Université Grenoble Alpes - (UJF)
L2 Licence Sciences et Technologies - STA401 **Mars 2024**

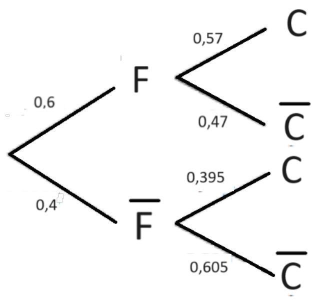
STA401 : CC2-Partiel correction

Exercice 1 :

1.

Evènements	Cancer	Non cancer	Total
Fumeur	342	258	600
Non fumeur	158	242	400
Total	500	500	1000

2. $P(F/C) = P(F \cap C)/P(C) = 0,684$ et $P(C/F) = P(F \cap C)/P(F) = 0,57$
3. $P(F/C) \neq P(F) = 0,6$ et $P(C/F) \neq P(C) = 0,5$ donc les 2 évènements ne sont pas indépendants.
4. Arbre pondéré :



Exercice 2:

1. $\widehat{\sigma}^2 = S'^2 = 0,95238095$ et $var(c(1,2,2,2,3,3,4))$
2. a) $\mathcal{N}(0;1)$
   b) $Z = 3X - 2Y \rightsquigarrow \mathcal{N}(1;145)$ car indépendantes.
   $P(3X - 2Y > 0) = P(\frac{Z-1}{\sqrt{145}} > \frac{-1}{\sqrt{145}}) = P(\frac{Z}{\sqrt{145}} > -0,083045) = P(\frac{Z}{\sqrt{145}} < 0,083045) = 0,532$
3. a) $P(X < 35) = P(Y < \frac{35-30}{8}) = P(Y < 0.625) \approx 0,734$ (entre 0.7324 et 0.7357). Calculatrice : 0.7340145
   b) $P(X > a) = P(Y > \frac{a-30}{8}) \Longleftrightarrow P(Y < \frac{a-30}{8}) = 0.02 \Longleftrightarrow P(Y < -(\frac{a-30}{8})) = 0.98 \Longleftrightarrow -u = 2.0537 \Longleftrightarrow a = 13,5704$
4. Par définition, le risque de première espèce est le risque minimisé dans un test : $\alpha = P_{\mathcal{H}_0}(\mathcal{H}_1)$ =P(accepter $\mathcal{H}_1$ alors qu'en réalité $\mathcal{H}_0$ est vraie).
   Donc α = P (faire une fausse joie à tort) = P (déclarer vainqueur / il n'est pas vainqueur) = $P(\ p > 0,5\ /\ p \leq 0,5\ /) = P_{\mathcal{H}_0}(\mathcal{H}_1)$. Donc : $\begin{cases} H_0 \ : \ p = \ 0.5 \\ H_1 \ : \ p > \ 0.5 \end{cases}$

Exercice 3 : (filières Min-Mat)

1. a) $S = \left\{ \begin{array}{cc} 1 & si\ A\ se\ réalise\ (1/2) \\ 0 & sinon \end{array} \right\}$ suit une loi Bernoulli de paramètre1/2. Donc X est la somme de n variables S_i de Bernoulli indépendantes, donc X suit donc une loi Binomiale de paramètres n et $1/2$. $E(X) = n/2$ et $V(X) = n/4$.
   b) Si n est grand, *le théorème Central Limite* donne une approximation par une loi Gaussienne [n est grand, n>50] : $\mu= n/2$ et $\sigma^2 = n/4$.

2. $F = X/n$; X avec $\mathcal{N}(50; 25)$. On trouve $p \simeq 0.9545$
3. Si p est inconnue, on cherche n pour que $p = P(0,4 < F < 0,6) = P(0,4n < X < 0,6n) > 0,9$
   Inégalité de Bienaymé Tchebychev (X variable réelle positive et a>0) :
   $P(\mid X - E(X) \mid \geq a) \leq \frac{V(X)}{a^2} \quad \Leftrightarrow \quad P(\mid X - n/2 \mid \geq a) \leq \frac{n}{4a^2} \quad \Leftrightarrow \quad P(-a \leq X - n/2 \leq a) \geq 1 - \frac{n}{4a^2} \quad \Leftrightarrow \quad P(n/2 - a \leq X \leq n/2 + a) \geq 1 - \frac{n}{4a^2}$Ici, on veut $n/2 - a = 0,4n$ et $n/2 + a = 0,6n$, et $0,9 = 1 - \frac{n}{4a^2}$. La résolution du système donne $a = 25$, et $n = 10 * a = 250$. Il faut donc lancer au moins 250 fois la pièce.

Exercice 3 : (filière Inm)

1. a) $S = \left\{ \begin{array}{cc} 1 & si\ A\ se\ réalise\ (1/2) \\ 0 & sinon \end{array} \right\}$ suit une loi Bernoulli de paramètre1/2. Donc X est la somme de 100 variables S_i de Bernoulli indépendantes, donc X suit donc une loi Binomiale de paramètres 100 et 1/2. $E(X) = 50$ et $V(X) = 25$. Avec cette loi Binomiale, $P(40 < X < 60) = P(X \leq 59) - P(X \leq 40) \simeq 0,94311$
   b) Si n est grand, *le théorème Central Limite* donne une approximation par une loi Gaussienne [n est grand, n>50] : $\mu= 50$ et $\sigma^2 = 25$. Avec cette nouvelle loi, on obtient : $P(40 < X < 60) \simeq 0,9545$

2. Inégalité de Bienaymé Tchebychev (X variable réelle positive et a>0) :
   $P(\mid X - E(X) \mid \geq a) \leq \frac{V(X)}{a^2} \quad \Leftrightarrow \quad P(\mid X - 16 \mid \geq a) \leq \frac{16}{a^2} \quad \Leftrightarrow \quad P(-a \leq X - 16 \leq a) \geq 1 - \frac{16}{a^2} \quad \Leftrightarrow \quad P(16 - a \leq X \leq 16 + a) \geq 1 - \frac{16}{a^2}$. Ici, on veut $16 - a = 10$ et $16 + a = 22$, soit $a = 6$.
   Donc : $P(10 < X < 22) \geq 1 - 16/36 = 0,5555$.

Exercice 4 :

PARTIE A :

1. Moyenne et variance empirique : $\overline{x} = 4,72$, $s^2 = 1,8716$.
   Moyenne et variance estimée sans biais : $\widehat{\mu} = \overline{x} = 4,72$, $\widehat{\sigma^2} = s'^2 = 2,0795555$.
2. Les paramètres μ et σ^2 du modèle sont inconnues, donc l'intervalle de confiance de μ au niveau de 95% est : $[\overline{x} \pm t^{n-1}_{1-\alpha/2} \frac{s'}{\sqrt{n}}]$. On trouve: $[4,72 \pm 2,2622 * \sqrt{2,079555/10}] \simeq [3,6884; 5,7516]$
3. Intervalle de confiance de la variance, $\alpha = 0,05$: $\left[\frac{ns^2}{z^{(n-1)}_{1-\alpha/2}}; \frac{ns^2}{z^{(n-1)}_{\alpha/2}} \right] \approx \left[\frac{18,716}{19,02}; \frac{18,716}{2,7} \right]$. Donc l'intervalle de l'écart type est : [0,991976 ; 2,63284]
4. On suppose que $\sigma = 1,4$ donc il est connu ! L'intervalle est maintenant : $[\overline{x} \pm u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}]$.
   Pour avoir une précision de l'intervalle de $\pm 0,5$, il faut : $u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} = 0,5 \quad \Rightarrow$
   $\sqrt{n} = 2,5758 * 1,4/0,5 = 7,21224 \ \Rightarrow \ n = 52,016$. Il faut donc un échantillon de taille 53 individus.

PARTIE B :

1. $\overline{X} \rightsquigarrow \mathcal{N}(4,5; 2/100)$. Pour la suite, on note $Y = (\overline{X} - 4.5)/\sqrt{2/100} \rightsquigarrow \mathcal{N}(0;1)$.
2. $P(\overline{X} < 5) = P(Y < \dfrac{5 - 4.5}{\sqrt{2/100}}) = P(Y < 3,5355) \approx 0,999796$ (calculatrice)
3. $P(\mid \overline{X} - 4.5 \mid < 1) = P(-1/\sqrt{2/100} < Y < 1/\sqrt{2/100}) = 2 * P(Y < 1/\sqrt{2/100}) - 1 \simeq 1$
4. On calcule l'intervalle de fluctuation de la moyenne pour vérifier si cet échantillon est représentatif ou pas à 98% : $I = [\mu \pm u_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}}] = [4.5 \pm 2,3263\dfrac{\sqrt{2}}{\sqrt{100}}] =]4.1710; 4.82899[$. De plus $4,7 \in I$, on conclut que pour un niveau de 98%, cet échantillon a une moyenne représentative de la loi de cette population.