Université Paris Descartes
UFR de Mathématiques et Informatique

Probabilités et statistiques pour l'informatique - Licence MIA 2e année
Examen du 6 janvier 2015 - Correction

Exercice 1

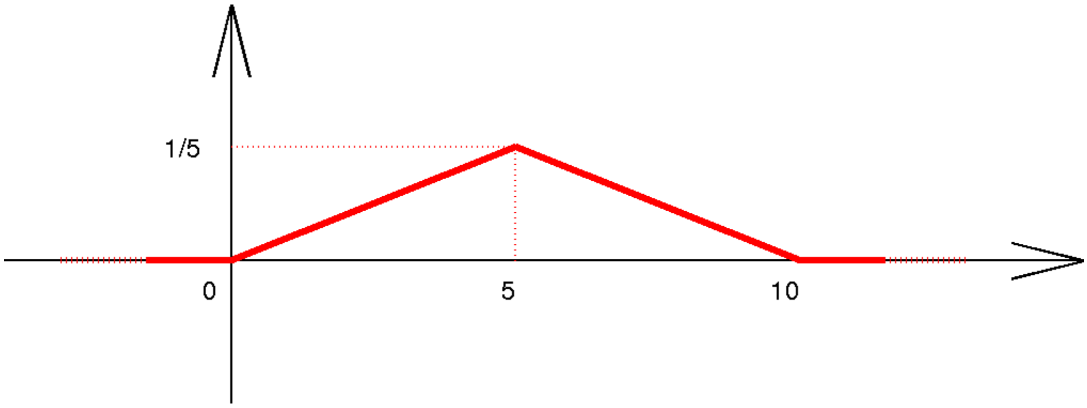
1. Les variables X_i sont indépendantes et toutes de même loi que X, et $E(X)$ est bien définie. Donc d'après la loi des grands nombres, la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ converge vers $\mu = E(X)$. Par conséquent $\bar{X}_n$ est un estimateur convergeant de μ.
   Il est sans biais car
   $$E(\bar{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n E(X_i) = \frac{1}{n} \times n \times \mu = \mu,$$
   par linéarité de l'espérance.
   D'après les valeurs numériques données dans l'énoncé on a $\bar{X}_n \frac{1}{100} \times 1.7712 \times 10^4 = 1.7712 \times 10^2$. La taille moyenne estimée est donc d'environ $177cm$.
2. On sait que $S_n = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}_n^2$ est un estimateur convergeant de σ^2. On a ici
   $$S_n = \frac{1}{100} \times 3.1416 \times 10^6 - (1.7712 \times 10^2)^2 \simeq 3.1416 \times 10^4 - 3.1371 \times 10^4 = 0.0045 \times 10^4 = 45.$$
   L'estimation de l'écart-type est donc $\sqrt{S_n} \simeq \sqrt{45} \simeq 6.71cm$.
3. Les variables X_i sont indépendantes et toutes de même loi que X, et $E(X) = \mu$ et $V(X) = \sigma^2$ sont bien définies. Le théorème central limite s'applique donc : pour tous $\alpha < \beta$, on a
   $$P\left(\alpha \leq \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \leq \beta\right) \xrightarrow[n\to\infty]{} P(\alpha \leq Z \leq \beta),$$
   où Z est une variable de loi normale $\mathcal{N}(0,1)$. On a donc, pour $n = 100$, $\alpha = -1.96$ et $\beta = 1.96$,
   $$P\left(-1.96 \leq \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \leq 1.96\right) \simeq P(-1.96 \leq Z \leq 1.96) \simeq 0.95,$$
   soit encore
   $$P\left(-1.96\frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq 1.96\frac{\sigma}{\sqrt{n}}\right) \simeq 0.95,$$
   $$P\left(\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right) \simeq 0.95.$$
   On obtient ainsi un intervalle de confiance au niveau 95% pour μ :
   $$\left[\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}} \ , \ \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right].$$
   Sa valeur numérique est ici :
   $$\begin{aligned}\left[177.12 - 1.96\frac{7}{\sqrt{100}} \ , \ 177.12 + 1.96\frac{7}{\sqrt{100}}\right] &= \left[177.12 - \frac{13.72}{10} \ , \ 177.12 - \frac{13.72}{10}\right] \\ &= [177.12 - 1.372 \ , \ 177.12 + 1.372] \simeq [175.7 \ , \ 178.5].\end{aligned}$$

4. L'erreur sur la valeur de μ est au maximum de $1.37cm$ au niveau de confiance 95%. C'est supérieur à 1 cm, donc on ne peut pas dire qu'on a estimé la taille moyenne avec une erreur inférieure à 1 cm pour ce niveau de confiance. Pour obtenir une estimation plus précise, il faudrait mesurer plus que 100 personnes.

Exercice 2 *Les trois parties de l'exercice sont très largement indépendantes.*

<u>*Partie 1*</u> *(4 points)*

1. .



2. f est une fonction positive, et

$$\begin{aligned}
\int_{\mathbb{R}} f(x)dx &= \int_0^5 \frac{x}{25}dx + \int_5^{10}\left(-\frac{x}{25}+\frac{2}{5}\right)dx \\
&= \frac{1}{25}\left[\frac{x^2}{2}\right]_0^5 - \frac{1}{25}\left[\frac{x^2}{2}\right]_5^{10} + \frac{2}{5}(10-5) \\
&= \frac{1}{25}\frac{25}{2} - \frac{1}{25}\left(\frac{100}{2}-\frac{25}{2}\right)+2 \\
&= \frac{1}{2} - 2 + \frac{1}{2} + 2 = 1.
\end{aligned}$$

   On a donc bien $\int_{\mathbb{R}} f = 1$, et donc f est bien une densité de probabilité.

3. On peut voir directement que le graphe de f est symétrique par rapport à l'axe vertical d'abscisse 5, autrement dit que $f(5+x) = f(5-x)$ pour tout x, ce qui implique que nécessairement $E(Z) = 5$.

Si on veut effectuer le calcul, ça donne :

$$\begin{aligned}
E(Z) &= \int_{\mathbb{R}} xf(x)dx = \int_0^5 \frac{x^2}{25}dx + \int_5^{10}\left(-\frac{x^2}{25}+\frac{2}{5}x\right)dx \\
&= \frac{1}{25}\left[\frac{x^3}{3}\right]_0^5 - \frac{1}{25}\left[\frac{x^3}{3}\right]_5^{10} + \frac{2}{5}\left[\frac{x^2}{2}\right]_5^{10} \\
&= \frac{1}{25}\frac{5^3}{3} - \frac{1}{25}\left(\frac{10^3}{3}-\frac{5^3}{3}\right) + \frac{2}{5}\left(\frac{10^2}{2}-\frac{5^2}{2}\right) \\
&= \frac{5}{3} - \left(\frac{40}{3}-\frac{5}{3}\right) + (20-5) \\
&= \frac{5}{3} - \frac{40}{3} + \frac{5}{3} + 20 - 5 \\
&= -10 + 20 - 5 = 5.
\end{aligned}$$

4. La fonction de répartition F_Z de la variable Z est définie par $F_Z(x) = P(Z \le x)$ pour tout $x \in \mathbb{R}$. On a ici

$$F_Z(x) = P(Z \le x) = \int_{-\infty}^x f(t)dt.$$

On doit distinguer suivant les valeurs de x.
– Si $x \le 0$. Alors $\int_{-\infty}^x f(t)dt = 0$.
– Si $0 \le x \le 5$, alors

$$\begin{aligned}
\int_{-\infty}^x f(t)dt &= \int_0^x \frac{t}{25}dt = \frac{1}{25}\left[\frac{x^2}{2}\right]_0^x \\
&= \frac{1}{25}\frac{x^2}{2} = \frac{x^2}{50}.
\end{aligned}$$

– Si $5 \le x \le 10$, alors

$$\begin{aligned}
\int_{-\infty}^x f(t)dt &= \int_0^5 \frac{t}{25}dt + \int_5^x\left(-\frac{t}{25}+\frac{2}{5}\right)dt \\
&= \frac{1}{25}\left[\frac{t^2}{2}\right]_0^5 - \frac{1}{25}\left[\frac{t^2}{2}\right]_5^x + \frac{2}{5}(x-5) \\
&= \frac{1}{25}\frac{5^2}{2} - \frac{1}{25}\left(\frac{x^2}{2}-\frac{5^2}{2}\right) + \frac{2}{5}(x-5) \\
&= \frac{1}{2} - \frac{x^2}{50} + \frac{1}{2} + \frac{2}{5}x - 2 \\
&= -\frac{x^2}{50} + \frac{2}{5}x - 1.
\end{aligned}$$

– Si $x \ge 10$ alors $\int_{-\infty}^x f(t)dt = 1$.
Au final on a donc

$$F_Z(x) = \begin{cases} 0 \text{ si } x \le 0 \\ \frac{x^2}{50} \text{ si } 0 \le x \le 5 \\ -\frac{x^2}{50} + \frac{2}{5}x - 1 \text{ si } 5 \le x \le 10 \\ 1 \text{ si } x \ge 10. \end{cases}$$

Partie 2

5.
```
Temps_Attente = function()
{
  # temps d'attente de l'ascenseur au premier tour
  X = runif(1,0,5)
  # on tire au sort pour savoir si l'étudiant passe au premier ou deuxième tour
  tour = sample(c(1,2),1) # tour vaut 1 ou 2 avec proba 1/2 pour chacun
  if(tour==1)
    T = X
  else
  {
    # temps d'attente de l'ascenseur au deuxième tour
    Y = runif(1,0,5)
    T = X + Y
  }
  T
}
```
   On peut écrire aussi cette fonction plus simplement : on fait la somme de n variables uniformes sur [0,5], n valant 1 ou 2 avec proba 1/2 pour chaque
```
Temps_Attente = function()
{
  n = sample(1:2,1)
  sum(runif(n,0,5))
}
```
   Ou encore en une seule ligne :
```
Temps_Attente = function()
{
  sum(runif(sample(1:2,1),0,5))
}
```

6.
```
T = rep(0,100)
for(i in 1:100)
  T[i] = Temps_Attente()
sum(T>6)
```

Partie 3

7. Notons D l'événement "l'étudiant passe au deuxième tour". On a $P(D) = 1/2$, et on peut écrire grâce à la formule des probabilités totales :
$$\begin{aligned}P(T \leq t) &= P(T \leq t|D^c)P(D^c) + P(T \leq t|D)P(D) \\ &= P(X \leq t)P(D^c) + P(X + Y \leq t)P(D) \\ &= \frac{1}{2}P(X \leq t) + \frac{1}{2}P(X + Y \leq t)\end{aligned}$$

8. On a
$$\begin{aligned}F_T(t) &= P(T \leq t) = \frac{1}{2}P(X \leq t) + \frac{1}{2}P(Z \leq t) \\ &= \frac{1}{2}F_X(t) + \frac{1}{2}F_Z(t).\end{aligned}$$

X suit la loi uniforme sur $[0, 5]$ donc sa densité s'écrit

$$f_X(x) = \begin{cases} \frac{1}{5} \text{ si } 0 \leq x \leq 5 \\ 0 \text{ sinon.} \end{cases}$$

Sa fonction de répartition vaut $F_X(t) = P(X \leq t) = \int_{-\infty}^{t} f_X(x)dx$.
- Si $t \leq 0$ alors $F_X(t) = 0$.
- Si $0 \leq t \leq 5$ alors $F_X(t) = \int_0^t \frac{1}{5}dx = \frac{1}{5}t$.
- Si $t \geq 5$ alors $F_X(t) = 1$.

On obtient donc finalement

$$F_T(t) = \frac{1}{2}F_X(t) + \frac{1}{2}F_Z(t) = \begin{cases} 0 \text{ si } t \leq 0 \\ \frac{t^2}{100} + \frac{t}{10} \text{ si } 0 \leq t \leq 5 \\ -\frac{t^2}{100} + \frac{1}{5}t - \frac{1}{2} \text{ si } 5 \leq t \leq 10 \\ 1 \text{ si } t \geq 10. \end{cases}$$

9. F_T est une fonction continue et dérivable sauf en quelques points ; donc T est une variable à densité et sa densité est égale à la dérivée de F_T aux points où elle est dérivable. Les valeurs aux autres points peuvent être choisies arbitrairement. On a donc

$$f_T(t) = F_T'(t) = \begin{cases} 0 \text{ si } t < 0 \\ \frac{t}{50} + \frac{1}{10} \text{ si } 0 \leq t < 5 \\ -\frac{t}{50} + \frac{1}{5} \text{ si } 5 \leq t < 10 \\ 0 \text{ si } t \geq 10. \end{cases}$$