Exercice 1

1. Une estimation ponctuelle de p est $\widehat{p} = f_e = \dfrac{13}{215} \approx 6\%$, la fréquence observée par le groupe de vignerons.

2. On prend $\alpha = 1\%$ et $n = 215$. On sait que si $(X_1, X_2, \ldots, X_n)$ est un échantillon de la loi de Bernoulli de paramètre p et si $F = \dfrac{1}{n}\sum_{k=1}^{n} X_k$, alors, asymptotiquement

$$U = \frac{F-p}{\sqrt{p(1-p)/n}} \hookrightarrow \mathcal{N}(0,1).$$

Soit alors t_α le nombre réel tel que

$P(-t_\alpha \leqslant U \leqslant t_\alpha) = 1-\alpha$ c'est-à-dire $P(U \leqslant t_\alpha) = 1 - \dfrac{\alpha}{2} = 0,995$

En cherchant dans la table de la loi normale centrée réduite, on lit $t_\alpha \simeq 2,576$.
Ainsi:

$$\begin{aligned} 0,99 &\simeq P\left(-t_\alpha \leqslant \frac{F-p}{\sqrt{p(1-p)/n}} \leqslant t_\alpha\right) \\ &= P\left(F - t_\alpha\sqrt{\frac{p(1-p)}{n}} \leqslant p \leqslant F + t_\alpha\sqrt{\frac{p(1-p)}{n}}\right) \end{aligned}$$

On en déduit **un** intervalle de confiance observé pour p au niveau 99% :

$$\boxed{I_{99\%,\text{obs}} = \left[f_e - t_\alpha\sqrt{\frac{f_e(1-f_e)}{n}}\ ,\ f_e + t_\alpha\sqrt{\frac{f_e(1-f_e)}{n}}\right] \simeq [1,8\%\ ;\ 10,2\%].}$$

3. On pose $\alpha = 1\%$, $n = 215$ et $p_0 = 0,03$. On souhaite tester :

$$\mathcal{H}_0\ :\ p = p_0 \qquad \text{contre} \qquad \mathcal{H}_1\ :\ p > p_0.$$

On pose $Z = \dfrac{F-p_0}{\sqrt{p_0(1-p_0)/n}}$. Puisque $n > 50$, $np_0 \geqslant 13$ et $np_0(1-p_0) > 10$, on peut approcher la loi de Z par la loi normale centrée réduite $\mathcal{N}(0,1)$.

On cherche $z_\alpha \in \mathbb{R}$ tel que $P(Z \leqslant z_\alpha) = 1 - \alpha = 0,99$. En regardant dans la table de la loi normale, on lit $z_\alpha \simeq 2,33$.

ce qui donne le domaine d'acceptation :

$$D_0 = \left[0\ ;\ p_0 + z_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}\right] \simeq [0\ ;\ 5,8\%]$$

• Prise de décision : ici $\widehat{p} \approx 6\%$. $\quad \widehat{p} \notin D_0$. **On décide de rejeter** l'hypothèse $\mathcal{H}_0$ avec un risque d'erreur de 1%.

Remarque : dans le cas d'un test bilatéral, on aurait testé

$$\mathcal{H}_0\ :\ p = p_0 \qquad \text{contre} \qquad \mathcal{H}_1\ :\ p \neq p_0.$$

On garderait la même variable de test Z mais on chercherait le réel z_α tel que $P(Z \leqslant z_\alpha) = 1 - \dfrac{\alpha}{2} = 0,995$. On obtiendrait $z_\alpha = \Phi^{-1}(0,995) \simeq 2,576$ et le domaine d'acceptation :

$$D_0 = \left[p_0 - z_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}\ ,\ p_0 + z_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}\right] \simeq [0,003\%\ ;\ 5,997\%] \simeq [0\%\ ;\ 6\%]$$

Pour l'échantillon considéré, $\widehat{p} \approx 6,05\%$. De nouveau, $\quad \widehat{p} \notin D_0$.
On rejetterait encore l'hypothèse $\mathcal{H}_0$ avec un risque d'erreur de 1%.

Exercice 2

1. Dans tout l'exercice, $n = 11$. On pose $x_1 = 1,71$; $x_2 = 0,89$; $\ldots$; $x_{11} = -0,19$.

   (a) Estimation ponctuelle de μ : $\quad \widehat{\mu} = \overline{x_e} = \dfrac{1}{n}\sum_{i=1}^{n} x_i \approx 0,676$

   (b) Estimation ponctuelle de σ^2 :

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x_e})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2\right) - \frac{n}{n-1}\overline{x_e}^2 \approx 0,860$$

   (c) Dans cette question, $\alpha = 5\%$. $\quad (X_1, X_2, \ldots, X_n)$ est un échantillon de taille n de la variable parente X où $X \hookrightarrow \mathcal{N}(\mu, \sigma)$.

   La moyenne empirique d'échantillon est: $\overline{X} = \dfrac{1}{n}\sum_{k=1}^{n} X_k$.

   On sait que $\overline{X} \hookrightarrow \mathcal{N}\left(\mu\ ,\ \dfrac{\sigma}{\sqrt{n}}\right)$

La variance empirique «corrigée» est : $S^2 = \dfrac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2$.

On sait que la variable aléatoire $T = \dfrac{\overline{X}-\mu}{S/\sqrt{n}}$ suit une loi de Student $\mathrm{T}_{(n-1)}$ à $n-1$ degrés de liberté.

On cherche donc le réel t_α tel que $P(-t_\alpha \leqslant T \leqslant t_\alpha) = 1-\alpha$. En notant F la fonction de répartition de la loi de Student à $n-1$ degrés de liberté, on obtient :

$$F(t_\alpha) - F(-t_\alpha) = 1-\alpha \quad \text{d'où} \quad 2F(t_\alpha) - 1 = 1-\alpha \quad \text{puis} \quad F(t_\alpha) = 1 - \frac{\alpha}{2} = 0,975$$

La table de la loi de Student donne : $t_\alpha \simeq 2,228$.

Donc $0,95 = P\left(-t_\alpha \leqslant \dfrac{\overline{X}-\mu}{S/\sqrt{n}} \leqslant t_\alpha\right) = P\left(\overline{X} - t_\alpha \dfrac{S}{\sqrt{n}} \leqslant \mu \leqslant \overline{X} + t_\alpha \dfrac{S}{\sqrt{n}}\right)$

On en déduit un intervalle de confiance aléatoire à 95% pour μ :

$$I_{95\%} = \left[\overline{X} - t_\alpha \frac{S}{\sqrt{n}} \ ,\ \overline{X} + t_\alpha \frac{S}{\sqrt{n}}\right]$$

puis un intervalle de confiance observé à 95% pour μ :

$$\boxed{I_{95\%\,,\,obs} = \left[\overline{x_e} - t_\alpha \frac{\widehat{\sigma}}{\sqrt{n}} \ ;\ \overline{x_e} + t_\alpha \frac{\widehat{\sigma}}{\sqrt{n}}\right] \simeq [0.053\ ,\ 1.300]}$$

(d) Dans cette question, $\alpha = 1\%$. La moyenne étant inconnue, elle est estimée à partir des données. On utilise la variable aléatoire

$$Z = \frac{(n-1)S^2}{\sigma^2} = \sum_{k=1}^{n}\left(\frac{X_k - \overline{X}}{\sigma}\right)^2$$

qui suit la loi du χ^2_{n-1} à $n-1$ degrés de liberté.

On cherche des réels positifs c et d tels que $P(c \leqslant Z \leqslant d) = 1-\alpha$.

Contrairement aux lois normale et de Student, la densité de la loi du χ^2 n'est pas une fonction paire. Mais on a l'habitude de répartir le risque de façon symétrique en prenant :

$$P(Z \leqslant c) = \frac{\alpha}{2} = 0,005 \quad \text{et} \quad P(Z \leqslant d) = 1 - \frac{\alpha}{2} = 0,995$$

La table des lois du χ^2 donne : $c \simeq 2,156$ et $d \simeq 25,188$

Donc $0,99 = P\left(c \leqslant \dfrac{(n-1)S^2}{\sigma^2} \leqslant d\right) = P\left(\dfrac{(n-1)\,S^2}{d} \leqslant \sigma^2 \leqslant \dfrac{(n-1)\,S^2}{c}\right)$

On en déduit un intervalle de confiance aléatoire à 99% pour σ^2 :

$$J_{99\%} = \left[\frac{(n-1)\,S^2}{d}\ ,\ \frac{(n-1)\,S^2}{c}\right]$$

puis un intervalle de confiance observé à 99% pour σ^2 :

$$\boxed{J_{90\%\,,\,obs} = \left[\frac{(n-1)\,\widehat{\sigma}^2}{d}\ ;\ \frac{(n-1)\,\widehat{\sigma}^2}{c}\right] \simeq [0.341\ ;\ 3.990]}$$

2. Notons `randn()` l'appel à la fonction générant des nombres distribués suivant la loi $\mathcal{N}(0,1)$. À la place de Bob, nous aurions pu écrire une nouvelle fonction retournant $\sigma\times$`randn()`$+\mu$. En effet, d'après les propriétés de la loi normale, si $X \hookrightarrow \mathcal{N}(0,1)$, alors

$$\sigma X + \mu \hookrightarrow \mathcal{N}(\mu, \sigma).$$

Exercice 3

1. (a) Une densité est la fonction f définie sur $\mathbb{R}$ par

$$f(t) = \begin{cases} \lambda\,\mathrm{e}^{-\lambda\,t} & \text{si} \quad t \geqslant 0 \\ 0 & \text{si} \quad t < 0 \end{cases}$$

Sa fonction de répartition est F définie sur $\mathbb{R}$ par $F(x) = 0$ si $x < 0$ et $F(x) = \displaystyle\int_0^x \lambda\,\mathrm{e}^{-\lambda\,t}\,\mathrm{d}t = \left[-\mathrm{e}^{-\lambda\,t}\right]_0^x = 1 - \mathrm{e}^{-\lambda\,x}$ pour $x \geqslant 0$

(b) La durée moyenne de fonctionnement entre deux pannes consécutives, est l'espérance mathématique de X (ou de Y), à savoir : $E(X) = E(Y) = \dfrac{1}{\lambda}$

2. Notons A l'événement : «chacune des 2 périodes de fonctionnement de la machine dure plus de 3 années.»

Alors $A = [X > 3] \cap [Y > 3]$.

D'où $P(A) = P\big([X>3]\cap[Y>3]\big) = P(X>3)\,P(Y>3)$ car X et Y sont indépendantes par hypothèse.

Donc $P(A) = (1 - F(3))^2 = \big(1 - (1 - \mathrm{e}^{-3\lambda})\big)^2 = \big(\mathrm{e}^{-3\lambda}\big)^2$.

Ainsi $\boxed{P(A) = \mathrm{e}^{-6\,\lambda}}$

3. On a posé $S = X + Y$.

   (a) • Par linéarité de l'espérance,
   $E(S) = E(X+Y) = E(X) + E(Y) = 2\,E(X) = \dfrac{2}{\lambda}$.
   • Comme X et Y sont indépendantes,
   $V(S) = V(X+Y) = V(X) + V(Y) = 2\,V(X) = \dfrac{2}{\lambda^2}$

   (b) Comme les variables aléatoires X et Y sont indépendantes, leur somme S est une variable à densité, dont une densité est obtenue par convolution :

$$\begin{aligned}\forall\, x \in \mathbb{R}, \quad h(x) &= \int_{-\infty}^{+\infty} f_X(t)\, f_Y(x-t)\,\mathrm{d}t \\ &= \int_{-\infty}^{+\infty} f(t)\, f(x-t)\,\mathrm{d}t \\ &= \int_{0}^{+\infty} f(t)\, f(x-t)\,\mathrm{d}t\end{aligned}$$

   Si $x < 0$ alors $\forall\, t \in \mathbb{R}^+,\ f(x-t) = 0$ donc $h(x) = 0$.

   Supposons $x \geqslant 0$. Alors pour tout réel $t > x,\ x - t < 0 \Longrightarrow f(t) = 0$.

$$\begin{aligned}\text{Donc } h(x) &= \int_0^x f(t)\, f(x-t)\,\mathrm{d}t \\ &= \int_0^x \lambda\,\mathrm{e}^{-\lambda\, t}\,\lambda\,\mathrm{e}^{-\lambda\,(x-t)}\,\mathrm{d}t \\ &= \int_0^x \lambda^2\,\mathrm{e}^{-\lambda\, x}\,\mathrm{d}t \\ &= \lambda^2\,\mathrm{e}^{-\lambda\, x}\int_0^x \mathrm{d}t \\ &= \lambda^2\,\mathrm{e}^{-\lambda\, x} \times x \\ &= \lambda^2\, x\,\mathrm{e}^{-\lambda\, x}\end{aligned}$$

   (c) Soit x un réel positif fixé. Calculons d'abord $P(S \leqslant x)$.

$$P(S \leqslant x) = \int_0^x h(t)\,\mathrm{d}t = \int_0^x \lambda^2\, t\,\mathrm{e}^{-\lambda\, t}\,\mathrm{d}t = \lambda\int_0^x t\,\left(\lambda\mathrm{e}^{-\lambda\, t}\right)\,\mathrm{d}t$$

   On effectue une **intégration par parties** en posant :

$$\begin{cases} u(t) &= t \\ v'(t) &= \lambda\mathrm{e}^{-\lambda\, t}\end{cases} \quad \text{d'où} \quad \begin{cases} u'(t) &= 1 \\ v(t) &= -\mathrm{e}^{-\lambda\, t}\end{cases}$$

   Comme les fonctions u et v sont de classe $\mathcal{C}^1$ sur $\mathbb{R}$, il vient

$$P(S \leqslant x) = \lambda\int_0^x u(t)\, v'(t)\,\mathrm{d}t = \lambda\left(\left[-t\,\mathrm{e}^{-\lambda\, t}\right]_0^x - \int_0^x -\mathrm{e}^{-\lambda\, t}\,\mathrm{d}t\right)$$

$$= \lambda\left(-x\,\mathrm{e}^{-\lambda\, x} - \left[\frac{\mathrm{e}^{-\lambda\, t}}{\lambda}\right]_0^x\right) = \lambda\left(-x\,\mathrm{e}^{-\lambda\, x} - \frac{\mathrm{e}^{-\lambda\, x}}{\lambda} + \frac{1}{\lambda}\right)$$

   $= 1 - \mathrm{e}^{-\lambda\, x}\,(1 + \lambda\, x)$. On en déduit que $P(S > x) = 1 - P(S \leqslant x)$

$$\boxed{P(S > x) = \mathrm{e}^{-\lambda\, x}\,(1 + \lambda\, x)}$$

4. APPLICATION NUMÉRIQUE : la durée moyenne de fonctionnement de la machine est : $E(S) = \dfrac{2}{\lambda} = 4$. D'où $\lambda = \dfrac{1}{2}$.

   L'énoncé demande la probabilité $P(S > 6)$. D'après le résultat obtenu en 3.(c),

$$P(S > 6) = \mathrm{e}^{-6\lambda}\,(1 + 6\lambda) = 4\,\mathrm{e}^{-3} \approx 0,20$$

   On pourra remarquer que $([X > 3] \cap [Y > 3]) \subset [S > 6]$

   et que $P\big([X > 3] \cap [Y > 3]\big) < P(S > 6)$