Université de Nice Sophia-Antipolis
L2M
TD de Statistique

Année Universitaire 2011/2012
**Statistique**

# Correction TD n° 3.

**Exercice 1 :** On utilisera le lemme suivant

1 Lemme *Soit $X$ une variable aléatoire continue telle que sa fonction de répartition $F$ est dérivable sauf aux points $x_1, \dots, x_n$. Alors une densité $f$ de $X$ est définie par $f(x) = F'(x)$ pour tout $x \in \mathbb{R} \setminus \{x_1, \dots, x_n\}$ et $f(x_i)$ vaut une valeur arbitraire positive.*

*De plus en tout point $x$ où $f$ est continue, $F$ est dérivable en $x$ et vérifie $F'(x) = f(x)$.*

*Si $f_1$ est une densité de $X$ et $f_2$ une fonction positive telle que pour tout $x \in \mathbb{R} \setminus \{x_1, \dots, x_n\}$, $f_1(x) = f_2(x)$, alors $f_2$ est une densité de $X$.*

1. Notons $F_Y$ la fonction de répartition de $Y$ et $F_X$ celle de $X$. Alors, pour tout $t \in \mathbb{R}$,
$$F_Y(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(-4X + 3 \leq t) = \mathbb{P}\left(X \geq \frac{3-t}{4}\right) = 1 - F_X\left(\frac{3-t}{4}\right)$$
Puisque $f$ est continue sauf aux points 0 et 4, la fonction $F_X$ est dérivable sur $\mathbb{R}$ sauf éventuellement en 0 et 4 et de plus pour tout $x \notin \{0, 4\}$, $F'_X(x) = f(x)$. Par le lemme, une densité $f_Y$ de $Y$ est donnée pour tout $t \notin \{-13, 3\}$ par $f_Y(t) = F'_Y(t) = \frac{1}{4} f'\left(\frac{3-t}{4}\right)$. On peut donc prendre:
$$f_Y(t) = \begin{cases} \frac{1}{16} & \text{si } -13 < t < 3 \\ 0 & \text{sinon.} \end{cases}$$
La loi de $Y$ est donnée par la densité précédente.
Remarque: on aurait pu également définir $f_Y$ par
$$f_Y(t) = \begin{cases} \frac{1}{16} & \text{si } -13 \leq t \leq 3 \\ 0 & \text{sinon.} \end{cases}$$
ou bien par
$$f_Y(t) = \begin{cases} \frac{1}{16} & \text{si } -13 \leq t < 3 \\ 0 & \text{sinon.} \end{cases}$$
ou encore par
$$f_Y(t) = \begin{cases} \frac{1}{16} & \text{si } -13 < t \leq 3 \\ 0 & \text{sinon.} \end{cases}$$

2. Notons $F_Z$ la fonction de répartition de $Z$, $F_X$ la fonction de répartition de $X$. Alors, puisque $Z$ est une variable aléatoire positive, $F_Z(t) = 0$ pour $t < 0$. De plus, pour tout $t \geq 0$,
$$\begin{aligned} F_Z(t) &= \mathbb{P}(Z \leq t) \\ &= \mathbb{P}(X^2 \leq t) \\ &= \mathbb{P}\left(-\sqrt{t} \leq X \leq \sqrt{t}\right) \\ &= F_X(\sqrt{t}) - F_X(-\sqrt{t}) \\ &= F_X\left(\sqrt{t}\right) + 0 \end{aligned}$$
En particulier, pour tout $t \neq 0$ et $t \neq 16$, une densité $f_Z$ de $Z$ est donnée par $f_Z(t) = F'_Z(t)$. En particulier, pour tout $t > 0$ et $t \neq 16$,
$$f_Z(t) = \frac{1}{2\sqrt{t}} F'_X\left(\sqrt{t}\right) = \frac{1}{2\sqrt{t}} f\left(\sqrt{t}\right).$$
On prend donc:
$$f_Z(t) = \begin{cases} \frac{1}{8\sqrt{t}}. & \text{si } 0 < t < 16 \\ 0 & \text{sinon.} \end{cases}$$

**Exercice 2:**

1. Soit $x \in \mathbb{R}$. Alors,

$$F_X(x) \quad = \quad \int_{-\infty}^{x} f(t)\mathrm{d}t$$

Donc, si $x \leq 0$, $F_X(x) = 0$. Si désormais, $x > 0$,

$$F_X(x) \quad = \quad \int_0^x \frac{t}{\alpha^2} \exp\left(-\frac{t^2}{2\alpha^2}\right) \mathrm{d}t.$$

En faisant le changement de variable $t = u\sqrt{2}\alpha$,

$$\begin{aligned} F_X(x) &= \int_0^{\frac{x}{\sqrt{2}\alpha}} 2u \exp\left(-u^2\right) \mathrm{d}u. \\ &= \left[-e^{-u^2}\right]_0^{\frac{x}{\sqrt{2}\alpha}} \\ &= 1 - e^{\frac{x^2}{2\alpha^2}}. \end{aligned}$$

2. Par définition un $p \in \{0, 1, 2, \ldots, 100\}$ème percentile est un $x_p \in \mathbb{R}$ tel que $F_X(x_p) = p/100$. La résolution de l'équation donne alors pour $p \neq 0$,

$$x_p = \sqrt{-2\alpha^2 \ln\left(1 - \frac{p}{100}\right)},$$

et pour $p = 0$ tout $x_p \leq 0$ convient.

3. On admettra le résultat

$$\int_0^{\infty} e^{-x^2} \mathrm{d}x = \frac{\sqrt{\pi}}{2}.$$

L'espérance de $X$ existe car la fonction $t \mapsto t^2/\alpha^2 e^{-t^2/(2\alpha^2)}$ est intégrable sur $[0, +\infty[$. De plus,

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{\infty} \frac{t^2}{\alpha^2} \exp\left(-\frac{t^2}{2\alpha^2}\right) \mathrm{d}t \\ &= 2\sqrt{2}\alpha \int_0^{\infty} x^2 e^{-x^2} \mathrm{d}x \end{aligned}$$

en faisant le changement de variable $x = t/(\sqrt{2}\alpha)$. En écrivant

$$\mathbb{E}(X) \quad = \quad -\sqrt{2}\alpha \int_0^{\infty} x\left(-2xe^{-x^2}\right) \mathrm{d}x$$

et en effectuant une intégration par partie, on obtient

$$\mathbb{E}(X) \quad = \quad \sqrt{2}\alpha \int_0^{\infty} e^{-x^2} \mathrm{d}x = \alpha\sqrt{\pi/2}.$$

Pour calculer la variance, on montre que $\mathbb{E}(X^2)$ existe et on utilise la formule

$$\mathrm{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Les calculs sont laissés au lecteur.

**Exercice 3:**

1. On rappelle qu'une fonction de densité est une fonction *positive* $f$ telle que $\int_{\mathbb{R}} f(x)\mathrm{d}x = 1$. Dire que $f$ est positive se traduit par $c \geq 0$. De plus,
$$\int_{\mathbb{R}} f(x)\mathrm{d}x = \int_1^{\infty} cx^{-\alpha-1} = c\left[-\frac{1}{\alpha}x^{-\alpha}\right]_1^{\infty} = \frac{c}{\alpha}.$$
Donc $c = \alpha$. (qui est bien positif)
2. Notons $F$ la fonction de répartition de $X$. Si $x \leq 1$ alors $F(x) = 0$. De plus, si $x > 1$, on a
$$F(x) = \int_1^x \alpha t^{-\alpha-1}\mathrm{d}t = 1 - x^{-\alpha}.$$
3. Notons $F_Z$ la fonction de répartition de $Z$. Dire que $Z > t$ signifie que $X > t$ et $Y > t$. Alors,
$$\begin{aligned} F_Z(t) &= \mathbb{P}(Z \leq t) \\ &= 1 - \mathbb{P}(Z > t) \\ &= 1 - \mathbb{P}(X > t, Y > t) \\ &= 1 - \mathbb{P}(X > t)\,\mathbb{P}(Y > t). \end{aligned}$$
où nous utilisons dans la dernière égalité le fait que $X$ et $Y$ sont indépendantes. De plus, puisque $X$ et $Y$ ont même loi, $\mathbb{P}(X > t) = \mathbb{P}(Y > t)$. Donc,
$$\begin{aligned} F_Z(t) &= 1 - \mathbb{P}(X > t)^2 \\ & \quad 1 - (1 - F(t))^2. \end{aligned}$$
Donc la fonction de répartition est donnée par:
$$F_Z(t) = \begin{cases} 0 & \text{si } t < 1 \\ 1 - t^{-2\alpha} & \text{si } t \geq 1. \end{cases}$$
En utilisant le lemme, on peut prendre une densité $f_Z$ définie pour $t \neq 1$, par $f_Z(t) = F_Z'(t)$ et $f_Z(1)$ choisie arbitrairement. On peut donc prendre
$$f_Z(t) = \begin{cases} 0 & \text{si } t < 1 \\ 2\alpha t^{-2\alpha-1} & \text{si } t \geq 1. \end{cases}$$
4. La fonction $x \mapsto x^{-\alpha}$ est intégrable sur $[1, +\infty[$ (car $\alpha > 1$) et donc $\mathbb{E}(X)$ existe. De plus,
$$\mathbb{E}(X) = \int_1^{\infty} \alpha x^{-\alpha} = \frac{\alpha}{\alpha - 1}.$$
D'autre part, pour caculer la variance, puisque $x \mapsto x^{-\alpha+1}$ est intégrable (car $\alpha - 1 > 1$), $\mathbb{E}(X^2)$ existe. De plus,
$$\mathbb{E}(X^2) = \int_1^{\infty} \alpha x^{-\alpha+2} = \frac{\alpha}{\alpha - 2}.$$
D'autre part,
$$\mathrm{var}X = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{\alpha}{\alpha - 2} - \left(\frac{\alpha}{\alpha - 1}\right)^2.$$
5. L'équation $\mathbb{E}(X) = 4/3$ se réécrit $\alpha/(\alpha - 1) = 4/3$. Cette équation a une unique solution, $\alpha = 4$.

**Exercice 4:**
On rappelle qu'une variable aléatoire $X$ suit une loi exponentielle de paramètre $\lambda > 0$ si elle admet une densité de la forme $f(x) = \lambda e^{-\lambda x}$ si $x \geq 0$ et $f(x) = 0$ si $x < 0$.

1. Puisque $X$ est une variable aléatoire positive (c'est à dire $\mathbb{P}(X \geq 0) = 1$), l'ensemble des valeurs que peut prendre $T$ est $T(\Omega) = \{1, 2, 3, \dots\} = \mathbb{N}^\star$. Cela implique que $T$ est une variable aléatoire discrète. Pour déterminer sa loi, il faut calculer $\mathbb{P}(T = k)$ pour tout $k \in T(\Omega)$. Soit donc $k \in \mathbb{N}^\star$,
$$\begin{aligned} \mathbb{P}(T = k) &= \mathbb{P}([X] + 1 = k) \\ &= \mathbb{P}([X] = k - 1) \\ &= \mathbb{P}(k - 1 \leq X < k) \\ &= \int_{k-1}^{k} \lambda e^{-\lambda x}\mathrm{d}x \\ &= e^{-\lambda(k-1)} - e^{-\lambda k} \\ &= q^{k-1}p. \end{aligned}$$

2. Notons $Z$ la variable aléatoire $Z = \min(X, Y)$, $F_Z$ sa fonction de répartiton et $F$ la fonction de répartition d'une loi exponentielle de paramètre $\lambda$. Alors, pour tout $t \in \mathbb{R}$,

$$\begin{aligned} F_Z(t) &= \mathbb{P}\left(\min(X,Y) \leq t\right) \\ &= 1 - \mathbb{P}\left(\min(X,Y) > t\right) \\ &= 1 - \mathbb{P}\left(X > t, Y > t\right) \\ &= 1 - \mathbb{P}\left(X > t\right)\mathbb{P}\left(Y > t\right) \\ &= 1 - (1 - F(t))^2 \end{aligned}$$

où nous utilisons à la 3ème égalité le fait que $X$ et $Y$ sont indépendantes et à la 4ème que $X$ et $Y$ suivent une loi exponentielle de paramètre $\lambda$. De plus, on laisse au lecteur le fait que $F(t) = 1 - e^{-\lambda t}$ pour tout $t \geq 0$ et $F(t) = 0$ sinon. La loi de $Z$ est donnée par sa fonction de répartition $F_Z$, donc par

$$F_Z(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 - e^{-2\lambda t} & \text{si } t \geq 0. \end{cases}$$

Si l'on souhaite donner une densité $f_Z$ de $Z$, en utilisant le lemme, on peut prendre

$$f_Z(t) = \begin{cases} 0 & \text{si } t < 0 \\ 2\lambda e^{-2\lambda t} & \text{si } t \geq 0. \end{cases}$$

Le calcul de la loi de $Z' = \max(X, Y)$ se fait de manière analogue. Notons $F_{Z'}$ la fonction de répartition de $Z'$. Pour tout $t \in \mathbb{R}$,

$$F_{Z'}(t) = \mathbb{P}(Z' \leq t) = \mathbb{P}(X \leq t, Y \leq t) = F(t)^2.$$

Donc, en particulier,

$$F_{Z'}(t) = \begin{cases} 0 & \text{si } t < 0 \\ \left(1 - e^{-\lambda t}\right)^2 & \text{si } t \geq 0. \end{cases}$$

Une densité $f_{Z'}$ de $Z'$, en utilisant le lemme, est

$$f_{Z'}(t) = \begin{cases} 0 & \text{si } t < 0 \\ \lambda e^{-\lambda t}\left(1 - e^{-\lambda t}\right) & \text{si } t \geq 0. \end{cases}$$

**Exercice 5:**

1. Soit $F_Y$ la fonction de répartition de $Y$. Alors, pour tout $t \in \mathbb{R}$,

$$F_Y(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(aX + b \leq t).$$

Donc, si $a > 0$,

$$F_Y(t) = \mathbb{P}\left(X \leq \frac{t-b}{a}\right) = F\left(\frac{t-b}{a}\right).$$

Si $a < 0$,

$$F_Y(t) = \mathbb{P}\left(X \geq \frac{t-b}{a}\right) = 1 - \mathbb{P}\left(X < \frac{t-b}{a}\right) = 1 - \mathbb{P}\left(X \leq \frac{t-b}{a}\right) = 1 - F\left(\frac{t-b}{a}\right)$$

car la variable aléatoire $X$ est continue.
Puisque $f$ est continue, $F$ est dérivable. De plus, séparant les cas $a < 0$ et $a > 0$, par le lemme, une densité $f_Y$ de $Y$ est donnée par

$$f_Y(t) = F_Y'(t) = \frac{1}{|a|} f\left(\frac{t-b}{a}\right)$$

2. Soit $F_Z$ la fonction de répartition de $Z$ définie pour $t \in \mathbb{R}$ par $F_Z(t) = \mathbb{P}(Z \leq t) = \mathbb{P}(|X| \leq t)$. Pour tout $t < 0$, $F_Z(t) = 0$ car $|X|$ est une variable aléatoire positive ($\mathbb{P}(|X| \geq 0) = 1 = 1 - F_Z(0)$ , donc $F_Z(0) = 0$ et par croissance et positivité de $F_Z$, $F_Z(t) = 0$ pour $t < 0$). De plus, pour $t \geq 0$,

$$F_Z(t) = \mathbb{P}(-t \leq X \leq t) = F(t) - F(-t).$$

On peut donc prendre une densité $f_Z$ de $Z$ par

$$f_Z(t) = \begin{cases} 0 & \text{si } t < 0 \\ f(t) + f(-t) & \text{si } t \geq 0. \end{cases}$$

3. De même, notons $F_T$ la fonction de répartition de $T$. Alors, pour tout $t \in \mathbb{R}$,

$$F_T(t) = \mathbb{P}(T \le t) = \mathbb{P}(|X| \le e^t) = F(e^t) - F(-e^t).$$

et on peut prendre comme densité la fonction $f_T$ définie sur $\mathbb{R}$ par

$$f_T(t) = e^t f(e^t) + e^t f(-e^t), \quad \forall t \in \mathbb{R}$$

4. La fonction $F$ est continue et strictement croissante (car sa dérivée est strictement positive). Par conséquent, elle réalise une bijection entre $\mathbb{R}$ et $F(\mathbb{R}) = ]0,1[$. Notons $F_U$ la fonction de répartition de $U$. Alors, pour $t \in \mathbb{R}$,

$$F_U(t) = \mathbb{P}\left(F(X) \le t\right).$$

Donc si $t \le 0$, $F_U(t) = 0$ et si $t \ge 1$, $F_U(t) = 1$. Pour $t \in ]0,1[$,

$$F_U(t) = \mathbb{P}\left(X \le F^{-1}(t)\right).$$

En effet, puisque $F$ est bijective et strictement croissante, $F^{-1}$ existe et est strictement croissante. Donc,

$$F_U(t) = F(F^{-1}(t)) = t.$$

La fonction de répartition $F_U$ est donc définie par

$$F_U(t) = \begin{cases} 0 & \text{si } t \le 0 \\ t & \text{si } 0 < t < 1. \\ 1 & \text{si } t \ge 1. \end{cases}$$

On reconnait la fonction de répartition d'une loi uniforme sur $[0,1]$. Donc $U$ suit une loi uniforme sur $[0,1]$.

5. On rappelle que dans cette feuille d'exercice, pour tout réel $x \in \mathbb{R}$, $[x]$ est la partie entière de $x$. Donc, attention, ici $V$ n'est pas une variable aléatoire continue. Puisque $V(\Omega) = \mathbb{Z}$, $V$ est une variable aléatoire discrète. Elle n'admet donc pas de densité. En revanche, on peut calculer sa loi. Elle est donnée par: $V(\Omega) = \mathbb{Z}$ et pour tout $k \in \mathbb{Z}$,

$$\mathbb{P}(V = k) = \mathbb{P}\left(k \le X < k+1\right) = F(k+1) - F(k).$$

Pour $t \in \mathbb{R}$,

$$\begin{aligned} F_V(t) &= \sum_{k \in \mathbb{Z}, k \le t} \mathbb{P}(V = k) \\ &= \sum_{k=-\infty}^{[t]} \mathbb{P}(V = k) \\ &= \sum_{k=-\infty}^{[t]} F(k+1) - F(k) \\ &= F\left(1 + [t]\right). \end{aligned}$$

**Exercice 6:**
On rappelle la propriété suivante: si $X$ suit une loi normale $\mathcal{N}(m, \sigma^2)$ alors pour tous réels, $a, b$, $aX + b$ suit une loi normale $\mathcal{N}(am + b, a^2\sigma^2)$. On note dans cette exercice $F$ la fonction de répartition d'une loi $\mathcal{N}(0,1)$.

Remarque: pour tout $t \in \mathbb{R}$, $F(t) = 1 - F(-t)$. En effet, si $Y$ est une variable aléatoire suivant la loi $\mathcal{N}(0,1)$. Alors $F(t) = \mathbb{P}(Y \le t) = \mathbb{P}(-Y \ge -t) = 1 - \mathbb{P}(-Y < -t)$. De plus, puisque $Y \sim \mathcal{N}(0,1)$, $-Y \sim \mathcal{N}(0,1)$ (appliquer la remarque au dessus avec $a = -1$ et $b = 0$). Donc, $F(t) = 1 - F(-t)$.

Dans cet exercice, $X \sim \mathcal{N}(1, 25)$, donc la variable aléatoire $Y = (X-1)/5$ suit une loi $\mathcal{N}(0,1)$

- Pour tout $a \ge 1$,

$$\mathbb{P}(1 < X < a) = \mathbb{P}(0 < Y < (a-1)/5) = F((a-1)/5) - F(0) = F((a-1)/5) - 1/2$$

On cherche donc $a$ tel que $F((a-1)/5) = 1.45$. Cela est impossible car $F(x) \le 1$ pour tout $x$.

- Soit $a \geq 0$.

$$\mathbb{P}(-a+1 < X < a+1) = \mathbb{P}(-a/5 < Y < a/5) = F(a/5) - F(-a/5) = 2F(a/5) - 1$$

On résout donc $F(a/5) = 0.975$. D'après la table $a/5 \simeq 1.96$ donc $a \simeq 9.825$.

- On a:

$$\mathbb{P}(X > a+1) = \mathbb{P}(Y > a/5) = 1 - F(a/5) = 0.02$$

. Donc $F(a/5) = 0.98$. D'après la table, $a/5 \simeq 2.056$ donc $a \simeq 10.28$ .

- Pour $a \geq 0$,

$$\mathbb{P}(|X-1| > a) = 1 - \mathbb{P}(-a+1 < X < a+1)$$

On résout donc $\mathbb{P}(-a+1 < X < a+1) = 0.95$. On l' a déjà fait, on a trouvé $a \simeq 9.825$.

**Exercice 7:**
Dans le corrigé du TD précédent on a expliqué comment approcher une loi Binomiale par une loi de Poisson. Dans ce corrigé, on approchera une loi Binomiale par une loi Normale.

2 Lemme *Soit $Y$ une variable aléatoire suivant une loi normale $\mathcal{N}(0,1)$. Soit pour tout $n \in \mathbb{N}^\star$, $X_n$ une variable aléatoire discrète suivant une loi Binomiale de paramètres $n$ et $p \in ]0,1[$. Alors, pour tout $a \in \mathbb{R} \cup \{-\infty\}$ et $b \in \mathbb{R} \cup \{+\infty\}$, $a \leq b$,*

$$\lim_{n \to +\infty} \mathbb{P}\left(a < \frac{X_n - np}{\sqrt{np(1-p)}} < b\right) = \mathbb{P}(a < Y < b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \mathrm{d}t.$$

Ce théorème est une conséquence du Théorème Central Limite, que l'on écrira plus loin. Cette version est aussi connu sous le nom du théorème de Moivre-Laplace. Ce théorème peut être utilisé pour approcher lorsque $n$ est grand une loi Binomiale $\mathcal{B}(n,p)$ par une loi Normale $\mathcal{N}(np, np(1-p))$.

En effet, si $Y$ suit une loi $\mathcal{N}(0,1)$ alors $Y' = np + \sqrt{np(1-p)}\,Y$ suit une loi $\mathcal{N}(np, np(1-p))$. Donc pour tout $\alpha \leq \beta$,

$$\begin{aligned}
\mathbb{P}(\alpha < X_n < \beta) &= \mathbb{P}\left(\frac{\alpha - np}{\sqrt{np(1-p)}} < \frac{X_n - np}{\sqrt{np(1-p)}} < \frac{\beta - np}{\sqrt{np(1-p)}}\right) \\
&\simeq \mathbb{P}\left(\frac{\alpha - np}{\sqrt{np(1-p)}} < Y < \frac{\beta - np}{\sqrt{np(1-p)}}\right) \\
&\simeq \mathbb{P}(\alpha < Y' < \beta).
\end{aligned}$$

En pratique cette majoration est correcte pour $n$ grand et $p$ ni trop proche de 0 ni trop proche de 1. Typiquement, on l'accepte quand $n > 50$, $np > 5$ et $n(1-p) > 5$.

Dans cet exercice, on a $n = 1100$ sacs, que l'on numérote de 1 à $n$. On considère la variable aléatoire $X_i$ qui vaut 1 si le $i$ème sac doit être remplacé, et $X_i = 0$ sinon. Les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes et suivent une loi de Bernouilli d'un paramètre $p$ à déterminer. Dire que $X_1 = 1$ c'est dire que dans un sac de 50 graines, 3 graines n'ont pas germées. Numérotons ces graines de 1 à $m = 50$, et notons $Y_i = 1$ si la $i$ème graine n'a pas germée, $Y_i = 0$ si elle a germée. Le nombre de graines qui n'ont pas germées dans le sac est donc $\sum_{i=1}^m Y_i$. Donc,

$$p = \mathbb{P}(X_1 = 1) = \mathbb{P}\left(\sum_{i=1}^m Y_i \geq 3\right)$$

De plus les $Y_i$ sont indépendantes, et obeissent à une loi de Bernouilli de paramètre 0.01. Donc la variable aléatoire $Y = \sum_{i=1}^m Y_i$ suit une loi Binomiale de paramètres $m = 50$ et 0.01. D'après le TD précédent, $Y$ s'approxime par une variable aléatoire $Z$ suivant une loi de poisson de paramètre $50 \times 0.01 = 0.5$. Alors,

$$p \simeq 1 - \mathbb{P}(Z \leq 2) \simeq 1 - \left(e^{-1/2} + (1/2)e^{-1/2} + (1/2)^2/2! e^{-1/2}\right) \simeq 0.0144.$$

De plus, le nombre de sacs que le marchand doit remplacer est donné par $\sum_{i=1}^n X_i$. On cherche donc à calculer la probabilité $\mathbb{P}(\sum_{i=1}^n X_i \geq 40)$.

Or, puisque $\sum_{i=1}^n X_i$ est une somme de variable aléatoires indépendantes suivant une loi de bernouilli de paramètre $p$, $\sum_{i=1}^n X_i$ suit une loi Binomiale $\mathcal{B}(n,p)$. On vérifie que $n > 50$ , $np > 5$ et $n(1-p) > 5$. On est donc dans les conditions d'application du lemme. Soit $Z$ une variable aléatoire de loi $\mathcal{N}(0,1)$. Alors, par le lemme,

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq 40\right) &= \mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \geq \frac{40-np}{\sqrt{np(1-p)}}\right) \\
&\simeq \mathbb{P}\left(Z \geq \frac{40-np}{\sqrt{np(1-p)}}\right) \\
&\simeq \mathbb{P}\left(Z \geq 6.12\right)
\end{aligned}$$

Cette dernière probabilité se lit dans la table donnée.

$$\mathbb{P}\left(Z \geq 6.12\right) \simeq 0.$$

**Exercice 8:**
Supposons que l'on fasse $n$ réservations. Numérotons les clients de 1 à $n$, et notons $X_i = 1$ si le $i$ ème client se présente à l'embarquement, $X_i = 0$ sinon. D'après l'énoncé, $X_i$ suit une loi de Bernouilli de paramètre $p = 0.8$. On supposera que $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes, ce qui signifie que les passagers se présentent indépendemment les uns des autres. On est en sur-réservation si $\sum_{i=1}^{n} X_i \geq 400$. On cherche donc $n$, le plus grand possible tel que

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i > 400\right) \leq 0.05.$$

La variable aléatoire $\sum_{i=1}^{n} X_i$ suit une loi Binomiale $\mathcal{B}(n,p)$. Si $n$ est assez grand, elle s'approxime, d'après le lemme, par une loi Normale $\mathcal{N}(np, np(1-p))$. Plus précisément, notons $Y$ une variable aléatoire suivant une loi $\mathcal{N}(0,1)$. Alors,

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{n} X_i > 400\right) &= \mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} > \frac{400-np}{\sqrt{np(1-p)}}\right) \\
&\simeq \mathbb{P}\left(Y > \frac{400-np}{\sqrt{np(1-p)}}\right)
\end{aligned}$$

Notons

$$q_n = \frac{400-np}{\sqrt{np(1-p)}} = \frac{1000-2n}{\sqrt{n}}.$$

On sait, puisque la fonction $x \mapsto \mathbb{P}(Y > x)$ est une fonction bijective de $\mathbb{R}$ dans $]0,1[$ (car continue, strictement décroissante) que l'équation $P(Y > x) = 0.05$ admet une unique solution. D'après la table, pour $x = 1.65$, $\mathbb{P}(Y > x) \leq 0.05$ et $\mathbb{P}(Y > x) \simeq 0.05$. On cherche donc $n \in \mathbb{N}^\star$ le plus grand possible tel que $q_n \geq 1.65$. L'équation $q_x = 1.65$ admet dans $\mathbb{R}$ la solution $x = 481.89$. On prend donc $n = 481$.

On peut donc accepter jusqu'à 481 réservations.

**Exercice 10:**
Rappelons le Théorème Central Limite:

3 Lemme *Soit $Y$ une variable aléatoire suivant une loi normale $\mathcal{N}(0,1)$. Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes et de même loi tel que $\mathbb{E}(X_1^2)$ existe. Notons $m = \mathbb{E}(X_1)$ , $\sigma^2 = \mathrm{var}X_1$ et $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Alors, pour tout $a \in \mathbb{R}\cup\{-\infty\}$ et $b \in \mathbb{R}\cup\{+\infty\}$, $a \leq b$,*

$$\lim_{n\to+\infty} \mathbb{P}\left(a < \frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - m\right) < b\right) = \mathbb{P}\left(a < Y < b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\mathrm{d}t.$$

Notons $Z$ une variable aléatoire suivant une loi normale $\mathcal{N}(m,\sigma)$. Le théorème précédent nous dit que $\bar{X}_n = 1/n\sum_{i=1}^{n} X_i$ obeit approximativement à une loi $\mathcal{N}(m, \sigma^2/n)$. En effet, on déduit du fait que $Z = \sigma/\sqrt{n}\,Y + m$

obeit à $\mathcal{N}(m,\sigma^2/n)$ que, pour tout $\alpha \in \mathbb{R}\cup\{-\infty\}$ et $\beta \in \mathbb{R}\cup\{+\infty\}$, $\alpha \leq \beta$,

$$\begin{aligned}
\mathbb{P}\left(\alpha < \bar{X}_n < \beta\right) &= \mathbb{P}\left(\frac{\sqrt{n}}{\sigma}(\alpha - m) \leq \frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - m\right) \leq \frac{\sqrt{n}}{\sigma}(\beta - m)\right) \\
&\simeq \mathbb{P}\left(\frac{\sqrt{n}}{\sigma}(\alpha - m) \leq Y \leq \frac{\sqrt{n}}{\sigma}(\beta - m)\right) \\
&\simeq \mathbb{P}(\alpha < Z < \beta)
\end{aligned}$$

En écrivant qu'une variable aléatoire suivant une loi Binomiale, s'écrit comme une somme de variables aléatoires indépendantes de Bernouilli, nous laissons au lecteur la preuve que ce théorème implique le théorème de Moivre Laplace.

1. Dans cet exercice, on a $X_1,\ldots,X_n$ (avec $n = 49$) variables aléatoires indépendantes et de même loi. La loi de chaque $X_i$ est celle de $X$. On va appliquer le Théorème Central Limite. Pour cela, caculons $\mathbb{E}(X)$ et $\mathrm{var}X$. Puisque la fonction $x \mapsto xf(x)$ est intégrable sur $\mathbb{R}$, $\mathbb{E}(X)$ existe et vaut:

$$\mathbb{E}(X) = \int_0^\infty x^2 e^{-x}\mathrm{d}x$$

   Cette intégrale se calcule en faisant des intégrations par parties. On trouve $\mathbb{E}(X) = 2$.
   De même, $\mathbb{E}(X^2)$ existe car $x \mapsto x^2 f(x)$ est intégrable sur $\mathbb{R}$. De plus,

$$\begin{aligned}
\mathbb{E}(X^2) &= \int_0^\infty x^3 e^{-x}\mathrm{d}x \\
&= 6
\end{aligned}$$

   Par conséquent, $\mathrm{var}X = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 6 - 4 = 2$.
   Le théorème central limite nous dit que $\bar{X}_n$ suit approximativement une loi $\mathcal{N}(m,\sigma^2/n)$ avec $m = 2$, $n = 49$ et $\sigma^2 = 2$. C'est à dire $\mathcal{N}(2, 2/49)$.
2. Notons donc $Z$ une variable aléatoire de loi normale $\mathcal{N}(2, 2/49)$. Alors,

$$\mathbb{P}\left(1.8 \leq \bar{X}_n \leq 2.1\right) \simeq \mathbb{P}(1.8 \leq Z \leq 2.1)$$

   Puisque $Z \sim \mathcal{N}(2, 2/49)$, $Y = \sqrt{49/2}\,(Z - 2) \sim \mathcal{N}(0,1)$. Donc,

$$\mathbb{P}\left(1.8 \leq \bar{X}_n \leq 2.1\right) \simeq \mathbb{P}(-0.0404 \leq Y \leq 0.0202)$$

   Cette dernière probabilité se calcule à l'aide de la table.

$$\mathbb{P}(-0.0404 \leq Y \leq 0.0202) = F(0.0202) - F(-0.0404)$$

   où $F$ est la fonction de répartition de $Y$. De plus, puisque $Y$ suit une loi normale, $F(t) = 1 - F(-t)$ pour tout $t \in \mathbb{R}$. Donc,

$$\mathbb{P}(-0.0404 \leq Y \leq 0.0202) = F(0.0202) + F(0.0404) - 1$$

   La table donne $F(0.0202) \simeq 0.508$ et $F(0.0404) \simeq 0.516$. Finalement,

$$\mathbb{P}\left(1.8 \leq \bar{X}_n \leq 2.1\right) \simeq 0.024.$$

**Exercice 11:**
Changement d'énoncé: déterminer $n$ tel que $\mathbb{P}(0.05 \leq X/n \leq 0.15) \geq 0.95$.

On a donc $n$ variables aléatoires $X_1,\ldots X_n$ indépendantes et de meme loi. Dire que $X_i = 1$ signifie que le $i$ème produit est défectueux, et $X_i = 0$ s'il ne l'est pas. Alors $X_i$ obeit à une loi de Bernouilli $\mathcal{B}(p)$. D'après l'énoncé, la probabilité qu'un produit soit défectueux est $p = 0.1$.

De plus, le nombre de produits défectueux est donné par $X = \sum_{i=1}^n X_i$. D'après le Théorème Central Limite (ou le théorème de Moivre-Laplace), $X/n$ suit approximativement une loi $\mathcal{N}(0.1, 0.1(1-0.1)/n)$. Soit donc $Y$ variable aléatoire ayant pour loi $\mathcal{N}(0.1, 0.1(1-0.1)/n) = \mathcal{N}(0.1, 0.09/n)$. Alors

$$\begin{aligned}
\mathbb{P}(0.05 \leq X/n \leq 0.15) &\simeq \mathbb{P}(0.05 \leq Y \leq 0.15) \\
&\simeq \mathbb{P}\left(-0.1667\sqrt{n} \leq Z \leq 0.1667\sqrt{n}\right)
\end{aligned}$$

où $Z = \sqrt{n/0.09}(Y - 0.1)$ obeit à une loi normale $\mathcal{N}(0,1)$. Si $F$ est la fonction de répartition d'une loi $\mathcal{N}(0,1)$, on a donc:

$$\mathbb{P}\left(0.05 \leq X/n \leq 0.15\right) \simeq F\left(0.1667\sqrt{n}\right) - F\left(-0.1667\sqrt{n}\right) = 2F\left(0.1667\sqrt{n}\right) - 1$$

Il faut donc choisir $n$ tel que

$$2F\left(0.1667\sqrt{n}\right) - 1 \geq 0.95$$

cad

$$F\left(0.1667\sqrt{n}\right) \geq 0.975.$$

D'après la table si $x \simeq 1.96$ alors $F(x) \simeq 0.975$. On prend $n$ tel que $0.1667\sqrt{n} \geq 1.96$. Il faut donc que la taille soit au moins égale à $n = 13825$.