

Probabilités

Exercices

1. Un examen comporte quinze questions, chacune ayant trois réponses possibles. Supposons que 70% des étudiants passant l'examen sont bien préparés et répondent correctement à chaque question avec une probabilité de 0.8; les 30% restants répondent au hasard.
 - (a) Caractériser la distribution de S , la note de chaque étudiant, si un point est accordé à chaque bonne réponse.
 - (b) Il faut huit bonnes réponses pour réussir l'examen. Conditionnellement au fait qu'un étudiant réussisse un examen, quelle est la probabilité qu'il était bien préparé?
2. Le paradoxe de Simpson fournit une illustration de la nécessité d'une approche conditionnelle en Statistique. Soient deux traitements médicaux, $T1$ et $T2$, $T1$ étant appliqué à cinquante patients et $T2$ à cinquante autres. Le résultat de cette expérience donne les pourcentages de survie suivants : 40% pour le traitement $T1$, 32% pour le traitement $T2$. Donc le traitement $T1$ semble meilleur puisqu'il entraîne un taux de survie plus élevé. Cependant, si on prend l'âge en compte, et l'on sépare les patients entre juniors (50) et seniors (50), les taux de succès sont donnés dans la table suivante :

	$T1$	$T2$
junior	40%	50%
senior	10%	35%

et $T1$ est moins bon que $T2$ dans les deux cas. Expliquer ce paradoxe en utilisant la probabilité conditionnelle et le théorème de Bayes.

3. Prenons le cas du dépistage du cancer du col d'utérus par frottis sanguin. Ce test bien que simple à réaliser, est assez imprécis: il y a un sensibilité (la proportion des personnes atteintes pour lequel il est positif) d'environ 70%, et une spécificité (la proportion des personnes non atteintes pour lequel il est négatif) d'environ 98%. De plus l'incidence de la maladie est d'environ d'une femme sur 10000. Quelle est la probabilité qu'une personne testée soit atteinte d'un cancer si le test est positif?

4. Des estimations de la prévalence de la maladie d'Alzheimer ont été récemment fournies par Pfeffer et al. [8] Les estimations sont présentées au tableau

Prévalence de la maladie d'Alzheimer (cas pour 100 habitants)

Tranche d'âge	hommes	femmes
65–69	1.6	0.0
70–74	0.0	2.2
75–79	4.9	2.3
80–84	8.6	7.8
85+	35.0	27.9

Supposons qu'un homme de 77 ans, une femme de 76 ans et une femme de 82 ans non liés soient choisis dans une communauté.

- Quelle est la probabilité que ces trois personnes soient atteintes de la maladie d'Alzheimer?
 - Quelle est la probabilité qu'au moins une des femmes soit atteinte de la maladie d'Alzheimer?
 - Quelle est la probabilité qu'au moins une des trois personnes soit atteinte de la maladie d'Alzheimer?
 - Quelle est la probabilité qu'une des trois personnes soit atteinte de la maladie d'Alzheimer?
 - Supposons que l'une des trois personnes soit atteinte de la maladie d'Alzheimer, sans savoir laquelle. Quelle est la probabilité conditionnelle que la personne touchée soit une femme?
 - Supposons que nous savons que deux des trois personnes sont atteintes de la maladie d'Alzheimer. Quelle est la probabilité conditionnelle qu'il s'agisse de deux femmes?
 - Supposons que nous savons que deux des trois personnes sont atteintes de la maladie d'Alzheimer. Quelle est la probabilité conditionnelle qu'ils aient tous les deux moins de 80 ans?
5. Une étude de validation doit être réalisée dans un hôpital local pour vérifier l'exactitude de l'évaluation de l'infection (*INF*) acquise en milieu hospitalier après un pontage coronarien (pontage coronarien ou pontage coronarien). Au cours d'une année donnée, l'hôpital effectue 1100 procédures de pontage aortique. Un algorithme des Centres de contrôle et de prévention des maladies (*CDC*) est actuellement utilisé pour classer les sujets comme ayant un *INF*. Pour valider cet algorithme, tous les sujets *CDC+* ($N = 100$) et un échantillon aléatoire de sujets *CDC-* ($N = 1000$) seront vérifiés par un boursier spécialiste des maladies infectieuses (ID) et une enquête détaillée sera effectuée, y compris une analyse comparative. examen des dossiers et documentation de l'utilisation des antibiotiques. Supposons que la détermination du boursier est correcte. Supposons que

100 sujets $CDC+$ soient identifiés, et que le boursier en confirme 80. En raison du grand nombre de sujets $CDC-$ (1000), seul un échantillon de 100 est étudié, et le boursier en confirme 90.

- (a) Quel est le $P(INF|CDC+)$ de l'algorithme CDC ?
 - (b) Quel est le $P(NINF|CDC-)$ de l'algorithme CDC ?
 - (c) Quelle est la sensibilité de l'algorithme CDC ?
 - (d) Quelle est la spécificité de l'algorithme CDC ?
6. Le cancer du sein est largement considéré comme une maladie hormonale. L'estradiol est une hormone importante dans la recherche sur le cancer du sein. Les données du tableau sur les taux sériques d'estradiol ont été obtenues chez 213 cas de cancer du sein et 432 témoins du même âge. Toutes les femmes étaient âgées de 50 à 59 ans.

Données sur le sérum-estradiol

Serum estradiol (pg/mL)	Cases ($N = 213$)	Contrôles ($N = 432$)
1-4	28	72
5-9	96	233
10-14	53	86
15-19	17	26
20-24	10	6
25-29	3	5
30+	6	4

Supposons qu'un taux sérique d'estradiol supérieur à 20 pg/ml soit proposé comme critère de dépistage pour l'identification des cas de cancer du sein.

- (a) Quelle est la sensibilité de ce test?
- (b) Quelle est la spécificité de ce test?
- (c) L'échantillon précédent a été sélectionné pour sur échantillonner les cas. Dans la population générale, la prévalence du cancer du sein est d'environ 2% chez les femmes âgées de 50 à 59 ans. Quelle est la probabilité de cancer du sein chez les femmes de la population générale âgées de 50 à 59 ans présentant un taux sérique d'estradiol ≥ 20 pg / mL?