Université de Montpellier **HLMA408**
2020/2021 **Joseph Salmon**

TD n° 1 : Introduction et rappels

Exercice 1. Croissance exponentielle (intérêt)

Un jeune salarié cherche à constituer une épargne pour un achat immobilier 10 ans plus tard. Deux banques lui proposent chacune un dispositif d'épargne qui bloque l'épargne sur cette période.

La Banque du Sud propose un taux d'intéret constant de 1% par an sur 10 ans. La Banque de l'Ouest propose un taux d'intérêt de 2% pendant 5 ans puis de 0% pendant les 5 ans qui suivent.

a) Quelle banque conseilleriez-vous ?

b) Quel taux d'intérêt la banque du Sud devrait-elle choisir pour rémunérer son client autant que la banque de l'Ouest au bout de 10 ans ?

Correction:

a) $(1.01)^{10} = 1.1046$ (Banque du Sud) ; $(1.02)^5 = 1.1040$ (Banque de l'Ouest) ;

b) $(1.02)^5 = x^{10} \implies 10\log(x) = 5\log(1.02) + 5\log(1) \implies$
$x = \exp\left(\frac{5\log(1.02)+5\log(1)}{10}\right)$
$x = \text{Moy. Geom}(1.02, 1.02, 1.02, 1.02, 1.02, 1, 1, 1, 1, 1)$
Application numérique : 0.995% par an.

Exercice 2. Croissance exponentielle (virus)

Deux variants d'un virus existe. Le premier (Virus1) se reproduit et sa population augmente de 20% tous les ans. Le second (Virus2) se reproduit avec un taux de croissance de 140% une année sur 5, et suit ensuite une décroissance de 1% les 4 années qui suivent.

— Quel est le variant qui se répandera le plus rapidement au bout de 5 ans ?

— Pour quel taux de décroissance le Virus2 aura la même croissance au bout de 5 ans que le Virus1 ?

Correction:

— $1.2^5 = 2.488 =$ et $2.4 \times 0.99^4 = 2.305$

— $(2.4) * x^4 = 1.2^5 \implies x = (\frac{1.2^5}{2.4})^{1/4} = 1.0090756983$, soit un taux de croissance de 0.90756983

Exercice 3. Moyenne harmonique Un oiseau migrateur part de La Havanna (Cuba) pour Toronto (Canada) pour passer son été au frais. Il revient ensuite à Cuba pour passer son hivers au chaud. À l'aller sa vitesse il vole à la vitesse $v_1 = 100\text{km/h}$. Au retour, porté par les vents favorables, il revient avec une vitesse $v_2 = 200\text{km/h}$.

a) Quelle est v la vitesse moyenne de l'oiseau sur son trajet aller-retour ?

b) (Bonus) Même question si l'on rajoute une étape et que le voyage devient : "La Havanna → Toronto → Lubbock (USA)→ La Havanna" (on suppose les 3 villes équidistantes), et que la vitesse entre La Havanna et Toronto est $v_1 = 100km/h$, que celle entre Toronto et Lubbock est $v_2 = 300km/h$, et que celle entre Lubbock et La Havanna est $v_2 = 200km/h$.

Correction:

1) L : distance entre les deux villes.

$$v_1 \times d_1 = L = v_2 \times d_2 \text{ (égalité des distances)} \quad (1)$$
$$v \times (d_1 + d_2) = 2L \text{ (égalité des distances)} \quad (2)$$
$$v \times (d_1 + d_2) = v_1 \times d_1 + v_2 \times d_2 \text{ (égalité des distance)} \quad (3)$$
$$v = \frac{v_1 \times \frac{L}{v_1} + v_2 \times \frac{L}{v_2}}{\frac{L}{v_1} + \frac{L}{v_2}} \quad (4)$$

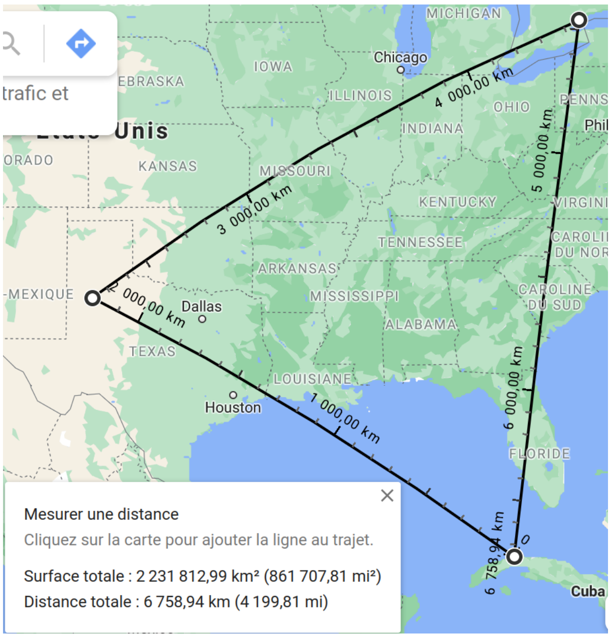

FIGURE 1 – Visualisation du voyage : La Havanna → Toronto → Lubbock → La Havanna.

$$v = \frac{2}{\frac{1}{v_1} + \frac{1}{v_2}} \tag{5}$$

2)

$$\begin{align}
v_1 \times d_1 = L = v_2 \times d_2 = v_3 \times d_3 \text{ (égalité des distances)} \tag{6}\\
v \times (d_1 + d_2 + d_3) = 3L \text{ (égalité des distances)} \tag{7}\\
v \times (d_1 + d_2 + d_3) = v_1 \times d_1 + v_2 \times d_2 + v_3 \times d_3 \tag{8}\\
v = \frac{v_1 \times \frac{L}{v_1} + v_2 \times \frac{L}{v_2} + v_3 \times \frac{L}{v_3}}{\frac{L}{v_1} + \frac{L}{v_2} + \frac{L}{v_3}} \tag{9}\\
v = \frac{3}{\frac{1}{v_1} + \frac{1}{v_2} + \frac{1}{v_3}} \tag{10}
\end{align}$$

Pour des distances L_1, L_2, et L_3 entre les villes cela donnerait :

$$\begin{align}
v_1 \times d_1 = L_1,\ v_2 \times d_2 = L_2,\ v_3 \times d_3 = L_3 \text{ (égalité des distances)} \tag{11}\\
v \times (d_1 + d_2 + d_3) = (L_1 + L_2 + L_3) \text{ (égalité des distances)} \tag{12}\\
v \times (d_1 + d_2 + d_3) = v_1 \times d_1 + v_2 \times d_2 + v_3 \times d_3 \tag{13}\\
v = \frac{v_1 \times \frac{L_1}{v_1} + v_2 \times \frac{L_2}{v_2} + v_3 \times \frac{L_3}{v_3}}{\frac{L_1}{v_1} + \frac{L_2}{v_2} + \frac{L_3}{v_3}} \tag{14}\\
v = \frac{L_1 + L_2 + L_3}{\frac{L_1}{v_1} + \frac{L_2}{v_2} + \frac{L_3}{v_3}} \tag{15}
\end{align}$$

Rappel sur les lois normales. On note Φ (resp. φ) la fonction de répartition (resp. la densité) d'une variable aléatoire X qui suit une loi normale centrée réduite :

$$\varphi(t) := \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \quad \text{et} \quad \Phi(x) = \int_{-\infty}^{x} \varphi(t)dt \ .$$

En particulier, on notera que $\Phi(0) = 1/2$ (symétrie) et que $\lim_{x\to+\infty} \Phi(x) = 1$ et $\lim_{x\to-\infty} \Phi(x) = 0$

Exercice 4. Centrer / Réduire

On rappelle que, si $Z = \sigma X + \mu$ pour $X \sim \mathcal{N}(0,1)$ (*i.e.*, X suit une loi normale centrée réduite), alors Z suit une loi normale $\mathcal{N}(\mu, \sigma^2)$ d'espérance μ et de variance σ^2. Donner la fonction de répartition $\Phi_{\mu,\sigma}$ et la densité $\varphi_{\mu,\sigma}$ de la variable Z. Donner également le lien entre $\varphi_{\mu,\sigma}$ et $\Phi_{\mu,\sigma}$.

Correction:

On a vu en cours l'expression de $\varphi_{\mu,\sigma}$:

$$\varphi_{\mu,\sigma}(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

Nous allons retrouver cette égalité en utilisant d'une part l'expression de la densité φ pour une loi normale centrée réduite et d'autre part le lien entre la fonction de répartition et la densité qui peut s'écrire :

$$\Phi_{\mu,\sigma}(x) = \int_{-\infty}^{x} \varphi_{\mu,\sigma}(t)dt \tag{16}$$

Cette égalité est équivalente à :

$$\Phi'_{\mu,\sigma}(x) = \varphi_{\mu,\sigma}(x). \tag{17}$$

En particulier, notons que $\Phi' = \varphi$. Par définition d'une fonction de répartition, on a : $\Phi_{\mu,\sigma}(x) = \mathbb{P}(Z \leq x)$ et comme $Z = \sigma X + \mu$, on peut écrire :

$$\Phi_{\mu,\sigma}(x) = \mathbb{P}(Z \leq x) = \mathbb{P}(X \leq \frac{x-\mu}{\sigma}) = \Phi(\frac{x-\mu}{\sigma}).$$

Pour finir, on va utiliser une formule générale de dérivation : si f est une fonction dérivable, et si a et b sont des constantes, alors $(f(ax+b))' = af'(ax+b)$. En appliquant cette formule (avec $a = 1/\sigma$ et $b = -\mu/\sigma$) et en utilisant l'équation (17), on obtient :

$$\varphi_{\mu,\sigma}(x) = \Phi'_{\mu,\sigma}(x) = \left(\Phi(\frac{x-\mu}{\sigma})\right)' = \frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma}).$$

Pour finir, on obtient donc :

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma}) = \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\frac{((x-\mu)/\sigma)^2}{2}}.$$

On retrouve bien l'expression de $\Phi_{\mu,\sigma}$, puisque :

$$-\frac{((x-\mu)/\sigma)^2}{2} = -\frac{(x-\mu)^2}{2\sigma^2}.$$

On aurait pu aussi retrouver l'expression de $\Phi_{\mu,\sigma}$ à partir de Φ et en utilisant l'égalité (16) et un outil qui s'appelle le changement de variable dans une intégrale.

Exercice 5. (Probabilités d'évènement et lois gaussiennes)

Soit X une variable distribuée suivant une loi normale centrée réduite. Donner la valeur des probabilités ci-dessous en fonction de la fonction Φ donnée en rappel (et potentiellement de constantes numériques que vous préciserez).

1) $\mathbb{P}(X \leq 1.37)$, $\mathbb{P}(X > 1.37)$ et $\mathbb{P}(X = 1.37)$;
2) $\mathbb{P}(X < 0.52)$ et $\mathbb{P}(X \leq -0.52)$;
3) $\mathbb{P}(X > 1.79)$ et $\mathbb{P}(X > -1.79)$;
4) $\mathbb{P}(-0.155 < X < 1.60)$, $\mathbb{P}(-1.3 < X < 2.1)$ et $\mathbb{P}(0.06 < X < 0.8)$;
5) $\mathbb{P}(X < -1.9 \text{ ou } X > 2.1)$;
6) $\mathbb{P}(|X| < 1.64)$, $\mathbb{P}(|X| < 1)$.

Correction:

On va se ramener chaque fois à une expression ne dépendant que de la fonction Φ évaluée sur une valeur positive. Pour cela on va utiliser plusieurs identités. Tout d'abord, pour une variable aléatoire à densité, la probabilité que $X = a$ est égale à 0, et donc $\mathbb{P}(X \leq a) = \mathbb{P}(X < a)$. Ensuite pour une variable aléatoire quelconque, on a :

$$\mathbb{P}(X > a) = 1 - \mathbb{P}(X \leq a) \text{ et } \mathbb{P}(a \leq X \leq b) = \Phi(b) - \Phi(a).$$

Enfin, puisque la fonction densité de la loi normale est paire, on a l'égalité $\mathbb{P}(X < a) = \mathbb{P}(X > -a)$. Il est conseillé de faire un dessin avec le graphe de la fonction densité de la loi normale et les aires représentant les différentes probabilités pour illustrer les relations ci-dessus. Notons que pour donner une valeur approchée des probabilités demandées, on peut ensuite utiliser un logiciel de calcul comme Python, ou une application pour smartphone comme Probability Distribution.

1) $\mathbb{P}(X \leq 1.37) = \Phi(1.37) \approx 0.91$; $\mathbb{P}(X > 1.37) = 1 - \Phi(1.37) \approx 0.09$; $\mathbb{P}(X = 1.37) = 0$;
2) $\mathbb{P}(X < 0.52) = \mathbb{P}(X \leq 0.52) = \Phi(0.52) \approx 0.70$; $\mathbb{P}(X \leq -0.52) = \mathbb{P}(X \geq 0.52) = 1 - \mathbb{P}(X \leq 0.52) \approx 0.30$.
3) $\mathbb{P}(X > 1.79) = 1 - \Phi(1.79) \approx 0.04$; $\mathbb{P}(X > -1.79) = \mathbb{P}(X < 1.79) = \Phi(1.79) \approx 0.96$;
4)
$$\begin{aligned}\mathbb{P}(-0.155 < X < 1.60) &= \mathbb{P}(X < 1.60) - \mathbb{P}(X < -0.155) \\ &= \mathbb{P}(X < 1.60) - \mathbb{P}(X > 0.155) \\ &= \Phi(1.60) + \Phi(0.155) - 1 \approx 0.51\end{aligned}$$

   Par un calcul identique, on obtient :

$$\mathbb{P}(-1.3 < X < 2.1) = \Phi(2.1) + \Phi(1.3) - 1 \approx 0.89$$

   et enfin : $\mathbb{P}(0.06 < X < 0.8) = \Phi(0.8) - \Phi(0.06) \approx 0.26$;
5) Comme les deux événements $X < -1.9$ et $X > 2.1$ sont disjoints, on peut écrire : $\mathbb{P}(X < -1.9 \text{ ou } X > 2.1) = \mathbb{P}(X < -1.9) + \mathbb{P}(X > 2.1) = 2 - (\Phi(1.9) + \Phi(2.1)) \approx 0.05$;
6) D'après la définition de la valeur absolue : $\mathbb{P}(|X| < 1.64) = \mathbb{P}(-1.64 < X < 1.64)$. Puis en faisant un calcul identique au cas 4, on obtient $\mathbb{P}(|X| < 1.64) = 2\Phi(1.64) - 1 \approx 0.90$, De même : $\mathbb{P}(|X| < 1) = 2\Phi(1) - 1 \approx 0.68$.

EXERCICE 6. Probabilités d'évènement et lois gaussiennes, bis

Soit Z de loi normale d'espérance 20 et d'écart-type 5. Calculer $\mathbb{P}(Z < 18)$, $\mathbb{P}(Z \leq 39)$, $\mathbb{P}(Z > 37)$, $\mathbb{P}(Z > 11)$ et $\mathbb{P}(22 \leq Z \leq 31)$.

Correction:

Comme dans l'exercice précédent, on va se ramener à des évaluations de la fonction Φ. Pour cela on considère la variable aléatoire X telle que $X = (Z - 20)/5$ (ou de manière équivalente $Z = 5X + 20$) ; rappelons que X suit alors une loi normale centrée réduite. Les valeurs numériques sont arrondies au centième.

$$\mathbb{P}(Z < 18) = \mathbb{P}(5X + 20 < 18) = \mathbb{P}(X < -\frac{2}{5}) = \Phi(-0.4) \approx 0.34$$

$$\mathbb{P}(Z \leq 39) = \mathbb{P}(5X + 20 \leq 39) = \mathbb{P}(X \leq \frac{19}{5}) = \Phi(3.8) \approx 1$$

$$\mathbb{P}(Z > 37) = 1 - \mathbb{P}(Z < 37) = 1 - \mathbb{P}(5X + 20 < 37) = 1 - \Phi(3.4) \approx 0$$

$$\mathbb{P}(Z > 11) = 1 - \mathbb{P}(Z < 11) = 1 - \mathbb{P}(5X + 20 < 11) = 1 - \Phi(-1.8) \approx 0.96$$

$$\mathbb{P}(22 \leq Z \leq 31) = \mathbb{P}(Z \leq 31) - \mathbb{P}(Z \leq 22) = \Phi(2.2) - \Phi(0.2) \approx 0.41$$

EXERCICE 7. (Quantiles gaussiens)

Soit Z de loi normale d'espérance 130 et d'écart-type 5. Résoudre chacune des équations suivantes (l'inconnue est b) en utilisant $q = \Phi^{-1}$, la fonction quantile de la loi normale centrée réduite :

1) $\mathbb{P}(Z < b) = 0.975$,
2) $\mathbb{P}(Z > b) = 0.025$,
3) $\mathbb{P}(Z < b) = 0.305$.

Correction:
On pose $X = (Z - 130)/5$; alors X suit une loi normale centrée réduite.

1)

$$\begin{aligned}
\mathbb{P}(Z < b) = 0.975 \quad &\Leftrightarrow \quad \mathbb{P}(5X + 130 < b) = 0.975 \\
&\Leftrightarrow \quad \mathbb{P}(X < \tfrac{b-130}{5}) = 0.975 \\
&\Leftrightarrow \quad \Phi(\tfrac{b-130}{5}) = 0.975 \\
&\Leftrightarrow \quad \tfrac{b-130}{5} = q_{0.975} \\
&\Leftrightarrow \quad b = 5q_{0.975} + 130
\end{aligned}$$

Numériquement on obtient $b \approx 139.80$.

2) l'équation $\mathbb{P}(Z > b) = 0.025$ est équivalente à $1 - \mathbb{P}(Z < b) = 0.025$, soit $\mathbb{P}(Z < b) = 0.975$, on est donc ramené au cas précédent.

3) Par le même raisonnement que dans le premier cas, l'équation $\mathbb{P}(Z < b) = 0.305$ admet comme solution $b = 5q_{0.305} + 130$, soit numériquement $b \approx 127.45$.

Exercice 8. (Quartiles approchés et histogrammes)

Table 1 – Distribution du nombre de cigarettes fumées par jour pour les 484 mères fumeuses.

Nb de cig.	Nb. de fumeurs (%)
0–5	16
5–10	25
10–15	14
15–20	4
20–30	32
30–40	5
40–60	4
Total	100

a) A partir du Tableau 1 trouver des quartiles approchés de la distribution du nombre de cigarettes fumées par jour des mères fumeuses pendant la grossesse.

b) Combinez les quatre dernières classes de ce tableau et tracez l'histogramme qui estime la densité : on veillera à proposer une courbe d'aire égale à 1.

c) Regardez l'histogramme de la Figure 2. On a oublié de faire le graphique pour une classe, celle correspondant aux âges entre 35 et 40 ans, que l'on a rempli à tort avec un 0. Comblez cette lacune.

Exercice 9. (Moyennes et écart-types)

a) Dans une étude du Missouri, le poids moyen à la naissance des bébés issus de mères fumeuses est 3180g et l'écart-type de 500g. Quel est le poids moyen et l'écart-type en onces sachant qu'il y a 0.035 onces dans 1g.

b) Soient $x_1, \ldots, x_n$ quelques observations. Pour des raisons de commodités, Bob a changé les unités menant à de nouvelles observations

$$y_i = ax_i + b, \qquad i = 1, \ldots, n \;,$$

avec $a > 0$ et $b \in \mathbb{R}$. Exprimez la moyenne et l'écart-type des y_i en fonction de ceux des x_i, et des constantes a et b.

Correction:

FIGURE 2 – Histogramme de l'âge des pères dans l'étude CHDS vue en cours. Les nombres indiquent la hauteur (en %) de chaque rectangle. Le rectangle pour la classe $[35, 40[$ est manquant.

a) Notons n le nombre de données, (x_i) la mesure des poids des bébés en grammes et (y_i) le poids en onces. On a donc $y_i = 0.035x_i$. La moyenne des mesures en grammes est donnée par : $\bar{x}_n = 3180$ et grâce à la relation vue en cours, on a $\bar{y}_n = 0.035\bar{x}_n = 111.3$. De même on a la relation suivante entre les écarts-type : $\mathrm{var}_n(\mathbf{x}) = |0.035| \, \mathrm{var}_n(\mathbf{y})$, soit $\mathrm{var}_n(\mathbf{y}) = 0.035 \times 500 = 17.5$. Notons que la valeur absolue dans la relation sur les écarts-type n'est pas nécessaire ici, mais il ne faut pas l'oublier, sinon cela pourrait conduire à des écarts-type négatifs !

b) Il s'agit ici d'une simple application des relations vues en cours. On a d'une part $\bar{y}_i = a\bar{x}_i + b$ et $\mathrm{var}_n \, \mathbf{y} = |a| \, \mathrm{var}_n(\mathbf{x}) = a \, \mathrm{var}_n(\mathbf{x})$ puisque $a > 0$.

EXERCICE 10. La loi normale et boîte à moustache

a) On modélise la loi de la taille des mères (dans la même étude que celle vue en cours) par une loi normale d'espérance 64 pouces et d'écart-type 2.5 pouces. Utiliser cette approximation et la fonction Φ pour estimer la proportion de mères mesurant entre 61.5 et 64.5 pouces. Aide : on pourra utiliser les approximations $\Phi(0.2) \approx 0.5793$ et $\Phi(1) \approx 0.8413$.

b) Supposons que l'on dispose d'observations issues d'une loi normale centrée réduite. Quelle proportion des observations peut-on espérer voir en dehors des « moustaches » de la boîte à moustache ? Aide : on pourra utiliser les approximations $\Phi(0.675) \approx 0.75$ et $\Phi(2.7) \approx 0.9965$.

Correction:

a) Notons Z la variable aléatoire modélisant la taille des mères ; on veut donc calculer $\mathbb{P}(61.5 < Z < 64.5)$, pour cela posons $X = (Z - 64)/2.5$. La variable aléatoire X suit donc une loi normale centrée réduite. On peut donc écrire :

$$\begin{aligned}
\mathbb{P}(61.5 < Z < 64.5) &= \mathbb{P}(61.5 < 2.5X + 64 < 64.5) \\
&= \mathbb{P}(\tfrac{61.5-64}{2.5} < X < \tfrac{64.5-64}{2.5}) \\
&= \mathbb{P}(-1 < X < 0.2) \\
&= \mathbb{P}(X < 0.2) - \mathbb{P}(X < -1) \\
&= \Phi(0.2) + \Phi(1) - 1 \approx 0.4206
\end{aligned}$$

b) Notons X les observations ; on suppose donc que X suit une loi normale centrée réduite. Les deux « moustaches » ont une longueur (à partir de la boîte) donnée par une fois et demi l'écart interquartile. Commençons par calculer cet écart. Celui est donné par : $q_{0.75} - q_{0.25}$. Par symétrie de la loi normale centrée réduite, $q_{0.25} = -q_{0.75}$, et donc l'écart interquartile est égal à

$2q_{0.75} = 2.\Phi^{-1}(0.75) \approx 2 \times 0.675 = 1.35$. La proportion que l'on peut espérer en dehors des moustaches est donnée par :

$$\mathbb{P}(X > q_{0.75} + 1.5 \times 1.35) + \mathbb{P}(X < q_{0.25} - 1.5 \times 1.35) \ .$$

En utilisant encore une fois la symétrie de la distribution et l'approximation $q_{0.75} + 1.5 \times 1.35 \approx 2.7$, la proportion recherchée est égale à :

$$2 \times \mathbb{P}(X > 2.7) \approx 2 \times (1 - \mathbb{P}(X < 2.7)) \approx 0.007 \ .$$

On peut donc s'attendre à environ une proportion de 7 ‰ des données à l'extérieur des moustaches.