

Travaux pratiques II : réponses

- 1.1 $P(T) = P(T | M) \times P(M) + P(T | \bar{M}) \times P(\bar{M})$
 $= 1 \times 0.16 / 100 + (0.005 / 100) \times (1 - 0.16 / 100) = 0.00164992$ et
 $P(M | T) = P(T | M) \times P(M) / P(T) = 1 \times 0.16\% / P(T) \approx 97\%$.
- 1.2 $P(T) = 1 \times 0.16 / 100 + (0.1 / 100) \times (1 - 0.16 / 100) = 0.0025984$ et
 $P(M | T) = P(T | M) \times P(M) / P(T) = 1 \times 0.16\% / P(T) \approx 62\%$.
- 1.3 $P(T) = 1 \times 0.01 / 100 + (0.005 / 100) \times (1 - 0.01 / 100) = 0.000149995$
 $P(M | T) = P(T | M) \times P(M) / P(T) = 1 \times 0.01\% / P(T) \approx 67\%$.
- 1.4 $P(T) = 1 \times 0.01 / 100 + (0.1 / 100) \times (1 - 0.01 / 100) = 0.0010999$ et
 $P(M | T) = P(T | M) \times P(M) / P(T) = 1 \times 0.01\% / P(T) \approx 9\%$.
- 1.5 Oui. Comme $P(M | T) \approx 9\%$, pour 100 sujets avec test positif il y a en moyenne 9 porteurs et 91 non-porteurs, c'est à dire environ 10 non-porteurs pour chaque sujet porteur. Cette quantité n'est autre que le *risque de faux positifs* dont parle l'article:
 $\frac{P(\bar{M} | T)}{P(M | T)} \approx \frac{100\% - 9\%}{9\%} \approx 10$.
- 1.6 Sur l'axe horizontal sont représentées des valeurs réalistes du taux de faux positifs $P(T | \bar{M}) = 1 - \text{spécificité}$. L'axe vertical (de gauche) représente la valeur prédictive positive $P(M | T)$ calculée à l'aide de la formule de Bayes. Chaque courbe correspond à une prévalence différente. On remarque donc, par exemple, que pour les "femmes donneuses de sang" (prévalence très faible de 0.01%), $P(M | T)$ est faible pour presque toutes les valeurs de spécificités. Par contre, $P(M | T)$ est élevé pour les groupes à risque. Plus la valeur prédictive positive $P(M | T)$ est faible, plus le risque de faux positifs $\frac{P(\bar{M} | T)}{P(M | T)} = \frac{1 - P(M | T)}{P(M | T)}$ est élevé. Ce risque est représenté sur l'axe vertical de droite.
- En conclusion, le test est approprié pour dépister les individus infectés dans les groupes à risque. Toutefois, dans les groupes à faible prévalence, il produit un nombre important de faux positifs pour chaque infection correctement détectée. Avant d'utiliser ce test de façon généralisée, il faut bien planifier la prise en charge de ces cas.

- 2.1 A. On obtient le tableau suivant :

CK	Infarctus présent	Infarctus absent	Total
≥ 40 U.I./l	228	42	270
< 40 U.I./l	2	88	90
Total	230	130	360

Sensibilité = $228/230 \approx 0.99$

Spécificité = $88/130 \approx 0.68$

Valeur prédictive positive = $228/270 \approx 0.84$

Valeur prédictive négative = $88/90 \approx 0.98$

B. On obtient le tableau suivant :

CK	Infarctus présent	Infarctus absent	Total
≥ 80 U.I./l	215	16	231
< 80 U.I./l	15	114	129
Total	230	130	360

Sensibilité = $215/230 \approx 0.93$

Spécificité = $114/130 \approx 0.88$

Valeur prédictive positive = $215/231 \approx 0.93$

Valeur prédictive négative = $114/129 \approx 0.88$

C. On obtient le tableau suivant :

CK	Infarctus présent	Infarctus absent	Total
≥ 120 U.I./l	185	8	193
< 120 U.I./l	45	122	167
Total	230	130	360

Sensibilité = $185/230 \approx 0.80$

Spécificité = $122/130 \approx 0.94$

Valeur prédictive positive = $185/193 \approx 0.96$

Valeur prédictive négative = $122/167 \approx 0.73$

On remarque donc que plus la limite augmente (et donc plus il devient difficile d'être positif au test), plus la sensibilité diminue et plus la spécificité augmente. (Il n'est par contre pas forcé - comme c'est le cas ici - que la valeur prédictive positive augmente et que la valeur prédictive négative diminue.)

- 2.2 En ajoutant les 1940 patients supplémentaires dans la colonne « Infarctus absent » du tableau de la question 1.B., en gardant la même proportion de positifs et de négatifs, de façon à ce que la spécificité estimée reste la même, on obtient le tableau suivant (on a utilisé *Spécificité* = 0.88):

CK	Infarctus présent	Infarctus absent	Total
≥ 80 U.I./l	215	248	463
< 80 U.I./l	15	1822	1837
Total	230	2070	2300

Valeur prédictive positive = $215/463 \approx 0.46$

Valeur prédictive négative = $1822/1837 \approx 0.99$

On voit que les valeurs prédictives dépendent de la prévalence. Un même test (de mêmes sensibilité et spécificité) n'aura pas les mêmes valeurs prédictives dans une population où la maladie est fréquente (cas de la question 1.B.) et dans une population où elle est rare (cas présent).

Solutions des problèmes de génétique

1. Indiquons Pierre par \mathcal{P} , son épouse par \mathcal{E} , et leur enfant par \mathcal{G} . Comme le frère de \mathcal{P} est atteint, le mariage des parents de \mathcal{P} doit être du type A_4 . Comme \mathcal{P} est sain, il n'est pas rr et donc :

$$\begin{aligned}P(\mathcal{P} \text{ est } Rr \mid \mathcal{P} \text{ non } rr) &= P(\mathcal{P} \text{ est } Rr \cap \mathcal{P} \text{ non } rr) / P(\mathcal{P} \text{ non } rr) \\ &= (1/2) / (1/2 + 1/4) = 2/3, \\ P(\mathcal{P} \text{ est } RR \mid \mathcal{P} \text{ non } rr) &= 1/3.\end{aligned}$$

Analoguement, pour son épouse, en supposant qu'elle soit saine :

$$\begin{aligned}P(\mathcal{E} \text{ est } Rr \mid \mathcal{E} \text{ non } rr) &= 2/3, \\ P(\mathcal{E} \text{ est } RR \mid \mathcal{E} \text{ non } rr) &= 1/3.\end{aligned}$$

Grâce aux hypothèses A et B et à la formule de la probabilité totale on obtient

$$\begin{aligned}P(\mathcal{G} \text{ est } rr) &= P(\mathcal{G} \text{ est } rr \mid \text{le mariage de } \mathcal{P} \text{ et } \mathcal{E} \text{ est } A_4)P(\text{le mariage de } \mathcal{P} \text{ et } \mathcal{E} \text{ est } A_4) \\ &+ P(\mathcal{G} \text{ est } rr \mid \text{le mariage de } \mathcal{P} \text{ et } \mathcal{E} \text{ est } \bar{A}_4)P(\text{le mariage de } \mathcal{P} \text{ et } \mathcal{E} \text{ est } \bar{A}_4) \\ &= (1/4)(2/3)(2/3) + 0 = 1/9,\end{aligned}$$

car $P(\text{le mariage de } \mathcal{P} \text{ et } \mathcal{E} \text{ est } \bar{A}_4) = 0$.

2. Soit $p = 20\% = 1/5$ la fréquence de r . Indiquons l'individu sain par \mathcal{I} , son épouse par \mathcal{E} . Comme \mathcal{I} est sain et qu'un de ses parents est rr , \mathcal{I} est forcément Rr . Quant à son épouse on a, suivant l'hypothèse C, que

$$\begin{aligned}P(\mathcal{E} \text{ est } RR) &= (4/5)^2 = 16/25, \\ P(\mathcal{E} \text{ est } Rr) &= 2(1/5)(4/5) = 8/25, \text{ et} \\ P(\mathcal{E} \text{ est } rr) &= (1/5)^2 = 1/25.\end{aligned}$$

Les mariages entre \mathcal{I} et \mathcal{E} pourraient donc être de trois types: A_3 , A_4 , A_5 . Toutefois, selon l'hypothèse A, si le mariage est de type A_3 , il n'y a pas d'enfant rr ; cette possibilité est donc à écarter. Par contre, si le mariage est de type A_4 , l'enfant est rr avec probabilité $1/4$ et si le mariage est de type A_5 , l'enfant est rr avec probabilité $1/2$. En utilisant la règle de la probabilité totale on obtient la probabilité cherchée:

$$\begin{aligned}P(\text{enfant malade} \mid \text{un parent malade}) &= P(\text{enfant malade} \mid A_4)P(A_4) + P(\text{enfant malade} \mid A_5)P(A_5) \\ &= (1/4) \times (8/25) + (1/2) \times (1/25) = 10\%.\end{aligned}$$

Pour résoudre les problèmes 3, 5 et 6 il est nécessaire d'utiliser une extension de la formule de la probabilité totale. Supposons que la population U soit partagée en trois parties disjointes H_1, H_2 et H_3 : $U = H_1 \cup H_2 \cup H_3$ et soit A un événement. Alors:

$$P(A) = P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + P(A|H_3)P(H_3).$$

On peut étendre cette formule à des partages de la population en 4, 5, ... parties.

Démonstration intuitive. A peut être partagé en trois parties: $A \cap H_1, A \cap H_2$ et $A \cap H_3$ disjointes avec respectivement N_{AH_1}, N_{AH_2} et N_{AH_3} individus. Alors

$$\begin{aligned} P(A) &= \frac{N_A}{N} = \frac{N_{AH_1} + N_{AH_2} + N_{AH_3}}{N} = P(A \cap H_1) + P(A \cap H_2) + P(A \cap H_3) \\ &= P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + P(A|H_3)P(H_3). \end{aligned}$$

3. Les probabilités conditionnelles suivantes découlent de l'hypothèse A:

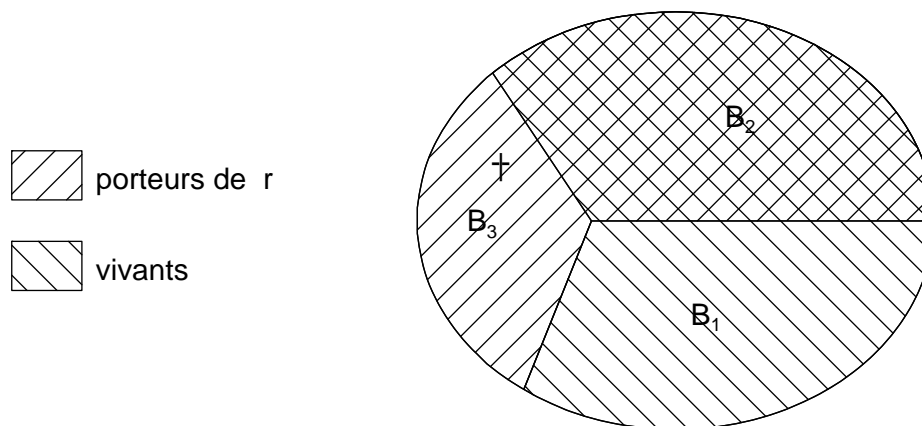
	$A = A_1$	$A = A_2$	$A = A_3$	$A = A_4$
$P(B_1 A)$	1	1/2	1/2	1/4
$P(B_2 A)$	0	1/2	1/2	1/2
$P(B_3 A)$	0	0	0	1/4

A l'aide de la formule de la probabilité totale on obtient:

$$\begin{aligned} P(B_1) &= P(B_1|A_1)P(A_1) + P(B_1|A_2)P(A_2) + P(B_1|A_3)P(A_3) + P(B_1|A_4)P(A_4) \\ &= 1 \times (1 - \lambda)^2 + \frac{1}{2}\lambda(1 - \lambda) + \frac{1}{2}\lambda(1 - \lambda) + \frac{1}{4}\lambda^2 \\ &= \left(1 - \frac{\lambda}{2}\right)^2, \end{aligned}$$

$$P(B_2) = \dots = \lambda\left(1 - \frac{\lambda}{2}\right),$$

$$P(B_3) = \dots = \left(\frac{\lambda}{2}\right)^2.$$



Génération suivante

Soit Q l'abréviation de "fils porteur de r " et V l'abréviation de "fils vivant". Alors la probabilité cherchée est

$$\begin{aligned} P(Q | V) &= P(Q \cap V)/P(V) = P(B_2)/P(\bar{B}_3) = P(B_2)/(1 - P(B_3)) \\ &= \frac{\lambda(1 - \lambda/2)}{1 - \lambda^2/4} = \frac{\lambda}{1 + \lambda/2} \end{aligned}$$

Nous remarquons que $\lambda/(1 + \lambda/2) < \lambda$. Donc, la fréquence de porteurs diminue d'une génération à l'autre. En d'autres termes, un facteur mortel à tendance à s'éliminer naturellement. En outre, le rapport entre λ et $\lambda/(1 + \lambda/2)$ est une fonction croissante en λ . On en déduit qu'un facteur mortel s'élimine d'autant plus vite qu'il est répandu.

4. Comme mon frère est rr , le mariage de mes parents est le mariage noté A_4 . Ecrivons, par exemple, $P_{A_4}(B_1) = P(\text{moi} = B_1 | \text{mes parents} = A_4)$. Avec cette notation, on a (voir Problème 3):

$$P_{A_4}(B_1) = \frac{1}{4}, \quad P_{A_4}(B_2) = \frac{1}{2}, \quad P_{A_4}(B_3) = \frac{1}{4}.$$

Donc:

$$\begin{aligned} P_{A_4}(\text{moi porteur} | \text{moi vivant}) &= P_{A_4}(\text{moi porteur} \cap \text{moi vivant})/P_{A_4}(\text{moi vivant}) \\ &= \frac{P_{A_4}(B_2)}{P_{A_4}(\bar{B}_3)} = \frac{1/2}{3/4} = \frac{2}{3}. \end{aligned}$$

5. Par le Problème 4, on a que:

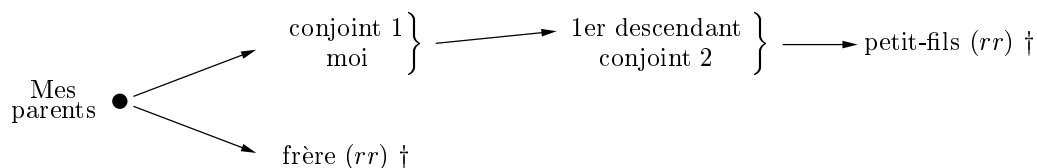
$$P(\text{moi} = Rr) = \frac{2}{3} \quad \text{et} \quad P(\text{moi} = RR) = \frac{1}{3}.$$

Alors:

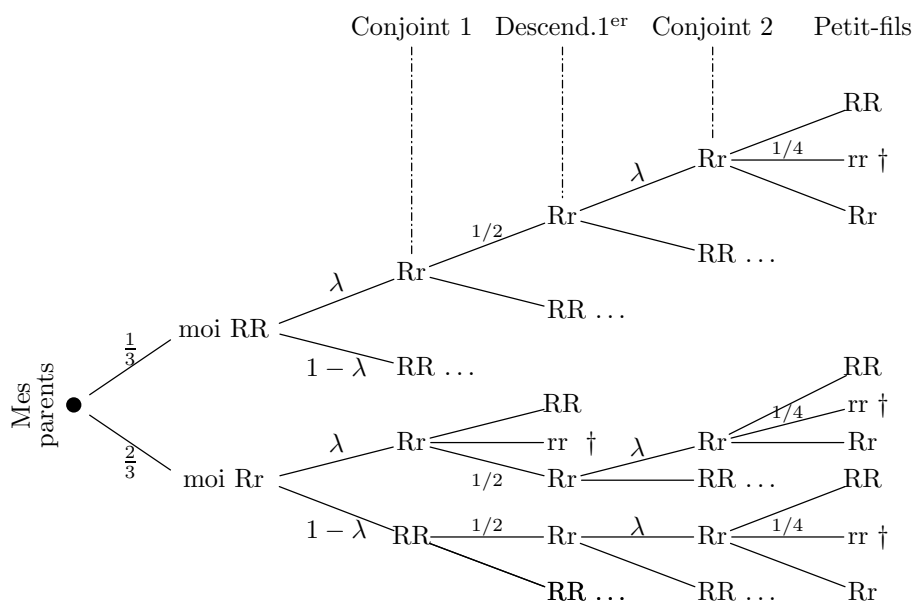
	$A = A_1$	$A = A_2$	$A = A_3$	$A = A_4$
$P(\text{mon mariage} = A) :$	$\frac{1}{3}(1 - \lambda)$	$\frac{1}{3}\lambda$	$\frac{2}{3}(1 - \lambda)$	$\frac{2}{3}\lambda$
$P(\text{mon fils} = rr A) :$	0	0	0	$\frac{1}{4}$

On en tire donc que $P(\text{mon fils} = rr) = \frac{1}{4} \cdot \frac{2}{3}\lambda = \frac{\lambda}{6}$.

6. Soit l'arbre généalogique qui nous intéresse:



Développons-le en marquant les branches qui se terminent par la mort du petit-fils:



Notons: A = "je porte r ", B = "mon petit-fils est rr ". Alors:

$$P(B|A) = \lambda \cdot \frac{1}{2} \cdot \lambda \cdot \frac{1}{4} + (1 - \lambda) \cdot \frac{1}{2} \cdot \lambda \cdot \frac{1}{4} = \frac{\lambda}{8} \quad (\text{branches du bas})$$

$$P(B|\bar{A}) = \lambda \cdot \frac{1}{2} \cdot \lambda \cdot \frac{1}{4} = \frac{\lambda^2}{8} \quad (\text{branches du haut})$$

Par le théorème de Bayes, on obtient la probabilité cherchée:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})} = \frac{\frac{2}{3} \cdot \frac{\lambda}{8}}{\frac{2}{3} \cdot \frac{\lambda}{8} + \frac{1}{3} \cdot \frac{\lambda^2}{8}} = \frac{2}{2 + \lambda}$$

Note. Selon le résultat du Problème 3, la fréquence de porteurs dans la deuxième génération est $\lambda/(1 + \lambda/2)$ mais, pour simplifier les calculs, nous avons supposé qu'elle est λ . Calculez la probabilité cherchée sans la simplification.