Probabilités – 3 IF

TD 4 : Variables aléatoires continues

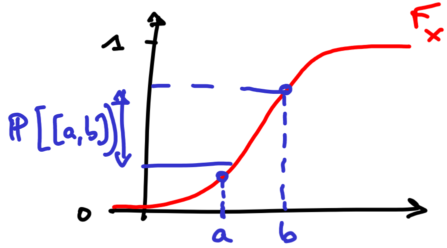
Rappel de cours

On considère l'univers Ω. avec une distribution de probabilité $\mathbb{P}$. Une **variable aléatoire continue** X sur Ω est une fonction $X : \Omega \to \mathbb{R}$ telle que $X(\Omega)$ est indénombrable. Dans ce cas, on ne peut pas définir $\mathbb{P}[X = x]$.

Exemple : Quelle est la probabilité qu'un individu pris au hasard mesure 1,735 m ?

On définit la distribution de probabilité sur des intervalles, grâce à la **fonction de répartition**.

$$\begin{aligned} F_X : \mathbb{R} &\to [0,1] \\ x &\mapsto \mathbb{P}[X \leq x] \end{aligned}$$



On a donc :

$$\mathbb{P}[X \in [a,b]] = F_X(b) - F_X(a)$$

Si la fonction de répartition est dérivable, on dit qu'elle admet une **densité**, définie comme:

$$\forall x, f_X(x) = F'_X(x)$$

Elle peut être vue comme un équivalent de $\mathbb{P}[X = x]$, pour un intervalle de largeur infinitésimal $\mathrm{d}x$:

$$f_X(x) = \lim_{\mathrm{d}x \to 0} \frac{F_X(x + \mathrm{d}x) - F_X(x)}{\mathrm{d}x} = \lim_{\mathrm{d}x \to 0} \frac{\mathbb{P}[X \in [x, x + \mathrm{d}x]]}{\mathrm{d}x}$$

C'est cette densité qui nous permet de redéfinir l'espérance :

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \cdot f_X(x) \,\mathrm{d}x$$

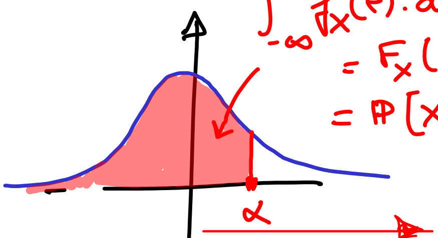
Une des lois continues les plus utilisées est la **loi normale**, d'espérance μ et de variance σ, notée $\mathcal{N}(\mu, \sigma)$, et de densité

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, on peut poser $Y = \frac{X-\mu}{\sigma}$ et on a $Y \sim \mathcal{N}(0,1)$. C'est ce qu'on appelle la **loi normale centrée réduite**. On a donc :

$$\begin{aligned} \mathbb{P}[X \leq x] &= \mathbb{P}[\sigma Y + \mu \leq x] \\ &= \mathbb{P}\left[Y \leq \frac{x-\mu}{\sigma}\right] \\ &\boxed{= F_Y\left(\frac{x-\mu}{\sigma}\right)} \end{aligned}$$

$$\int_{-\infty}^{\alpha} f_X(t).\mathrm{d}t = F_X(\alpha) = \mathbb{P}[X \leq \alpha]$$



Il suffit donc de connaître les valeurs de la fonction de répartition de la loi normale centrée réduite pour en déduire celle de n'importe quelle loi normale.

$$F_Y(-\alpha) = 1 - F_Y(\alpha)$$

α	0.01	0.02	0.03	...
0.0				
0.1				
0.2			F(0.23)	
⋮				

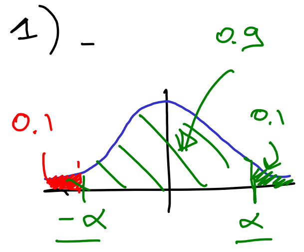
Exercice 1 :

La note obtenue par des étudiants à un examen est une v.a.r. normale $X \sim \mathcal{N}(7, 3^2)$.

1. Calculer le pourcentage d'individus ayant plus de 10, et la note en dessous de laquelle se trouvent 10 % des étudiants.

- $\mathbb{P}[X > 10] = 1 - \mathbb{P}[X \leq 10] = 1 - \mathbb{P}\left[Y \leq \frac{10-7}{3}\right] = 1 - F_Y(1) = \underline{0.1587}$

On pose $Y = \frac{X-7}{3}$ donc $Y \sim \mathcal{N}(0,1)$.



- On cherche k tel que $\mathbb{P}[X \leq k] = 0.1$

donc $\mathbb{P}\left[Y \leq \frac{k-7}{3}\right] = 0.1$ et $F_Y\left(\frac{k-7}{3}\right) = 0.1$

donc $F_Y\left(\frac{7-k}{3}\right) = 0.9$ et $\frac{7-k}{3} = 1.28$ et $\underline{k = 3{,}16}$

2. Compte tenu de ces résultats, on décide de revaloriser l'ensemble des notes par une transformation linéaire $Z = aX + b$. Quelles valeurs doit-on donner à a et b pour que les valeurs précédentes passent respectivement à 50 % et 7 ? (Indication : calculer $\mathbb{E}(Z)$ et $Var(Z)$ en fonction de $\mathbb{E}(X)$ et $Var(X)$).

- $E[Z] = a\mathbb{E}[X] + b = 7a + b \quad (1) \longleftarrow$ par linéarité de l'espérance
- $Var(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2$
$= a^2\mathbb{E}[X^2] + 2ab\,\mathbb{E}[X] + b^2 - (a\mathbb{E}[X] + b)^2$
$= a^2 Var(X) = (3a)^2 \quad (2).$

Avec les valeurs de l'énoncé : $(S) : \begin{cases} \mathbb{E}[Z] = 10 \\ \mathbb{P}[Z \leq 7] = 0.1 \end{cases} \iff \frac{7 - \mathbb{E}[Z]}{\sqrt{Var(Z)}} = -1.28$

$(S) \iff \begin{cases} 7a + b = 10 \\ \frac{7 - 7a - b}{3a} = -1.28 \end{cases} \iff \begin{cases} 7a + b = 10 \\ \frac{7 - 10}{3a} = -1.28 \end{cases} \iff \boxed{\begin{cases} b = 4.53 \\ a = \frac{1}{1.28} = 0.78 \end{cases}}$

Exercice 2 :

Montrer que si $X \sim \mathcal{E}(\lambda)$, alors pour tous réels positifs a et b, on a

$$\mathbb{P}(X > a + b | X > b) = \mathbb{P}(X > a)$$

$X \sim \mathcal{E}(\lambda)$: $f_X(t) : \begin{cases} 0 & \text{si } t < 0 \\ \frac{1}{\lambda} e^{-\frac{t}{\lambda}} & \text{sinon.} \end{cases}$

$\mathbb{P}[X \leq t] = F_X(t) = \int_0^t f_X(x)\,dx = \frac{1}{\lambda}\int_0^t e^{-\frac{x}{\lambda}}\,dx = -\left[e^{-\frac{x}{\lambda}}\right]_0^t = 1 - e^{-\frac{t}{\lambda}}$

$\mathbb{P}[X > a+b \mid X > b] = \frac{\mathbb{P}[X > a+b]}{\mathbb{P}[X > b]} = \frac{1 - F_X(a+b)}{1 - F_X(b)} = \frac{e^{-\frac{(a+b)}{\lambda}}}{e^{-\frac{b}{\lambda}}} = e^{-\frac{a}{\lambda}} = \underline{\mathbb{P}[X > a]}$

Si X représente par exemple la durée de vie d'une imprimante, que signifie cette propriété ?

La durée de vie déjà écoulée n'influe pas sur le temps restant à vivre → pas de vieillissement

Exercice 3 : (DS 2018)

On souhaite modéliser les revenus salariés dans les pays européens. Soit X le salaire des individus et r le revenu minimum fixé par le pays. On suppose que X suit la loi de Pareto $P(a, r)$, de densité

$$f(x) = ar^a x^{-(a+1)} \quad \text{si } x > r$$
$$0 \quad \text{sinon.}$$

1. Quelle est l'espérance de X ?

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x \cdot f_X(x) \, dx = \int_r^{+\infty} x \, f_X(x) \, dx$$

$$= \int_r^{+\infty} x \times a r^a x^{-a-1} \, dx = \int_r^{+\infty} a r^a x^{-a} \, dx \qquad \left(\text{définie ssi } a > 1\right)$$

$$= a r^a \left[\frac{1}{1-a} x^{1-a}\right]_r^{+\infty} = \frac{a}{1-a} r^a \left(0 - r^{1-a}\right) = \boxed{\frac{ar}{a-1}}$$

2.Donner la fonction de répartition de X, notée F_X. Montrer que $U = (\frac{r}{X})^a$ suit une loi uniforme.

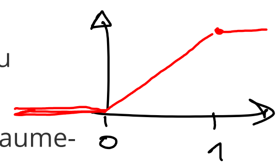
$$F_X(x) = \int_r^x f_X(t) \, dt = \int_r^x a r^a t^{-(a+1)} \, dt = \cancel{a} r^a \left[\frac{-1}{\cancel{a}} t^{-a}\right]_r^x$$

(pour $x \geq r$) $\quad = r^a \left[r^{-a} - x^{-a}\right] = 1 - \left(\frac{r}{x}\right)^a$.

Soit $U = \left(\frac{r}{X}\right)^a$

$F_U(u) = \mathbb{P}[U \leq u] = \mathbb{P}\left[\left(\frac{r}{X}\right)^a \leq u\right]$ si $u \leq 0 \to$ la proba est nulle. on suppose $u > 0$.

$$= \mathbb{P}\left[X \geq \frac{r}{\sqrt[a]{u}}\right] = 1 - \underbrace{F_X\left(\frac{r}{\sqrt[a]{u}}\right)}_{\text{si } \sqrt[a]{u} < 1 \text{ i.e. } u < 1} = 1 - \left[1 - \left(r \times \frac{\sqrt[a]{u}}{r}\right)^a\right] = u$$



3. D_9 correspond au revenu minimal des 10% les plus riches et D_1 correspond au revenu maximal des 10% les plus pauvres. Ces quantités sont données par $F_X(D_9) = 0.9$ et $F_X(D_1) = 0.1$. En 2014, le rapport D_9/D_1 observé vaut 2.81 en France et 3.56 au Royaume-Uni (données OCDE sur les gains bruts, https://www.oecd-ilibrary.org/employment/data/gains/rapport-inter-decile-des-gains-bruts_data-00302-fr).

   Calculer le rapport D_9/D_1 pour $r = 1$, en fonction de a. Pour quelle valeur de a a-t-on $D_9/D_1 = 2.8$? Pour quelle valeur de a a-t-on $D_9/D_1 = 3.56$? Comment interpréter le paramètre a ?

Pour $r = 1$, $\forall x > 1$, $F_X(x) = 1 - \left(\frac{1}{x}\right)^a$

Donc $F_X(D_9) = 0.9 \iff 1 - \left(\frac{1}{D_9}\right)^a = 0.9$

$$\iff \left(\frac{1}{D_9}\right)^a = 0.1$$

$$\iff D_9 = \sqrt[a]{10}$$

De la même manière, $F_X(D_1) = 0.1 \iff D_1 = \sqrt[a]{\frac{9}{10}}$

D'où $\frac{D_9}{D_1} = \sqrt[a]{9}$

Ainsi, $\frac{D_9}{D_1} = 2.8 \Leftrightarrow \sqrt[a]{9} = 2.8 \Leftrightarrow a \ln(2.8) = \ln(9)$

$\Leftrightarrow a = 2.13$.

et $\frac{D_9}{D_1} = 3.56 \Leftrightarrow a \ln(3.56) = \ln(9) \Leftrightarrow a = 1.73$

a permet donc de contrôler la dispersion de la distribution de probabilité. Plus il est petit et plus le rapport est élevé entre les 10 % de valeurs les plus faibles et les 10 % de valeurs les plus élevées.