Echantillonnage et estimation.

Corrigés d'exercices

Version du 12/06/2014

Les exercices du livre corrigés dans ce document sont les suivants :

N°22 page 419

1. Notons respectivement $\text{IF}\left(p=0,6\right)$ et $\text{IF}\left(p=0,5\right)$ les intervalles de fluctuation cherchés. Ici $n=100$ et il vient :

$$\begin{aligned}\text{IF}\left(p=0,6\right) &= \left[0,6-1,96\frac{\sqrt{0,6\times 0,4}}{\sqrt{100}}\,;\,0,6+1,96\frac{\sqrt{0,6\times 0,4}}{\sqrt{100}}\right]\\ &= \left[0,504\,0\,;\,0,696\,0\right]\end{aligned}$$

$$\begin{aligned}\text{IF}\left(p=0,5\right) &= \left[0,5-1,96\frac{\sqrt{0,5\times 0,5}}{\sqrt{100}}\,;\,0,5+1,96\frac{\sqrt{0,5\times 0,5}}{\sqrt{100}}\right]\\ &= \left[0,402\,;\,0,598\right]\end{aligned}$$

Puisque l'on s'intéresse à $p=0,6$ et $p=0,5$ on met donc l'accent sur les pierres rouges (les rubis) puisqu'il y en a théoriquement 60% dans chaque sac sauf dans celui du contrebandier où cette proportion vaut 50%.

2. *a*. On suppose que les douaniers ont prélevé 40 pierres rouges dans un sac.
La proportion de pierres rouges est donc de $40\%=0,4$.

Cette valeur n'appartient pas à $\text{IF}\left(p=0,6\right)$. Elle est même nettement inférieure à la borne inférieure de cet intervalle. En outre, elle est très proche de la borne inférieure de $\text{IF}\left(p=0,5\right)$.

On peut donc raisonnablement conclure :

$\boxed{\text{Le sac dans lequel le prélèvement a été effectué est celui du contrebandier.}}$

b. On suppose que les douaniers ont prélevé 55 pierres rouges dans un sac.
La proportion de pierres rouges est donc cette fois de $55\% = 0,55$.

Cette valeur appartient aux deux intervalles de fluctuation et on ne peut donc conclure !

La valeur 0,55 appartenant aux deux intervalles de fluctuation, on ne peut conclure !

3. $\mathrm{IF}\left(p=0,6\right)$ et $\mathrm{IF}\left(p=0,5\right)$ sont des intervalles respectivement centrés sur … $p=0,6$ et $p=0,5$. Pour *n* fixé, leurs longueurs valent : $2\times 1,96\dfrac{\sqrt{0,6\times 0,4}}{\sqrt{n}}$ et $2\times 1,96\dfrac{\sqrt{0,5\times 0,5}}{\sqrt{n}}$.
Ainsi, pour n suffisamment grand, la borne supérieure de $\mathrm{IF}\left(p=0,5\right)$ sera strictement inférieure à la borne inférieure de $\mathrm{IF}\left(p=0,6\right)$. On va donc résoudre :

$$0,5+1,96\frac{\sqrt{0,5\times 0,5}}{\sqrt{n}}<0,6-1,96\frac{\sqrt{0,6\times 0,4}}{\sqrt{n}}$$

On a :

$$\begin{aligned}
&0,5+1,96\frac{\sqrt{0,5\times 0,5}}{\sqrt{n}}<0,6-1,96\frac{\sqrt{0,6\times 0,4}}{\sqrt{n}}\\
&\Leftrightarrow 1,96\frac{\sqrt{0,5\times 0,5}}{\sqrt{n}}+1,96\frac{\sqrt{0,6\times 0,4}}{\sqrt{n}}<0,1\\
&\Leftrightarrow 19,6\left(\sqrt{0,5\times 0,5}+\sqrt{0,6\times 0,4}\right)<\sqrt{n}\\
&\Leftrightarrow n>19,6^2\left(\sqrt{0,5\times 0,5}+\sqrt{0,6\times 0,4}\right)^2
\end{aligned}$$

Or : $19,6^2\left(\sqrt{0,5\times 0,5}+\sqrt{0,6\times 0,4}\right)^2 \simeq 376,4$.
On choisit donc : $n=377$.

Pour que les intervalles de confiance soient disjoints, il convient de prélever au moins 377 pierres.

Pour cette valeur de *n*, les intervalles de confiances s'écrivent :

$$\begin{aligned}
\mathrm{IF}\left(p=0,6\right)&=\left[0,6-1,96\frac{\sqrt{0,6\times 0,4}}{\sqrt{377}}\,;\,0,6+1,96\frac{\sqrt{0,6\times 0,4}}{\sqrt{377}}\right]\\
&=\left[0,550\,55\,;\,0,649\,45\right]\\
\mathrm{IF}\left(p=0,5\right)&=\left[0,5-1,96\frac{\sqrt{0,5\times 0,5}}{\sqrt{377}}\,;\,0,5+1,96\frac{\sqrt{0,5\times 0,5}}{\sqrt{377}}\right]\\
&=\left[0,449\,53\,;\,0,550\,47\right]
\end{aligned}$$

N°23 page 420

Ici, on a $p = 0,03$.

1. Ici, on a $n = 200$.
   L'intervalle de fluctuation asymptotique au seuil de 95% s'écrit :

$$\begin{aligned}\text{IF} &= \left[0,03 - 1,96\frac{\sqrt{0,03\times 0,97}}{\sqrt{200}}\ ;\ 0,03 + 1,96\frac{\sqrt{0,03\times 0,97}}{\sqrt{200}}\right]\\ &= \left[0,006\,358\ ;\ 0,053\,642\right]\end{aligned}$$

$$\boxed{\text{IF} = \left[0,006\,358\ ;\ 0,053\,642\right]}$$

2. $n = 200 \geq 30$, $np = 200\times 0,03 = 6 \geq 5$ et $n(1-p) = 200\times 0,97 = 194 \geq 5$.

Les conditions permettant d'utiliser l'intervalle de fluctuation pour une prise de décision sont satisfaites.

3. Avec 14 excès de vitesse, on obtient une proportion observée égale à $\frac{14}{200} = 0,07$.

   Cette valeur n'appartient pas à l'intervalle IF et on peut raisonnablement en conclure (risque de 5% d'erreur) qu'il y a recrudescence des excès de vitesse et qu'il convient d'installer un panneau clignotant supplémentaire.

On peut raisonnablement conclure qu'il y a recrudescence des excès de vitesse et qu'il convient d'installer un panneau clignotant supplémentaire.

4. Ici $n = 500$.

   a. On a cette fois :

$$\begin{aligned}\text{IF} &= \left[0,03 - 1,96\frac{\sqrt{0,03\times 0,97}}{\sqrt{500}}\ ;\ 0,03 + 1,96\frac{\sqrt{0,03\times 0,97}}{\sqrt{500}}\right]\\ &= \left[0,015\,047\ ;\ 0,044\,953\right]\end{aligned}$$

$$\boxed{\text{IF} = \left[0,015\,047\ ;\ 0,044\,953\right]}$$

   b. L'installation d'un panneau supplémentaire ne sera pas décidée si la proportion d'excès de vitesse observée appartient à l'intervalle précédent.

Si on note N le nombre d'excès de vitesse observés, on doit donc avoir :

$$\frac{N}{500} \in \text{IF} = \left[0,015\,047\;;\;0,044\,953\right]$$

On a :

$$\begin{aligned} & \frac{N}{500} \in \text{IF} = \left[0,015\,047\;;\;0,044\,953\right] \\ \Leftrightarrow\ & 0,015\,047 \le \frac{N}{500} \le 0,044\,953 \\ \Leftrightarrow\ & 7,523\,5 \le N \le 22,476\,5 \end{aligned}$$

L'installation du panneau supplémentaire ne sera pas décidée
si on observe entre 8 et 22 excès de vitesse sur les 500 véhicules contrôlés.

N°26 page 421

1. *a*. Avec $f = 49\% = 0,49$ et $n = 500$, l'intervalle de confiance au niveau 95% s'écrit :

$$\text{IC} = \left[f - \frac{1}{\sqrt{n}}\;;\; f + \frac{1}{\sqrt{n}} \right] = \left[0,49 - \frac{1}{\sqrt{500}}\;;\; 0,49 + \frac{1}{\sqrt{500}} \right] = \left[0,445\,3\;;\;0,534\,7\right]$$

La proportion d'habitants ayant voté pour Louis Victorien peut être estimée au niveau 95% par l'intervalle de confiance :

$$\text{IC} = \left[0,445\,3\;;\;0,534\,7\right]$$

b. D'après l'intervalle précédent, on ne peut en rien garantir la défaite de Louis Victorien une part significative des valeurs de l'intervalle IC étant supérieure à 0,5.

On ne peut garantir la défaite de Louis Victorien.

2. On continue de raisonner avec $f = 49\% = 0,49$.
On pourra raisonnablement conclure à la défaite de Louis Victorien si la borne supérieure de l'intervalle de confiance est strictement inférieure à 0,5. On va donc résoudre :

$$0,49 + \frac{1}{\sqrt{n}} < 0,5$$

On a :

$$0,49 + \frac{1}{\sqrt{n}} < 0,5 \Leftrightarrow \frac{1}{\sqrt{n}} < 0,01 \Leftrightarrow \sqrt{n} > 100 \Leftrightarrow n > 10\,000$$

Pour une fréquence observée $f = 49\% = 0,49$, on aurait raisonnablement pu prévoir la défaite de Louis Victorien si 10 000 électeurs avaient été interrogés.

Ce résultat est TRES intéressant ! Il montre que plus une telle élection est serrée (les pourcentages des deux candidats sont proches de 50%) plus il faudrait interroger d'électeurs pour pouvoir prévoir de façon fiable le résultat final. Si on reprend la situation de l'exercice, il faut réaliser qu'interroger 10 000 électeurs coûte très cher (de tels sondages, en France, se font en général auprès de 1 000 électeurs environ !) te ne sera pas envisagé ! ☺ Mieux vaut donc attendre tranquillement la fin du dépouillement …

N°32 page 422

1. Pour *k* entier compris entre 1 et 6, on a : $\frac{4}{39} \le p_k \le \frac{9}{39}$ et donc $p_k \in [0\,;1]$.

   Les réels p_k peuvent être des probabilités.

   De surcroît, on a :

$$\sum_{k=1}^{6} p_k = \sum_{k=1}^{6} \frac{3+k}{39} = \sum_{k=1}^{6} \frac{3}{39} + \sum_{k=1}^{6} \frac{k}{39} = 6\times\frac{3}{39} + \frac{1}{39}\times\sum_{k=1}^{6} k = \frac{18}{39} + \frac{1}{39}\times\frac{6\times 7}{2} = \frac{18}{39} + \frac{21}{39} = \frac{39}{39} = 1$$

   Ainsi ;

Les réels p_k sont bien des probabilités élémentaires.

2. *a*. On a :

$$P(A) = P(\{2\,;4\,;6\}) = p_2 + p_4 + p_6 = \frac{3+2}{39} + \frac{3+4}{39} + \frac{3+6}{39} = \frac{5+7+9}{39} = \frac{21}{39} = \frac{7}{13}.$$

$$P(B) = P(\{3\,;4\,;5\,;6\}) = p_3 + p_4 + p_5 + p_6 = \frac{3+3}{39} + \frac{3+4}{39} + \frac{3+5}{39} + \frac{3+6}{39} = \frac{6+7+8+9}{39} = \frac{30}{39} = \frac{10}{13}$$

$$P(C) = P(\{3\,;4\}) = p_3 + p_4 = \frac{3+3}{39} + \frac{3+4}{39} = \frac{6+7}{39} = \frac{13}{39} = \frac{1}{3}$$

$P(A) = \frac{7}{13}$, $P(B) = \frac{10}{13}$ et $P(C) = \frac{1}{3}$.

   b. On cherche ici : $P_A(B)$.

   On a : $P_A(B) = \frac{P(A\cap B)}{P(A)}$.

   Comme $A = \{2\,;4\,;6\}$ et $B = \{3\,;4\,;5\,;6\}$ on a : $A\cap B = \{4\,;6\}$.

D'où : $P\left(A\cap B\right)=P\left(\left\{4\,;6\right\}\right)=p_4+p_6=\dfrac{3+4}{39}+\dfrac{3+6}{39}=\dfrac{7+9}{39}=\dfrac{16}{39}$, puis :

$$P_A\left(B\right)=\frac{P\left(A\cap B\right)}{P\left(A\right)}=\frac{\frac{16}{39}}{\frac{7}{13}}=\frac{16}{39}\times\frac{13}{7}=\frac{16}{3\times 7}=\frac{16}{21}$$

La probabilité que le nombre obtenu soit supérieur ou égal à 3 sachant qu'il est pair est égale à $\dfrac{16}{21}$.

c. On a d'une part : $P\left(A\cap B\right)=\dfrac{16}{39}$.

On a d'autre part : $P\left(A\right)\times P\left(B\right)=\dfrac{7}{13}\times\dfrac{10}{13}=\dfrac{70}{169}$.

Comme $P\left(A\cap B\right)=\dfrac{16}{39}$ n'est pas égale à $P\left(A\right)\times P\left(B\right)=\dfrac{70}{169}$, on en conclut immédiatement que les événements *A* et *B* ne sont pas indépendants.

On procède comme précédemment avec les événements *A* et *C*.
On a : $A=\left\{2\,;4\,;6\right\}$ et $C=\left\{3\,;4\right\}$.

D'où : $A\cap C=\left\{4\right\}$ et donc $P\left(A\cap C\right)=P\left(\left\{4\right\}\right)=p_4=\dfrac{3+4}{39}=\dfrac{7}{39}$.

Par ailleurs : $P\left(A\right)\times P\left(C\right)=\dfrac{7}{13}\times\dfrac{1}{3}=\dfrac{7}{39}$.

Comme $P\left(A\cap C\right)=P\left(A\right)\times P\left(C\right)$, on en conclut immédiatement que les événements *A* et *C* sont indépendants.

Les événements *A* et *B* ne sont pas indépendants. Les événements *A* et *C* sont indépendants.

3. Comme l'obtention d'un nombre pair correspond à l'événement *A*, pour un tirage, l'espérance du gain vaut : $1\times P\left(A\right)+0\times P\left(\overline{A}\right)=P\left(A\right)=\dfrac{7}{13}$.

   Si on lance le dé 60 fois, l'espérance *E* du gain total vaudra : $E=60\times\dfrac{7}{13}=\dfrac{420}{13}\simeq 32,3$.

L'espérance totale du gain total est de $\dfrac{420}{13}$ € soit environ 32,3€.

4. *a*. Des nombres pairs ont été obtenus $6+11+16=33$ fois. Le gain total du joueur est donc de 33€.

Le gain total du joueur est de 33€.

b. La fréquence observée de l'événement *A* vaut : $\frac{33}{60}=\frac{11}{20}=0,55$.

A la question 2, on avait obtenu : $P(A)=\frac{7}{13}=0,538$.

La fréquence observée de l'événement *A* est compatible avec la probabilité de cet événement.

c. On a ici : $f=0,55$ et $n=60$. D'où, en utilisant $\left[f-\frac{1}{\sqrt{n}}\,;\,f+\frac{1}{\sqrt{n}}\right]$:

$$\begin{aligned}\left[f-\frac{1}{\sqrt{n}}\,;\,f+\frac{1}{\sqrt{n}}\right]&=\left[0,55-\frac{1}{\sqrt{60}}\,;\,0,55+\frac{1}{\sqrt{60}}\right]\\&=\left[0,420\,9\,;\,0,679\,1\right]\end{aligned}$$

Un intervalle de confiance au niveau de confiance de 0,95 de la probabilité de l'événement *A* est : $\left[0,420\,9\,;\,0,679\,1\right]$.

N°34 page 423

La variable aléatoire *X* suit la loi $\mathscr{N}\left(10\,;\,0,03^2\right)$.

1. *a*. Soit *A* l'événement « Le rouleau est accepté ».
On a ici : $P(A)=P(X\geq 9,95)=1-P(X<9,95)\simeq 0,95$.

La probabilité qu'un rouleau pris au hasard dans la production soit accepté est égale à $0,95$ (valeur approchée à 10^{-2}).

b. Soit *R* l'événement « Le rouleau est refusé ».
On a immédiatement : $R=\overline{A}$ puis : $P(R)=P\left(\overline{A}\right)=1-P(A)=P(X<9,95)\simeq 0,05$.

La probabilité qu'un rouleau pris au hasard dans la production soit refusé est égale à $0,05$ (valeur approchée à 10^{-2}).

2. On utilise l'intervalle $\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right]$ avec $f = \frac{125}{20\,000} = 0,006\,25 = 6,25 \times 10^{-3}$ et $n = 20\,000$:

$$\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right] = \left[6,25 \times 10^{-3} - \frac{1}{\sqrt{20\,000}} ; 6,25 \times 10^{-3} + \frac{1}{\sqrt{20\,000}} \right]$$
$$= \left[-0,000\,8 ; 0,013\,3 \right]$$

La proportion ne pouvant être négative, on retiendra : $\left[0 ; 0,013\,3 \right]$.

> Un intervalle de confiance au niveau 0,95 pour la proportion des plaques de la livraison qui sont sans défaut est : $\left[0,84 ; 1 \right]$.

N°40 page 425

Partie A – Loi binomiale

1. L'événement E correspond au « SUCCES » d'une épreuve de Bernoulli (la plaque est défectueuse ou ne l'est pas) de paramètre $p = P(\mathrm{E}) = 0,02$.

   Le prélèvement pouvant être assimilé à un tirage avec remise de 50 plaques, ayant toutes la même probabilité p d'être défectueuses et dont les états sont indépendants, il peut être assimilé à un schéma de Bernoulli. La variable aléatoire X, qui comptabilise le nombre de « SUCCES » (i.e. le nombre de plaques défectueuses) suit donc la loi binomiale de paramètres $n = 50$ et $p = P(\mathrm{E}) = 0,02$.

> La variable aléatoire X suit la loi binomiale $\mathscr{B}(50 ; 0,02)$.

2. D'après ce qui précède :

$$P(X = 0) = \binom{50}{0} \times 0,02^0 \times 0,98^{50} = 0,98^{50} \simeq 0,364$$
$$P(X = 1) = \binom{50}{1} \times 0,02^1 \times 0,98^{49} = 50 \times 0,02 \times 0,98^{49} = 0,98^{49} \simeq 0,372$$

> $P(X = 0) \simeq 0,364$ et $P(X = 1) \simeq 0,372$
> (valeurs arrondies à 10^{-3})

3. On cherche ici : $P(X \leq 2)$.
   On a : $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$.
   $P(X = 0)$ et $P(X = 1)$ ont été calculées à la question précédente. Il reste à calculer $P(X = 2)$.

   On a : $P(X = 2) = \binom{50}{2} \times 0,02^2 \times 0,98^{48} = \frac{50 \times 49}{2} \times 0,02^2 \times 0,98^{48} = 0,49 \times 0,98^{48}$ (une valeur approchée ne nous intéresse pas ici).

   Il vient alors :

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= 0,98^{50} + 0,98^{49} + 0,49 \times 0,98^{48} \\ &= 0,98^{48} \times \left(0,98^2 + 0,98 + 0,49\right) \\ &= 2,4304 \times 0,98^{48} \\ &\simeq 0,922 \end{aligned}$$

La probabilité que au plus deux plaques soient défectueuses est égale à :
$P(X \leq 2) = 2,4304 \times 0,98^{48} \simeq 0,922$

Partie B – Loi normale

1. Pour la variable aléatoire L_1, on a : $\mu_1 = 550$ et $\sigma_1 = 1$.
   Ainsi, il vient immédiatement : $P(548 < L_1 < 552) = P(\mu_1 - 2\sigma_1 < L_1 < \mu_1 + 2\sigma_1) \simeq 0,954$.

$P(548 < L_1 < 552) = P(\mu_1 - 2\sigma_1 < L_1 < \mu_1 + 2\sigma_1) \simeq 0,954$

2. Ici, on a : $P(108 < L_2 < 112) = 0,95$.
   Remarque : pour deux variables aléatoires *X* et *Y* continues sur $\mathbb{R}$, l'indépendance se traduit par : $P(X \in [a\,;b] \cap Y \in [c\,;d]) = P(X \in [a\,;b]) \times P(Y \in [c\,;d])$.

   Pour que la plaque soit conforme pour la longueur et pour la largeur, il faut $548 < L_1 < 552$ et $108 < L_2 < 112$.
   En tenant compte de l'indépendance de L_1 et L_2, il vient alors :

$$P\left((548 < L_1 < 552) \cap (108 < L_2 < 112)\right) = P(548 < L_1 < 552) \times P(108 < L_2 < 112) \simeq 0,907$$

Partie C – Intervalle de confiance

On utilise l'intervalle $\left[f-\frac{1}{\sqrt{n}}\,;\, f+\frac{1}{\sqrt{n}} \right]$ avec $f=\frac{94}{100}=0,94$ et $n=100$:

$$\left[f-\frac{1}{\sqrt{n}}\,;\, f+\frac{1}{\sqrt{n}} \right]=\left[0,94-\frac{1}{\sqrt{100}}\,;\,0,94+\frac{1}{\sqrt{100}} \right]=\left[0,84\,;1,04\right]$$

La proportion ne pouvant être supérieure à 1, on retiendra : $\left[0,84\,;1\right]$.

Un intervalle de confiance au niveau 0,95 pour la proportion des plaques de la livraison qui sont sans défaut est : $\left[0,84\,;1\right]$.