_{kn} = \mathbb{P}(X = k, Y = n) = \frac{\lambda^k e^{-\lambda} \alpha^n (1-\alpha)^{k-n}}{n!(k-n)!} \mathbb{1}_{\{0\le n\le k\}}(k, n)$$

où $\lambda > 0$ et $0 \le \alpha \le 1$ sont des nombres réels donnés.

1) Quelle est la loi de la variable aléatoire X ?

2) Quelle est la loi de la variable aléatoire Y ? X et Y sont-elles indépendantes ?

3) Quelle est la loi de la variable aléatoire $Z = X - Y$?

4) Pour tout couple $(m, n) \in \mathbb{N}^2$, calculer $\mathbb{P}(Y = m \mid Z = n)$, qu'en déduire sur Y et Z ?

5) On fait l'hypothèse que le nombre d'enfants d'une famille, tirée au hasard dans une population donnée est une variable aléatoire de Poisson d'espérance 2, 2. On admet qu'à chaque naissance, la probabilité d'observer un garçon est égale à 1/2.

a. Montrer que, conditionnellement au fait que la famille a n enfants, le nombre de garçons suit une loi Binomiale $Bin(n, 1/2)$.

b. Quelle est la probabilité pour qu'une famille ainsi choisie contienne a enfants dont b garçons ?

c. Que peut-on dire des variables aléatoires représentant le nombre de garçons et le nombre de filles d'une famille choisie au hasard dans la population donnée ?

TD 2

1. Bernoulli

On considère une suite de variables de Bernoulli, indépendantes et identiquement distribuées $(X_1, \ldots, X_n, \ldots)$ avec $X_i \sim Ber(1/2)$.

1) Modéliser cette suite à l'aide de chaîne de Markov (espace d'état, matrice de transition).

2) Quelle est la mesure invariante de cette chaîne de Markov ?

2. Le Hamster

Un hamster paresseux ne connaît que 3 endroits dans sa cage :

– Les copeaux où il dort (état $(1, 0, 0)$)

– La mangeoire où il mange (état $(0, 1, 0)$)

– La roue où il fait de l'exercice (état $(0, 0, 1)$).

Ses journée sont semblables les unes aux autres et on représente son activité par une chaîne de Markov. Toutes les minutes, il peut soit changer d'activité, soit continuer celle qu'il était en train de faire :

– Quand il dort, il a 9 chances sur 10 de ne pas se réveiller la minute suivante.

– Quand il se réveille, il y a 1 chance sur 2 qu'il aille manger et 1 chance sur 2 qu'il parte faire de l'exercice.

– Le repas ne dure qu'une minute, après il fait autre chose.

– Après avoir mangé, il y a 3 chances sur 10 qu'il parte courir dans sa roue, mais surtout 7 chances sur 10 qu'il retourne dormir.

– Courir est fatigant, il a donc 8 chances sur 10 de retourner dormir au bout d'une minute, sinon (avec 2 chances sur 10) il continue en oubliant qu'il est déjà un peu fatigué.

1) Rappeler la matrice de transition de la chaîne de Markov $(X_t)_{t\in\mathbb{N}}$, qui représente l'activité du hamster.

2) Caractériser les 3 états (récurrents ou transitoires).

3) Trouver la loi de probabilité invariante pour $(X_t)_{t\in\mathbb{N}}$.

3. Chaîne périodique

On considère la chaîne de Markov $(X_t)_{t\in\mathbb{N}}$, avec $X_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ et la matrice de transition $M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

1) Calculer la trajectoire des 5 premiers états de la chaîne.

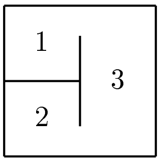
2) Quels sont les états récurrents et les états transitoires de cette chaîne ?

3) Trouver une mesure invariante μ pour cette chaîne.

4) A-t-on $M^n X_0 \to \mu$?

4. Une souris

Une souris évolue dans un appartement de plan succinct



en choisissant au bout de chaque minute de rester dans la pièce où elle était ou une seule ouverture pour se rendre dans une autre pièce.

i) Si elle est dans la pièce n°1 (état e_1), il y a une chance sur trois qu'elle reste dans cette pièce pour la minute suivante.

ii) Si elle est dans la pièces n°2 (état e_2) ou n°3 (état e_3), elle n'y reste pas pour la minute suivante.

iii) Si elle est dans la pièce n°3, il y a une chance sur deux qu'elle se rend dans la pièce n°1 pour la minute suivante.

Cette évolution peut être représentée par une chaîne de Markov notée $X_1, \ldots, X_n, \ldots$

1) Donner la matrice de transition. Tous les états sont ils récurrents ?

2) Calculer la mesure invariante.

3) Montrer par deux méthodes qu'il existe un chemin de longueur 2 qui va de e_1 à e_1. Calculer $\mathbb{P}(X_3 = e_3 \mid X_1 = e_3)$.

4) Supposons que la souris est dans la pièce n°1 pendant la première minute. Dans quelle pièce se trouvera t elle le plus probablement à la 180ème minute (soit au bout de 3 heures) ?

5. Chaîne de Markov

On considère la chaîne de Markov à deux états ($X_t \in E = \{(1,0)^T, (0,1)^T\}$), avec $\mathbb{P}(X_0 = (0,1)^T) = 1$ et la matrice de transition $M = \begin{pmatrix} 0,9 & 0,2 \\ 0,1 & 0,8 \end{pmatrix}$.

1) Calculer les probabilités des suites $(X_0, \ldots, X_3) = ((0,1)^T, (0,1)^T, (1,0)^T, (1,0)^T)$, $(X_2, X_3) = ((1,0)^T, (1,0)^T)$ et $X_3 = (1,0)^T$.

2) Calculer la probabilité de $X_2 = (1,0)^T$ sachant $X_3 = (1,0)^T$.

3) Trouver la loi de probabilité invariante pour $(X_t)_{t\in\mathbb{N}}$.

6. Acides aminés

On considère la séquence ADN suivante :

$AGGCGTGAAGGTTCCCCAAAATTTTTGGGCCCAAAAACCCTTGGGGGGGCCCCC$

ici $X_t \in E = \{A, G, C, T\}$. On supposera que X_t est une chaîne de Markov.

1) A partir de cette séquence construire la matrice de transition de X_t la plus vraisemblable possible.

2) Présenter deux méthodes pour calculer ou estimer la mesure invariante.

TD 3

1. Variables à densité

Soit (X, Y) un couple de variables aléatoires dont la densité est donnée par

$$f(x,y) = \frac{1}{a\sqrt{xy}} \mathbb{1}_{(0,1]\times(0,1]}(x,y) \quad \text{avec} \quad a > 0.$$

1) Calculer a.

2) Donner les lois marginales et les espérances mathématiques de X et Y, la matrice de variance-covariance de ce couple, ainsi que le coefficient de corrélation entre X et Y.

3) Les variables aléatoires X et Y sont elles indépendantes ? Donner la densité conditionnelle de X sachant $Y = y$.

2. Loi normale

Soit (X, Y) un vecteur aléatoire suivant une loi normale dont la densité est donnée par

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)}$$

avec $(\mu_1, \mu_2, \sigma_1, \sigma_2) \in \mathbb{R}^2 \times (\mathbb{R}^+)^2$.

1) Donner les lois marginales et les espérances mathématiques de X et Y, la matrice de variance-covariance de ce couple, les densités conditionnelles de X sachant $Y = y$ et de Y sachant $X = x$, ainsi que le coefficient de corrélation entre X et Y.

2) Calculer la loi de $Z = X + Y$.

3) Donner une condition nécessaire et suffisante pour que les variables aléatoires X et Y soient indépendantes.

3. Variables exponentielles

Soient X et Y deux variables aléatoires indépendantes de lois exponentielles de paramètres respectifs α et β.

1) Donner la densité de la variable $Z = X + Y$ lorsque $\alpha = \beta$ puis lorsque $\alpha \neq \beta$.

2) Donner la densité de la variable $U = X - Y$ lorsque $\alpha = \beta$.

4. Intervalles de confiance

On procède à une série de $n = 9$ mesures avec un même appareil. On suppose que le résultat d'une mesure est une variable aléatoire X qui suit une loi normale de moyenne m et de variance $\sigma^2 = 0.25$.

1) Construire un intervalle de confiance I_1 à 90% pour m.

2) Construire un intervalle de confiance I_2 à 98% pour m.

3) On a observé la série suivante : $(15.2, 15.8, 14.7, 15.3, 15.2, 15.1, 15.6, 14.9, 14.9)$. Donner les intervalles estimés de I_1 et I_2 et les comparer.

4) Comment pourrait-on obtenir un intervalle de confiance ayant même niveau de confiance que I_2 et même longueur que I_1 ?

5. Proportion

On veut estimer, dans un département donné, la proportion p des ménages qui possèdent une machine à laver à partir d'un échantillon de $n = 200$ ménages de ce département. Pour cela on introduit une variable aléatoire X_i qui vaut 1 si le ménage i a une machine à laver et 0 sinon. On suppose que l'échantillon $(X_1, \ldots, X_{200})$ est i.i.d..

1) Montrer que l'estimateur $\bar{X}$ est un estimateur convergent de p.

2) On a trouvé $\bar{X} = 0.64$. Donner un intervalle de confiance à 95% pour p.

a. En utilisant une approximation de la loi de $\bar{X}$ et en remplaçant $p(1-p)$ par $\bar{X}(1-\bar{X})$.

b. En utilisant une approximation de la loi de $\bar{X}$ et en majorant le produit $p(1-p)$.

6. Plus difficile

Soit $f(x, y) = C\mathbb{1}_{\{0\leq x, |y|\leq x\}}(x, y)e^{-\lambda x}$, avec $\lambda > 0$.

1) A quelle condition f est elle la densité d'un couple de variables aléatoires réelles $(X, Y) \in \mathbb{R}^2$?

2) Déterminer les lois marginales (support et densité).

TD 4

1. Estimation paramétrique

Soit (X_1, X_2) un échantillon i.i.d. de taille 2 de la variable aléatoire X dont la densité est donnée par

$$f_\theta(x) = \frac{2x}{\theta^2} \mathbb{1}_{[0,\theta]}(x)$$

avec un paramètre $\theta > 0$.

1) La fonction f_θ est elle vraiment une densité pour toute valeur du paramètre θ ?

2) Calculer la fonction de répartition de X. En déduire celle de $\max(X_1, X_2)$, puis la densité g_θ de $\max(X_1, X_2)$.

3) Montrer que les estimateurs de θ suivants sont sans biais,

$$\hat{\theta}_1 = \frac{3}{4}(X_1 + X_2), \quad \hat{\theta}_2 = \frac{5}{4}\max(X_1, X_2).$$

4) Calculer le risque quadratique moyen de $\hat{\theta}_1$ et $\hat{\theta}_2$.

5) Quel estimateur choisir ?

2. Loi normale

Soit $(X_1, \ldots, X_n)$ un échantillon i.i.d. d'une variable aléatoire suivant une loi normale $\mathcal{N}(0, \sigma^2)$. On veut estimer σ^2 inconnu. On définit la famille de statistiques $Q_a := a\sum_{i=1}^n X_i^2$.

1) Calculer le biais en fonction de a de Q_a comme estimateur de σ^2.

2) En déduire le risque quadratique moyen associé à Q_a.

3) Pour quelles valeurs de a ce risque est-il minimum ? L'estimateur correspondant est-il alors sans biais ?

3. Machine outil

Une machine outil fabrique en série des axes dont le diamètre est une variable aléatoire X. On suppose que X suit une loi normale d'espérance m et de variance σ^2 inconnues. On dispose d'un échantillon i.i.d. de taille n de X.

1) Donner les estimateurs du maximum de vraisemblance de m et de σ^2.

2) Montrer, sans utiliser les résultats du cours sur les estimateurs du maximum de vraisemblance, que les estimateurs sont convergents.

3) Donner l'estimateur, noté S^2, sans biais de σ^2. Comparer son risque quadratique avec celui du maximum de vraisemblance.

4) Application numérique : Pour un échantillon de $n = 9$ pièces, on a mesuré les diamètres et obtenu les mesures suivantes : $(398.4, 398.5, 399, 399.2, 399.8, 400, 400.1, 400.3, 400.6)$. Donner les estimations de m et σ^2.

4. Variables de Poisson

Soit $(x_1, \ldots, x_n)$ la réalisation d'un n-échantillon i.i.d. d'une variable aléatoire suivant une loi de Poisson $X \sim Pois(\theta)$.

1) Donner l'estimateur du maximum de vraisemblance $\hat{\theta}_{MV}$ de θ.

2) Propriété de $\hat{\theta}_{MV}$: biais, convergence.

5. Chaîne de Markov

On considère une chaîne de Markov de matrice de transition $M = \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}$ où $0 < p, q < 1$. On dispose d'une réalisation $(X_1, \ldots, X_n)$, avec $X_1 = e_1$.

1) Calculer la loi invariante de la chaîne en fonction de p et q.

2) Donner la vraisemblance d'une réalisation $(x_1, \ldots, x_n)$ de $(X_1, \ldots, X_n)$ en fonction de p et q.

3) Calculer les estimateurs du maximum de vraisemblance $\hat{p}$ et $\hat{q}$ de p et q.

6. Variables exponentielles

On considère un échantillon i.i.d. $X_1, \ldots, X_n$ de variable aléatoire exponentielle de paramètre $\frac{1}{\lambda}$, $\lambda \geq 0$, c'est-à-dire la densité est donnée par $f_\lambda(x) = \frac{1}{\lambda} \exp(-\frac{x}{\lambda}) \mathbb{1}_{\mathbb{R}^+}(x)$, $\mathbb{E}(X_1) = \lambda$ et $\mathrm{Var}\,(X_1) = \lambda^2$.

1) Trouver l'estimateur du maximum de vraisemblance $\hat{\lambda}_{MV}$ de λ.

2) En utilisant une approximation gaussienne de la loi de $\hat{\lambda}_{MV}$, donner un intervalle de confiance de niveau α pour λ.

3) Application : On observe la durée de vie de 100 ampoules de loi f_λ et on trouve une durée de vie moyenne de 1000 heures. Calculer l'intervalle de confiance de niveau $0,9$ pour λ.

7. Tirage aléatoire avec remise

On considère une urne contenant 100 boules $(B_1, \ldots, B_{100})$, dont k noires et $100-k$ blanches. On tire au hasard uniformément une boule, on note sa couleur et on la remet dans l'urne. Soit X la variable aléatoire qui vaut 1 si la boule tirée est noire et 0 sinon. On considère $(X_1, \ldots, X_n)$ les résultats de n tirages avec remise.

1) Exprimer en fonction de k la loi de X_1 et du couple (X_1, X_2).

2) Calculer la vraisemblance d'une réalisation $(x_1, \ldots, x_n)$ de $(X_1, \ldots, X_n)$ en fonction de k.

3) Quel est l'estimateur du maximum de vraisemblance $\hat{k}_{MV}$ du paramètre k ?

4) Montrer que $\hat{k}_{MV}$ est sans biais et convergent.

5) On suppose dans cette question que la taille de l'échantillon est $n = 100$. Quelle est la loi de $\hat{k}_{MV}$, en fonction de k ? Donner une majoration de sa variance et en déduire un intervalle de confiance à plus de 90% pour k en utilisant une approximation gaussienne.

TD 5

1. Bernoulli

Soit un couple i.i.d. (X_1, X_2) tel que $X_1 \sim Ber(p)$. On considère les hypothèses suivantes : $H_0 : p = 1/2$ contre $H_1 : p = 3/4$. En choisissant la statistique du test $T = X_1 + X_2$, faire la liste de toutes les régions de rejet tel que le niveau du test α soit plus petit que 1/2. Calculer respectivement leur puissance du test β. Déterminer celle qui minimise la somme des risques de 1ère et 2ème espèce $\alpha + \beta$.

2. Urne

Dans une urne on sait qu'il y a $N = 30$ boules dont R sont rouges et $N - R$ sont noires. On tire avec remise $n = 15$ boules et on note leur couleur, teste avec un niveau α de moins de 5% les hypothèses $H_0 : R = 15$ contre $H_1 : R = 20$. Quel est alors la puissance du test ?

3. Test gaussien

Soit X une variable aléatoire normale d'écart type inconnu. On veut choisir entre les hypothèses $H_0 : m = m_0$ et $H_1 : m < m_0$ à l'aide d'un échantillon de taille n, $(X_1, \ldots, X_n)$.

1) Quelle est la statistique utilisée pour tester H_0 contre H_1 ? Quelle est sa loi sous H_0 ?

2) Quelle est alors la règle de décision de seuil 5% pour tester H_0 contre H_1 ?

4. Analyse de l'eau

L'eau rejetée par une usine contient des sels de cadmium qui peuvent être polluants. Plus précisément, la teneur moyenne en sels de cadmium de cette eau est de $10\mu g$ par litre alors que la concentration tolérée par la loi doit être strictement inférieure à $10\mu g$ par litre. Les responsables de l'usine expérimentent un procédé pour diminuer la pollution de cette eau. Après l'utilisation du procédé. On prélève, de manière indépendante, $n = 10$ fois un litre d'eau dans les effluents de l'usine et on note $(X_1, \ldots, X_n)$ les résultats de la teneur en sels de cadmium. On suppose que les X_i sont des réalisations de variables aléatoires indépendantes de loi normale d'espérance m et de variance $\sigma^2 = 1$.

A l'aide de cet échantillon de taille n, les responsables de l'usine veulent choisir entre les deux hypothèses suivantes : H_1 : { le procédé est efficace i.e. la teneur moyenne m en sels de cadmium par litre d'eau est strictement inférieure à 10 } auquel cas ils l'adopteront et H_0 : { le procédé est inefficace i.e. m reste égale à 10 } auquel cas ils feront appel à des procédés plus sophistiqués et coûteux.

1) Formaliser le problème de test ainsi posé.

2) Les responsables de l'usine veulent choisir la règle suivante : "Décider que le procédé est efficace si l'un des prélèvements a une teneur en sels de cadmium strictement inférieure à 9". Formaliser cette règle de décision et calculer, chacun des risques associés.

3) Un jeune ingénieur choisit la règle suivante : "Décider que le procédé est efficace si la moyenne des n teneurs en sels de cadmium des prélèvements est inférieure à 9.4". Formaliser cette règle de décision et calculer, chacun des risques associés.

4) Quelle est la règle de décision U.P.P. au seuil 5% pour tester H_0 contre H_1 ?

5. Test d'adéquation du χ^2

On lance 60 fois un dé et on note les observations suivantes : n_i le nombre de fois où la valeur i, $i \in \{1, \ldots, 6\}$, est sortie lors des 60 tirages.

i	1	2	3	4	5	6
n_i	9	6	11	6	12	16

1) Effectuer un test de seuil 5% pour déterminer si le dé est pipé ou non.

2) A partir de quelle taille d'échantillon les mêmes fréquences observées auraient-elles conduit à considérer le dé comme pipé ?