Chapitre 1

Probabilités de base

Rappels de cours

1. Analyse combinatoire

✦ **Permutation** : le nombre de possibilités différentes de ranger n objets distincts dans un ordre donné est

$$n! = n(n-1)(n-2)\dots 1.$$

NB : par convention, $0! = 1$.

✦ **Arrangement** : le nombre de possibilités d'ordonner p objets pris parmi n est

$$A_n^p = \frac{n!}{(n-p)!}.$$

✦ **Combinaison** : le nombre de possibilités de choisir p objets pris parmi n, sans tenir compte de l'ordre, est

$$C_n^p = \frac{n!}{p!\,(n-p)!}.$$

NB : C_n^p est parfois noté $\binom{n}{p}$.

2. Probabilités

2.1 Notations

✦ Événement (ou résultat d'une expérience aléatoire) : A, B, C, D...

✦ Événement impossible : $\varnothing$.

✦ Ensemble des événements (univers) : Ω.

✦ Événement complémentaire (ou contraire) de A : $\overline{A}$.

✦ Événement A ou Événement B : $A \cup B$.

✦ Événement A et Événement B : $A \cap B$.

✦ Probabilité que l'événement A se produise : $\mathbb{P}(A)$.

2.2 Définitions et propriétés

✦ Deux événements A et B sont **indépendants** si et seulement si la réalisation de A n'influe pas sur la réalisation de B (et inversement).

✦ Deux événements A et B sont **incompatibles** si et seulement s'ils ne peuvent se réaliser en même temps ($A \cap B = \varnothing$).

✦ **Complémentaire de l'union** : $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

✦ **Complémentaire de l'intersection** : $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

✦ Pour un événement A, $\mathbb{P}(A) \in [\,0\,;\,1\,]$, et $\mathbb{P}\left(\overline{A}\right) = 1 - \mathbb{P}(A)$.

✦ $\mathbb{P}(\varnothing) = 0$, $\mathbb{P}(\Omega) = 1$.

2.3 Calcul de probabilités

Soient A et B deux événements.

✦ Si tous les cas possibles sont **équiprobables**,

$$\mathbb{P}(A) = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}.$$

✦ **Probabilité conditionnelle** : $\mathbb{P}\left(A \,|\, B\right) = \dfrac{\mathbb{P}\left(A \cap B\right)}{\mathbb{P}(B)}$.

✦ **Probabilité de l'intersection** : $\mathbb{P}\left(A \cap B\right) = \mathbb{P}\left(A \,|\, B\right) \cdot \mathbb{P}(B)$.
Si A et B sont indépendants, $\mathbb{P}\left(A \cap B\right) = \mathbb{P}\left(A\right) \cdot \mathbb{P}(B)$.
Pour trois événements, $\mathbb{P}\left(A \cap B \cap C\right) = \mathbb{P}\left(A \,|\, B \cap C\right) \cdot \mathbb{P}\left(B \,|\, C\right) \cdot \mathbb{P}(C)$.

✦ **Probabilité de l'union** : $\mathbb{P}\left(A \cup B\right) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}\left(A \cap B\right)$.
Si A et B sont incompatibles, $\mathbb{P}\left(A \cup B\right) = \mathbb{P}(A) + \mathbb{P}(B)$.

✦ **Théorème de Bayes** :

$$\mathbb{P}(B\,|A\,) = \frac{\mathbb{P}(A\,|B\,)\cdot\mathbb{P}(B)}{\mathbb{P}(A)}.$$

✦ **Théorème des probabilités totales** :

$$\mathbb{P}(A) = \mathbb{P}(A\,|B\,)\cdot\mathbb{P}(B) + \mathbb{P}(A\,|\overline{B}\,)\cdot\mathbb{P}(\overline{B}).$$

Pour des événements $B_1,\ B_2 \ldots B_n$ deux à deux incompatibles et tels que $B_1 \cup B_2 \cup \ldots \cup B_n = \Omega$,

$$\mathbb{P}(A) = \mathbb{P}(A\,|B_1\,)\cdot\mathbb{P}(B_1) + \mathbb{P}(A\,|B_2\,)\cdot\mathbb{P}(B_2)\ldots + \mathbb{P}(A\,|B_n\,)\cdot\mathbb{P}(B_n).$$

3. Applications

3.1 Diagnostic

Pour un test **T** de diagnostic d'une maladie **M**, soient les événements M « être malade », $\overline{M}$ « ne pas être malade », T « avoir un test positif » et $\overline{T}$ « avoir un test négatif ».

✦ **Prévalence** p de la maladie : $\mathbb{P}\,(M)$.

✦ **Sensibilité (Se)** du test : $\mathbb{P}\,(T\,|M\,)$.

✦ **Spécificité (Sp)** du test : $\mathbb{P}\left(\overline{T}\,\middle|\overline{M}\,\right)$.

✦ **Taux de faux positifs (FP)** du test : $\mathbb{P}\left(T\,\middle|\overline{M}\,\right)$.

✦ **Taux de faux négatifs (FN)** du test : $\mathbb{P}\left(\overline{T}\,|M\,\right)$.

✦ **Valeur prédictive positive (VPP)** du test : $\mathbb{P}\,(M\,|T\,)$.

✦ **Valeur prédictive négative (VPN)** du test : $\mathbb{P}\left(\overline{M}\,\middle|\overline{T}\,\right)$.

✦ **Courbe ROC** : graphe de Se en fonction de 1−Sp.
Le test optimal correspond au couple (1−Sp,Se) le plus proche du point (0,1).

3.2 Concordance entre deux jugements

Pour deux jugements $\mathbf{T}_1$ et $\mathbf{T}_2$, soient les événements T_i « jugement i positif » et $\overline{T_i}$ « jugement i négatif » pour $i = 1, 2$. Dans un échantillon de taille n, on observe :

		$\mathbf{T}_2$: $\overline{T_2}$	T_2	
$\mathbf{T}_1$	$\overline{T_1}$	a	b	$a+b=\overline{n}_1$
	T_1	c	d	$c+d=n_1$
		$a+c=\overline{n}_2$	$b+d=n_2$	n

✦ **Taux de concordance observé** :

$$p_o = \frac{a+d}{n}.$$

✦ **Taux de concordance calculé**, *i.e.* taux qui serait observé par hasard si les deux jugements étaient indépendants :

$$p_c = \frac{1}{n}\left(\frac{\overline{n}_1 \cdot \overline{n}_2}{n} + \frac{n_1 \cdot n_2}{n}\right).$$

✦ **Cœfficient de concordance kappa** observé dans l'échantillon :

$$\kappa = \frac{p_o - p_c}{1 - p_c}.$$

3.3 Épidémiologie

On s'intéresse à la liaison entre un facteur d'exposition **E** (présent : événement « E » ou absent : événement « $\overline{E}$ ») et une maladie **M** (présente : événement « M » ou absente : événement « $\overline{M}$ »). Dans un échantillon de taille n, on observe :

	$\overline{M}$	M	
$\overline{E}$	a	b	$a+b=\overline{e}$
E	c	d	$c+d=e$
	$a+c=\overline{m}$	$b+d=m$	n

Les mesures qui comparent le risque de maladie des sujets présentant le facteur d'exposition à celui des sujets ne le présentant pas sont :

✦ **Odds-ratio** dans l'échantillon :

$$\mathrm{OR} = \frac{\mathbb{P}(M\,|E)\,/\mathbb{P}\left(\overline{M}\,|E\right)}{\mathbb{P}\left(M\,|\overline{E}\right)/\mathbb{P}\left(\overline{M}\,|\overline{E}\right)} = \frac{a\cdot d}{b\cdot c}.$$

✦ **Risque relatif** dans l'échantillon :

$$\mathrm{RR} = \frac{\mathbb{P}(M\,|E)}{\mathbb{P}\left(M\,|\overline{E}\right)} = \frac{d/e}{b/\overline{e}}.$$

NB : si la maladie est rare, RR $\approx$ OR.

3.4 Données censurées

Dans une population, on s'intéresse au délai de survie T.

✦ **Fonction de survie** : le délai T est décrit par la fonction de survie S telle que $S(x) = \mathbb{P}(T > x)$ pour tout $x \geq 0$.

✦ **Données censurées** : dans un échantillon de taille n, m sujets décèdent pendant la période d'observation, et les délais de survie observés sont $t_1 \leq t_2 \leq \ldots \leq t_m$. Les sujets qui ne décèdent pas pendant la période de suivi (ou qui décèdent d'une cause autre que celle d'intérêt) sont dits *censurés*.
NB : il s'agit d'un seul type de censure, qui ne couvre pas tous les cas.

✦ **L'estimateur de Kaplan-Meier** : à chaque temps de décès t_i, pour $i = 1, \ldots, m$, correspondent r_i, le nombre de sujets à risque de décéder juste avant t_i, et d_i, le nombre de sujets décédés en t_i. Alors la fonction de survie S peut être estimée par l'estimateur du produit-limite de Kaplan-Meier $\widehat{S}$ tel que

$$\widehat{S}(x) = \prod_{t_i \leq x} \mathbb{P}(T > t_i\,|T > t_{i-1}) = \prod_{t_i \leq x}\left(1 - \frac{d_i}{r_i}\right)$$

pour tout $x \geq 0$ (où $\prod$ symbolise le *produit* des différentes probabilités).

QCMs avec corrigés détaillés

QCM 1

Dans un certain village, 70% des villageois sont des menteurs. À un croisement où deux choix sont possibles, vous demandez votre route à l'un de ces villageois. Dans 20% des cas, celui-ci ne connaît pas la bonne route ; il vous indique alors une route au hasard. Dans les autres cas, connaissant la bonne route, il vous donnera systématiquement une mauvaise indication s'il est menteur ou systématiquement la bonne dans le cas contraire. Le fait d'être un menteur est indépendant du fait de connaître la bonne route.
La probabilité, à 10^{-2} près, que le villageois vous indique la bonne direction est égale à

A. 0.18 **B.** 0.34 **C.** 0.50 **D.** 0.66 **E.** 0.82.

QCM 2

Un cadenas est sécurisé par un code secret à 4 chiffres (de 0 à 9, avec répétitions possibles). Sachant que le code contient un seul « 9 », le nombre de codes possibles est

A. 10 **B.** 100 **C.** 9^3 **D.** 4×9^3 **E.** 10 000.

---

20 personnes adultes, également réparties entre les deux sexes, participent à un essai clinique. On suppose que l'équilibre prévaut, chez les hommes et chez les femmes, entre fumeurs et non fumeurs. On procède à un tirage au sort afin de constituer deux groupes de 10 personnes chacun affectés à deux traitements distincts.

QCM 3 (avec calculatrice)

La probabilité, à 10^{-3} près, pour que, dans chacun des groupes, il y ait autant de femmes que d'hommes est égale à

A. 0 **B.** 0.118 **C.** 0.250 **D.** 0.344 **E.** 0.500.

QCM 4 (avec calculatrice)

La probabilité, à 10^{-3} près, pour que, dans chacun des groupes, il y ait à la fois autant de femmes que d'hommes et autant de fumeurs que de non fumeurs est égale à

A. 0.500 **B.** 0.250 **C.** 0.115 **D.** 0.125 **E.** 0.

Une étude est menée afin d'identifier des facteurs prédictifs de décès après un infarctus du myocarde. L'étude commence un 1er janvier et se termine le 31 décembre. Trois sujets sont inclus : un sujet est inclus le 1er janvier et décède le 1er février ; un sujet est inclus le 1er février et décède le 1er novembre ; un sujet est inclus le 1er juin et ne décède pas pendant l'étude.

QCM 5

Concernant les données observées, la seule affirmation correcte est

- **A.** Aucun sujet n'est censuré le 31 décembre
- **B.** Tous les sujets sont censurés le 31 décembre
- **C.** Un sujet est censuré le 31 décembre, et sa durée de suivi est égale à sept mois
- **D.** Un sujet est censuré le 31 décembre, et sa durée de suivi est égale à un an
- **E.** Un sujet est censuré le 31 décembre, et sa durée de suivi n'est pas calculable.

QCM 6

On estime la fonction de survie par la méthode de Kaplan-Meier.
La probabilité, au % près, de survivre 8 mois ou plus est estimée à

A. 0% **B.** 33% **C.** 50% **D.** 67% **E.** 100%.

QCM 7 (avec calculatrice)

Afin de déceler les sujets maniaco-dépressifs d'une population, un questionnaire est mis au point. Si le score obtenu dépasse une certaine valeur, le sujet est considéré comme maniaco-dépressif. Pour optimiser cette prévision, une étude a été réalisée auprès d'un échantillon représentatif de 130 personnes dont 30 maniaco-dépressives. Après avoir fixé différentes valeurs seuils du score, les personnes ont été classées de la manière suivante :

Seuil	Nombre de faux négatifs	Nombre de faux positifs
50	1	18
60	2	15
70	3	12
80	5	8

Le seuil optimisant la sensibilité et la spécificité du test est

A. 50 **B.** 60 **C.** 70 **D.** 80 **E.** aucun.

QCM 8

Dans une population, un test de dépistage d'une certaine maladie est mis au point.
Connaissant la prévalence p de cette maladie, la sensibilité Se et la spécificité Sp du test, la valeur prédictive positive (VPP) et la valeur prédictive négative (VPN) du test sont respectivement égales à

A. $\text{VPP} = \dfrac{(1-p)\cdot \text{Sp}}{(1-p)\cdot \text{Sp} + p\cdot(1-\text{Se})}$, $\text{VPN} = \dfrac{p\cdot \text{Se}}{p\cdot \text{Se} + (1-p)\cdot(1-\text{Sp})}$

B. $\text{VPP} = \dfrac{(1-p)\cdot \text{Se}}{p\cdot \text{Se} + (1-p)\cdot(1-\text{Sp})}$, $\text{VPN} = \dfrac{p\cdot \text{Sp}}{p\cdot(1-\text{Se}) + (1-p)\cdot \text{Sp}}$

C. $\text{VPP} = \dfrac{(1-p)\cdot \text{Se}}{(1-p)\cdot \text{Se} + p\cdot(1-\text{Sp})}$, $\text{VPN} = \dfrac{p\cdot \text{Sp}}{(1-p)\cdot(1-\text{Se}) + p\cdot \text{Sp}}$

D. $\text{VPP} = \dfrac{p\cdot \text{Se}}{p\cdot \text{Se} + (1-p)\cdot(1-\text{Sp})}$, $\text{VPN} = \dfrac{(1-p)\cdot \text{Sp}}{(1-p)\cdot \text{Sp} + p\cdot(1-\text{Se})}$

E. $\text{VPP} = \dfrac{p\cdot \text{Sp}}{p\cdot \text{Se} + (1-p)\cdot(1-\text{Sp})}$, $\text{VPN} = \dfrac{(1-p)\cdot \text{Se}}{p\cdot(1-\text{Se}) + (1-p)\cdot \text{Sp}}$.

QCM 9

Dans une salle d'attente se trouvent 3 personnes. On s'intéresse à leur mois de naissance.
Le nombre de cas possibles et le nombre de cas où ces personnes sont toutes nées un mois différent sont, respectivement

A. 1 728 et 220 **B.** 1 728 et 1 320 **C.** 1 320 et 220
D. 531 441 et 1 728 **E.** 531 441 et 220.

QCM 10

Dans une population, une pathologie est symptomatique dans 10% des cas, et asymptomatique dans 20% des cas ; 70% de la population sont indemnes de cette maladie. Avec un certain test de dépistage, le diagnostic est positif dans 90% des cas chez les malades symptomatiques, dans 40% des cas chez les malades asymptomatiques, et dans 10% des cas chez les non-malades.
Un individu a un diagnostic positif. La probabilité, à 10^{-2} près, qu'il soit malade asymptomatique est égale à

A. 0.12 **B.** 0.33 **C.** 0.50 **D.** 0.67 **E.** 0.95.
