Université de Rennes 1 Année 2013–2014

PRSB (Probabilités et Statistiques, L1 Bio)

Examen de rattrapage
19 juin 2014 - Durée : 2 heures

Vous pouvez utiliser des calculatrices et une feuille A4 avec vos notes.
Les téléphones portables sont interdits. Justifiez toutes vos réponses.

Exercice 1

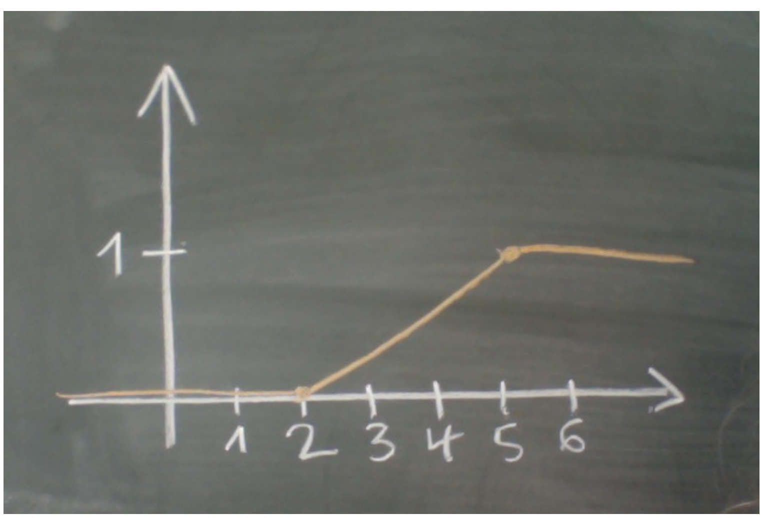
(a) Il y a 6 écoles spécialisées dans un certain domaine scientifique, et vous devez écrire une liste de préférence : vos 3 écoles préférées, classées dans l'ordre de préférence. Combien de listes possibles y a-t-il ? Ensuite, si parmi les 6 écoles, il y a 4 privées et 2 publiques, combien de listes contiennent exactement une école publique ? Réponse : $6 \cdot 5 \cdot 4 = 120$ et $3 \cdot 2 \cdot 4 \cdot 3 = 72$

(b) Soient A et B deux événements, et supposons que $\mathbb{P}(B) \neq 0$. Donner la définition de *la probabilité conditionnelle de A sachant que B est réalisé.*

(c) Soit X une variable aléatoire qui suit une loi uniforme (continue) sur l'intervalle $[2, 5]$, c.à.d., $X \sim \mathcal{U}([2, 5])$. Calculer la valeur de $\mathbb{P}(X \in [2, 3])$ et de $\mathbb{P}(X \in [0, 3])$. Réponse : $\frac{1}{3}$ et $\frac{1}{3}$

(d) J'ai une machine qui peut générer des nombres aléatoires uniformément distribués dans l'intervalle $[0, 1]$ (selon la loi $\mathcal{U}([0, 1])$). Comment fais-je pour créer des nombres aléatoires uniformes dans l'intervalle $[2, 5]$? Réponse : à partir d'un nombre x obtenu de ma machine, je calcule $2 + 3x$.

(e) Dessiner le graphe de la fonction de répartition de la loi $\mathcal{U}([2, 5])$.



(f) Une loterie nationale a un tirage par semaine, et toutes les semaines il y a le même, très grand, nombre de joueurs. En moyenne, il y a 1 grand gagnant, c.à.d. un joueur qui a tous les numéros exacts (mais bien sûr, parfois il y en a aucun, et parfois plusieurs). Soit X le nombre de grands gagnants la semaine prochaine. Cette variable aléatoire suit quelle loi ? (Donnez le nom et les paramètres de la loi.) Solution : Poisson $\mathcal{P}(1)$. Évidemment ce n'est pas exactement vrai : en réalité la loi de Poisson n'est qu'une approximation, qui est *vraiment* valable quand le nombre de joueurs tend vers $+\infty$.

(g) Si une variable aléatoire X suit une loi géométrique $\mathcal{G}(\frac{1}{3})$, déterminez $\mathbb{P}(X = 3)$. $\mathbb{P}(X = 3) = (\frac{2}{3})^2 \cdot \frac{1}{3} = \frac{4}{27}$.

(h) (1.) Soit X une variable aléatoire discrète qui prend les valeurs 0 et 1, chacune avec probabilité $\frac{1}{2}$. Calculer $\mathbb{E}(10 \cdot X)$ et $\mathbb{V}(10 \cdot X)$. Réponse : $\mathbb{E}(10 \cdot X) = 10 \cdot \mathbb{E}(X) = 10 \cdot \frac{1}{2} = 5$. $\mathbb{V}(10 \cdot X) = 10^2 \cdot \mathbb{V}(X) = 100 \cdot \frac{1}{4} = 25$.

(2.) Soient $X_1, \ldots, X_{10}$ des variables aléatoires *indépendantes*, toutes de la même loi que X (c.à.d. $X_i \sim \mathcal{B}(1, \frac{1}{2})$). Calculer $\mathbb{E}(X_1 + \ldots + X_{10})$ et $\mathbb{V}(X_1 + \ldots + X_{10})$. Réponse : $\mathbb{E}(X_1 + \ldots + X_{10}) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_{10}) = \frac{1}{2} + \ldots + \frac{1}{2} = 5$ et $\mathbb{V}(X_1 + \ldots + X_{10}) \overset{*}{=} \mathbb{V}(X_1) + \ldots + \mathbb{V}(X_{10}) = 2,5$, où l'égalité $\overset{*}{=}$ est vraie car les X_i sont indépendantes.

(i) Pour estimer la proportion d'individus dans une population qui est de groupe sanguin A, je prends un échantillon de 60 individus au hasard, et je teste la proportion d'individus de type A dans cet échantillon. Cela me donne une *estimation* raisonnable. Si maintenant je veux diviser par 10 l'erreur d'estimation attendu, je dois choisir un échantillon plus grand – quelle taille d'échantillon est nécessaire ? (On pourra utiliser le TCL.)Réponse : On a vu dans le TCL que l'erreur attendu d'une telle estimation se comporte comme $\frac{1}{\sqrt{n}}$, donc pour diviser l'erreur attendu par n il faut multiplier la taille n de l'échantillon par 100 : il me faut 6000 individus.

Exercice 2 On considère une pièce d'argent équilibrée (non-pipée) dont les deux côtés sont marqués "0" et "1". On jette cette pièce deux fois – on a donc un univers $\Omega = \{(0,0),\ (0,1),\ (1,0),\ (1,1)\}$, dont chaque élément apparaît avec probabilité $\frac{1}{4}$. Considérons les deux variables aléatoires

$$X((i,j)) = \min(i,j) \quad \text{et} \quad Y(i,j) = |i-j|$$

(a) Calculer l'espérance $\mathbb{E}(X)$ et $\mathbb{E}(Y)$. $\mathbb{E}(X) = 0 \cdot \frac{3}{4} + 1 \cdot \frac{1}{4} = \frac{1}{4}$ et $\mathbb{E}(Y) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}$

(b) Calculer la covariance $Cov(X,Y)$. $= \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y) \overset{*}{=} 0 - \frac{1}{4} \cdot \frac{1}{2} = -\frac{1}{8}$ où l'égalité $\overset{*}{=}$ vient du fait que pour tout couple (i,j) on a $X(i,j) = 0$ ou $Y(i,j) = 0$.

(c) Les deux variables aléatoires X et Y sont-elles positivement corrélées, négativement corrélées, ou non corrélées ? Sont elles dépendantes ou indépendantes ? Réponse : Dépendantes, car négativement corrélées ($Cov(X,Y) < 0$).

Exercice 3 Parmi 100 personnes prises au hasard, on observe 40 de groupe sanguin A. Donnez un intervalle de confiance pour la proportion des individus de groupe sanguin A dans la population, où le niveau de confiance exigé est de 99%. Réponse : Les bornes de l'intervalle de confiance sont $\frac{40}{100} \pm \frac{2{,}5758}{2\cdot\sqrt{100}} = 0,4 \pm 0,12879 = [0,27121,\ 0,52879]$. Donc la proportion est quelque part entre 27,1% et 52,9%.

Exercice 4 On observe deux populations d'une certaine espèce animale dans deux environnements différents. On mesure leurs tailles, et on obtient les résultats suivants :

	Population A	Population B
Petite taille	20	40
Taille moyenne	30	50
Grande taille	50	60

Tester au niveau de confiance de 90% l'hypothèse nulle selon laquelle la distribution des tailles dans les deux populations est la même (c.à.d. on a indépendance).

Solution : On fait les totaux (1er tableau) et on calcule la distribution théorique sous H_0 (2ème tableau) :

	A	B	Total
Petite taille	20	40	60
Taille moyenne	30	50	80
Grande taille	50	60	110
Total	100	150	250

	A	B	Total
Petite taille	24	36	60
Taille moyenne	32	48	80
Grande taille	44	66	110
Total	100	150	250

On calcule $\chi^2_{empirique} = \frac{20^2}{24} + \frac{30^2}{32} + \ldots + \frac{60^2}{66} - 250 = 2,683$ Il faut regarder la loi χ^2 avec $(2-1)(3-1) = 2$ degrés de liberté, c.à.d. $\chi^2(2)$. Le seuil critique pour rejeter l'hypothèse nulle est 4,60. Puisque 2,683 est bien plus petit que ce seuil, j'accepte l'hypothèse nulle.

Quantiles de la loi normale $\mathcal{N}(0,1)$. Notation : $q_p = \Phi^{-1}(p)$.

p	0,95	0,975	0,995	0,999	0,9995	0,9999	0,99995	0,99999
q_p	1,645	1,9600	2,5758	3,0902	3,2905	3,7190	3,8906	4,2649

Quantiles d'ordre $1-\alpha$ **de la loi** $\chi^2(k)$ (loi du χ^2 à k degrés de liberté)

χ^2	α=0,2	α=0,1	α=0,05	α=0,01	α=0,001
$k=1$	1.64	2.71	3.84	6.64	10.83
$k=2$	3.22	4.60	5.99	9.21	13.82
$k=3$	4.64	6.25	7.82	11.34	16.27