Amérique du sud 2015. Enseignement spécifique

EXERCICE 3 (5 points) (commun à tous les candidats)

Partie A

Le chikungunya est une maladie virale transmise d'un être humain à l'autre par les piqûres de moustiques femelles infectées.

Un test a été mis au point pour le dépistage de ce virus. Le laboratoire fabriquant ce test fournit les caractéristiques suivantes :

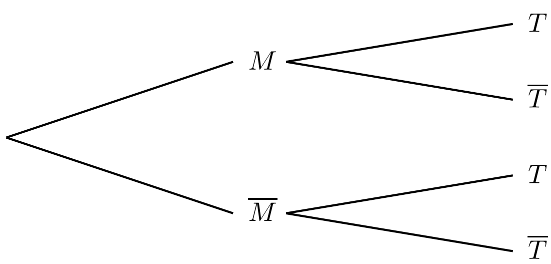
- la probabilité qu'une personne atteinte par le virus ait un test positif est de $0,98$;
- la probabilité qu'une personne non atteinte par le virus ait un test positif est de $0,01$.

On procède à un test de dépistage systématique dans une population « cible ». Un individu est choisi au hasard dans cette population. On appelle :

- M l'événement : « L'individu choisi est atteint du chikungunya »
- T l'événement : « Le test de l'individu choisi est positif »

On notera $\overline{M}$ (respectivement $\overline{T}$) l'événement contraire de l'événement M (respectivement T).
On note p $(0 \leqslant p \leqslant 1)$ la proportion de personnes atteintes par la maladie dans la population cible.

1) a) Recopier et compléter l'arbre de probabilité ci-dessous.



b) Exprimer $P(M \cap T)$, $P\left(\overline{M} \cap T\right)$ puis $P(T)$ en fonction de p.

2) a) Démontrer que la probabilité de M sachant T est donnée par la fonction f définie sur $[0,1]$ par :

$$f(p) = \frac{98p}{97p+1}.$$

b) Etudier les variations de la fonction f.

3) On considère que le test est fiable lorsque la probabilité qu'une personne ayant un test positif soit réellement atteinte du chikungunya est supérieure à $0,95$.
En utilisant les résultats de la question **2.**, à partir de quelle proportion p de malades dans la population le test est-il fiable ?

Partie B

En juillet 2014, l'institut de veille sanitaire d'une île, en s'appuyant sur les données remontées par les médecins, publie que $15\,\%$ de la population est atteinte par le virus.
Comme certaines personnes ne consultent pas forcément leur médecin, on pense que la proportion est en réalité plus importante.
Pour s'en assurer, on se propose d'étudier un échantillon de 1 000 personnes choisies au hasard dans cette île. La population est suffisamment importante pour considérer qu'un tel échantillon résulte de tirages avec remise.
On désigne par X la variable aléatoire qui, à tout échantillon de 1 000 personnes choisies au hasard, fait correspondre le nombre de personnes atteintes par le virus et par F la variable aléatoire donnant la fréquence associée.

1) a) Sous l'hypothèse $p = 0,15$, déterminer la loi de X.

b) Dans un échantillon de 1 000 personnes choisies au hasard dans l'île, on dénombre 197 personnes atteintes par le virus.
Quelle conclusion peut-on tirer de cette observation à propos du chiffre de $15\,\%$ publié par l'institut de veille sanitaire ?
Justifier. (On pourra s'aider du calcul d'un intervalle de fluctuation au seuil de $95\,\%$.)

2) On considère désormais que la valeur de p est inconnue.
En utilisant l'échantillon de la question **1. b.**, proposer un intervalle de confiance de la valeur de p, au niveau de confiance de 95 %.

Partie C

Le temps d'incubation, exprimé en heures, du virus peut être modélisé par une variable aléatoire T suivant une loi normale d'écart type $\sigma = 10$.
On souhaite déterminer sa moyenne μ.
La représentation graphique de la fonction densité de probabilité de T est donnée en annexe.

1) a) Conjecturer, à l'aide du graphique, une valeur approchée de μ.

b) On donne $p(T < 110) = 0,18$. Hachurer sur le graphique un domaine dont l'aire correspond à la probabilité donnée.

2) On note T' la variable aléatoire égale à $\dfrac{T-\mu}{10}$.

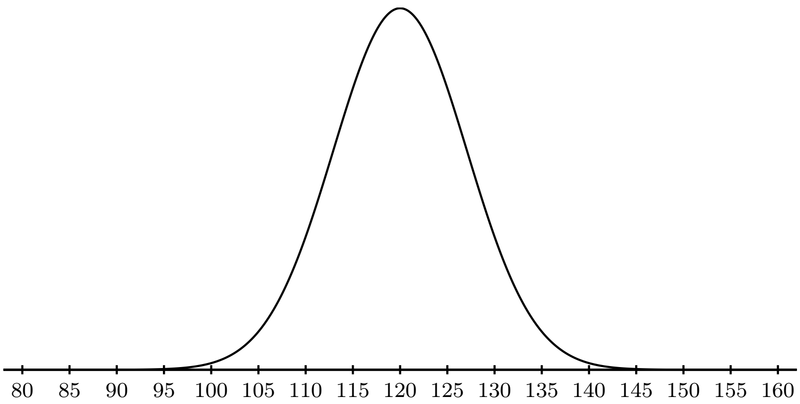
a) Quelle loi la variable aléatoire T' suit-elle ?

b) Déterminer une valeur approchée à l'unité près de la moyenne μ de la variable aléatoire T et vérifier la conjecture de la question 1.

FEUILLE ANNEXE (à rendre avec la copie)

Annexe, Exercice 3

Courbe représentative de la fonction densité de la loi normale $\mathscr{N}\left(\mu, 10^2\right)$

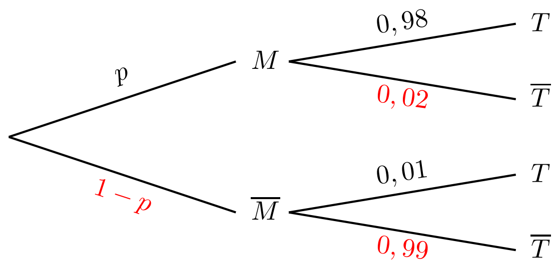


Amérique du Sud 2015. Enseignement spécifique

EXERCICE 3 : corrigé

Partie A

1) a) L'énoncé fournit $P_M(T) = 0,98$, $P_{\overline{M}}(T) = 0,01$ et $P(M) = p$. Représentons la situation par un arbre de probabilité.



b) $P(M \cap T) = P(M) \times P_M(T) = 0,98p$. $P\left(\overline{M} \cap T\right) = P\left(\overline{M}\right) \times P_{\overline{M}}(T) = 0,01(1-p) = 0,01 - 0,01p$. Enfin, d'après la formule des probabilités totales,

$$P(T) = P(M \cap T) + P\left(\overline{M} \cap T\right) = 0,98p + 0,01 - 0,01p = 0,97p + 0,01.$$

2) a)

$$P_T(M) = \frac{P(M \cap T)}{P(T)} = \frac{0,98p}{0,97p + 0,01} = \frac{0,98p \times 100}{(0,97p + 0,01) \times 100} = \frac{98p}{97p + 1}.$$

b) La fonction f est dérivable sur $[0,1]$ en tant que quotient de fonctions dérivables sur $[0,1]$ dont le dénominateur ne s'annule pas sur $[0,1]$ et pour tout réel p de $[0,1]$,

$$f'(p) = \frac{98(97p+1) - 98p \times 97}{(97p+1)^2} = \frac{98}{(97p+1)^2}.$$

La fonction f' est strictement positive sur $[0,1]$ et donc la fonction f est strictement croissante sur $[0,1]$.

3) Le test est fiable quand $P_T(M) \geqslant 0,95$ ou encore $f(p) \geqslant 0,95$.

$$f(p) \geqslant 0,95 \Leftrightarrow \frac{98p}{97p+1} \geqslant 0,95 \Leftrightarrow 98p \geqslant 92,15p + 0,95 \Leftrightarrow 5,85p \geqslant 0,95$$
$$\Leftrightarrow p \geqslant \frac{0,95}{5,85} \Leftrightarrow p \geqslant \frac{95}{585} \Leftrightarrow p \geqslant \frac{19}{117}.$$

Partie B

1) a) La variable aléatoire X est régie par un schéma de Bernoulli. En effet,

- 1000 expériences identiques et indépendantes sont effectuées ;
- chaque expérience a deux issues : « la personne est atteinte par le virus » avec une probabilité $p = 0,15$ et « la personne n'est pas atteinte par le virus » avec une probabilité $1 - p = 0,85$.

La variable aléatoire X suit donc une loi binomiale de paramètres $n = 1000$ et $p = 0,15$. On sait alors que pour tout entier k tel que $0 \leqslant k \leqslant 1000$,

$$P(X = k) = \binom{1000}{k}(0,15)^k(0,85)^{1000-k}.$$

b) On note tout d'abord que $n \geqslant 30$ puis $np = 150$ et donc $np \geqslant 5$ et aussi $n(1-p) = 850$ et donc $n(1-p) \geqslant 5$.

Un intervalle de fluctuation au seuil 95% est

$$\left[p - 1,96\sqrt{\frac{p(1-p)}{n}}, p + 1,96\sqrt{\frac{p(1-p)}{n}}\right] = \left[0,15 - 1,96\sqrt{\frac{0,15 \times (1-0,15)}{1000}}, 0,15 + 1,96\sqrt{\frac{0,15 \times (1-0,15)}{1000}}\right]$$
$$= [0,127; 0,173].$$

en arrondissant de manière à élargir un peu l'intervalle. La fréquence observée dans l'échantillon est $f = \dfrac{197}{1000} = 0,197$. Cette fréquence n'appartient pas à l'intervalle de fluctuation et donc on peut affirmer que le pourcentage de 15% publié par l'institut de veille sanitaire est faux au risque de se tromper de 5%.

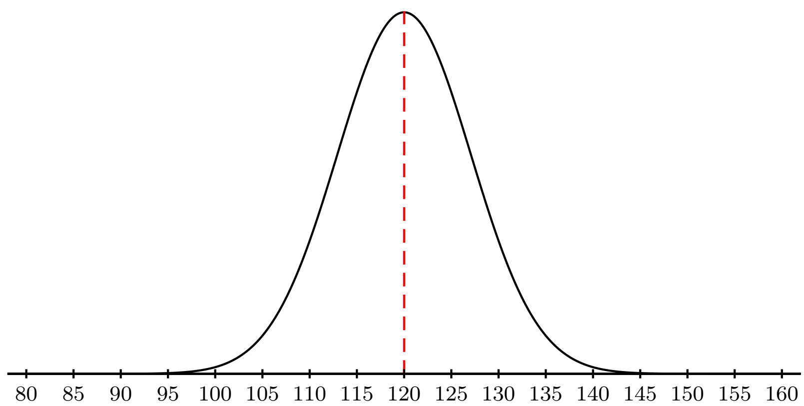
2) Ici, $n = 1000$ et $f = 0,197$. On est dans les conditions d'utilisation d'un intervalle de confiance car $n \geqslant 30$, $nf \geqslant 5$ et $n(1-f) \geqslant 5$. Un intervalle de confiance au niveau de confiance 95% est

$$\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right] = \left[0,197 - \frac{1}{\sqrt{1000}}, 0,197 + \frac{1}{\sqrt{1000}}\right] = [0,165; 0,229]$$

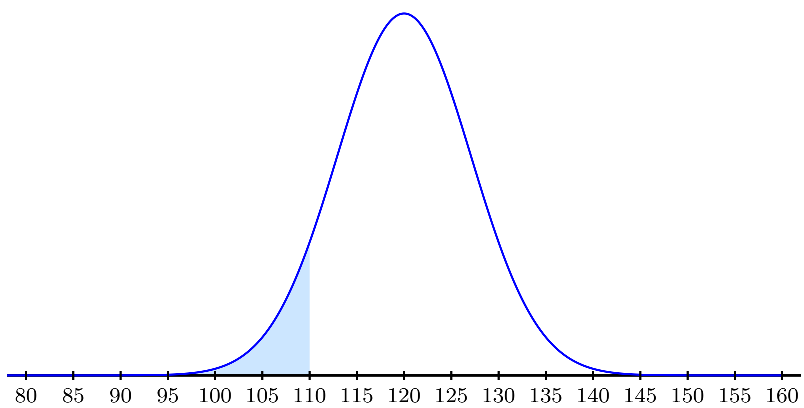
en arrondissant de manière à élargir un peu l'intervalle. La probabilité p appartient à l'intervalle $[0,165; 0,229]$ avec un niveau de confiance de 95%.

Partie C

1) a) Il semble que $\mu = 120$.



b)



2) a) On sait que T' suit la loi normale centrée réduite c'est-à-dire la loi normale de moyenne 0 et d'écart-type 1.

b)

$$T < 110 \Leftrightarrow T - \mu < 110 - \mu \Leftrightarrow \frac{T-\mu}{10} < \frac{110-\mu}{10} \Leftrightarrow$$

et donc $P(T < 110) = P\left(T' < \dfrac{110-\mu}{10}\right) = P\left(T' \leqslant \dfrac{110-\mu}{10}\right)$. La calculatrice fournit

$$P\left(T' \leqslant \frac{110-\mu}{10}\right) = 0,18 \Leftrightarrow \frac{110-\mu}{10} = -0,91\ldots$$

puis $\mu = 110 + 10 \times 0,91\ldots = 119,1\ldots$ à l'unité près par excès. Ceci valide la conjecture de la question 1)a).

