Dans la revue *Lancet Public Health*, les chercheurs affirment qu'au 11 mai 2020, 5,7 % des adultes français avaient déjà été infectés par la COVID 19.

On se servira de cette donnée pour les parties A et B de cet exercice.

Partie A

1. On prélève un individu dans la population française adulte au 11 mai 2020.
   On note I l'évènement : « l'adulte a déjà été infecté par la COVID 19 »
   Quelle est la probabilité que cet individu prélevé ait déjà été infecté par la COVID 19 ?
2. On prélève un échantillon de 100 personnes de la population supposées choisies de façon indépendante les unes des autres. On assimile ce prélèvement à un tirage avec remise.
   On appelle X la variable aléatoire qui compte le nombre de personnes ayant déjà été infectées.
   (a) Justifiez que X suit une loi binomiale dont on donnera les paramètres.
   (b) Calculer son espérance mathématique. Interpréter ce résultat dans le cadre de l'exercice.
   (c) Quelle est la probabilité qu'il n'y ait aucune personne infectée dans l'échantillon ?
   On donnera une valeur approchée à 10^{-4} près du résultat.
   (d) Quelle est la probabilité qu'il y ait au moins 2 personnes infectées dans l'échantillon ?
   On donnera une valeur approchée à 10^{-4} près du résultat.
   (e) Déterminer le plus petit entier n tel que $P(X \leqslant n) > 0,9$.
   Interpréter ce résultat dans le contexte de l'exercice.

Partie B :

Un test a été mis en place : celui-ci permet de déterminer (même longtemps après l'infection), si une personne a ou non déjà été infectée par la COVID 19.

Si le test est positif, cela signifie que la personne a déjà été infectée par la COVID 19.

Deux paramètres permettent de caractériser ce test : sa sensibilité et sa spécificité.

La **sensibilité** d'un test est la probabilité qu'il soit positif sachant que la personne a été infectée par la maladie. (Il s'agit donc d'un vrai positif).

La **spécificité** d'un test est la probabilité que le test soit négatif sachant que la personne n'a pas été infectée par la maladie. (Il s'agit donc d'un vrai négatif).

Le fabricant du test fournit les caractéristiques suivantes :

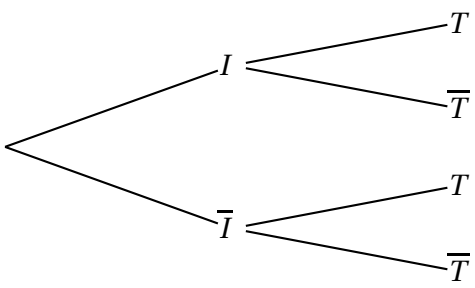
— Sa sensibilité est de 0,8.

— Sa spécificité est de 0,99.

On prélève un individu soumis au test dans la population française adulte au 11 mai 2020.

On note T l'évènement « le test réalisé est positif ».

1. Compléter l'arbre des probabilités ci-dessous avec les données de l'énoncé :



2. Montrer que $p(T) = 0,055\,03$.
3. Quelle est la probabilité qu'un individu ait été infecté sachant que son test est positif ?
   On donnera une valeur approchée à 10^{-4} près du résultat.

Partie C :

On considère un groupe d'une population d'un autre pays soumis au même test de sensibilité 0,8 et de spécificité 0,99.

Dans ce groupe la proportion d'individus ayant un test positif est de 29,44 %.

On choisit au hasard un individu de ce groupe ; quelle est la probabilité qu'il ait été infecté ?

Partie A

1. Puisque l'individu est choisi dans la population française, on suppose qu'il y a une situation d'équiprobabilité et que les proportions sont assimilables à des probabilités.

   Comme 5,7 % des adultes avaient déjà été infectés, d'après l'étude du *Lancet*, la probabilité que l'individu choisi ait déjà été infecté est de $P(I) = 0,057$.

2. (a) — On a une épreuve de Bernoulli, dont le succès : « l'individu choisi a déjà été infecté », a une probabilité $p = 0,057$;
   — Cette épreuve est répétée $N = 100$ fois, de façon identique et indépendante (car le prélèvement des 100 individus est assimilé à un tirage avec remise) ;
   — X est une variable aléatoire qui compte le nombre de succès parmi ces répétitions.

   Avec ces éléments, on peut affirmer que X suit la loi binômiale de paramètres $N = 100$ et $p = 0,057$.

   Remarque : le paramètre donnant le nombre de répétitions est souvent noté n, plutôt que N, mais dans la question **2. e.**, un entier n est introduit, donc pour éviter le conflit de notation, on utilise N ici.

   (b) Puisque X suit une loi binômiale, on a : $E(X) = N \times p = 100 \times 0,057 = 5,7$. On en déduit que dans un échantillon de 100 personnes adultes choisies au sein de la population française le 11 mai 2020, en moyenne, 5,7 d'entre eux avaient déjà été infectés par la COVID 19.

   (c) La probabilité demandée est celle de l'événement $\{X = 0\}$.

   Pour les variables aléatoires régies par la loi binômiale, on a (pour k entier naturel inférieur à n) : $P(X = k) = \binom{N}{k} \times p^k \times (1-p)^{N-k}$.

   On a donc, ici : $P(X = 0) = \binom{100}{0} \times 0,057^0 \times 0,943^{100} = 0,943^{100} \approx 0,0028$.

   (d) La probabilité qu'au moins deux personnes de l'échantillon aient été préalablement infectées est de :

   $P(X \geqslant 2) = 1 - P\left(\overline{X \geqslant 2}\right) = 1 - P(X \leqslant 1) = 1 - P(X = 0) - P(X = 1)$. (Certains modèles de calculatrices n'ont pas besoin de ce calcul).

   À la calculatrice, on obtient : $P(X \geqslant 2) \approx 0,9801$.

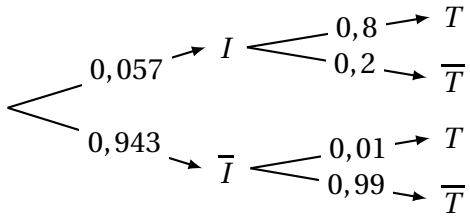
   (e) Par exploration à la calculatrice, on constate que pour $n \leqslant 8$, on a une probabilité inférieure ou égale à 0,9, avec $P(X \leqslant 8) \approx 0,8829$ et $P(X \leqslant 9) \approx 0,9408$.

   L'entier cherché est 9.

   Cela signifie que dans un échantillon de cent adultes choisis dans la population française le 11 mai 2020, il y a plus de neuf chances sur dix que le nombre d'entre eux préalablement infectés par la COVID 19 est inférieur ou égal à 9.

Partie B

1. L'événement I est l'événement déjà utilisé dans la **Partie A**. On avait $P(I) = 0,057$.
   — La description de la **sensibilité** nous fait comprendre qu'il s'agit de la probabilité conditionnelle : $P_I(T) = 0,8$;
   — celle de la **spécificité** indique que est : $P_{\overline{I}}\left(\overline{T}\right) = 0,99$.
   Avec ces informations, on peut compléter l'arbre probabilisé :

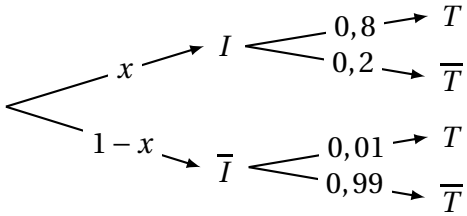


2. Les événements I et $\overline{I}$ partitionnent l'univers, donc, d'après la loi des probabilités totales :
   $P(T) = P\left(I \cap T\right) + P\left(\overline{I} \cap T\right) = 0,057 \times 0,8 + 0,943 \times 0,01 = 0,055\,03$.
3. La question posée est de calculer : $P_T(I)$.
   D'après la définition des probabilités conditionnelles :
   $P_T(I) = \dfrac{P(T \cap I)}{P(T)} = \dfrac{0,057 \times 0,8}{0,055\,03} = \dfrac{4560}{5503} \approx 0,828\,6$.
   Cela signifie qu'une personne dont le test est positif n'a qu'une probabilité de 0,828 6 (environ) d'avoir été préalablement infectée par la COVID 19.

Partie C

Si le test est le même, sa sensibilité et sa spécificité sont les mêmes, ce qui change, c'est donc la probabilité d'avoir été préalablement infecté. Cette probabilité n'est pas connue, notons la x.

On a donc l'arbre suivant :



Les événements I et $\overline{I}$ partitionnent toujours cet univers différent, donc, d'après la loi des probabilités totales :

$P(T) = P\left(I \cap T\right) + P\left(\overline{I} \cap T\right) = x \times 0,8 + (1-x) \times 0,01 = 0,8x + 0,01 - 0,01x = 0,79x + 0,01$.

D'après l'énoncé, 29,44 % des gens ont un test positif, donc la probabilité de choisir un individu dont le test est positif est de 0,2944.

On a donc une double égalité, dont on déduit l'équation suivante : $0,79x + 0,01 = 0,2944$

Résolvons : $0,79x + 0,01 = 0,294\,4 \iff 0,79x = 0,284\,4$

$$\iff x = \frac{0,284\,4}{0,79}$$

$$\iff x = \frac{2\,844}{7\,900}$$

La probabilité que la personne choisie ait été infectée est donc de $\dfrac{2\,943}{7\,900} = 0,36$.

Dans cet autre pays, la proportion de personnes préalablement infectées est donc d'environ 36 %.