Corrigé 10

Exercice 1. (a) On voit bien qu'il s'agit d'un intervalle de confiance : les bornes sont des variables aléatoires qui ne dépendent pas du paramètre inconnu. Il faut donc montrer que la confiance de I_n est de 95 %, c'est-à-dire que

$$\mathbb{P}(M_n \leq \theta \leq 0.05^{-1/n} \cdot M_n) = 0.95.$$

Notez que ce sont les bornes de l'intervalle, M_n et $0.5^{-1/n} \cdot M_n$, qui sont aléatoires, et non pas θ. Donc la probabilité ci-dessus est la probabilité que l'intervalle couvre θ, plutôt que la probabilité que θ soit dans l'intervalle.

Calculons donc

$$\begin{aligned}\mathbb{P}(M_n \leq \theta \leq 0.05^{-1/n} \cdot M_n) &= \mathbb{P}(\{M_n \leq \theta\} \cap \{\theta \leq 0.05^{-1/n} \cdot M_n\}) \\ &= \mathbb{P}(\{M_n \leq \theta\} \cap \{0.05^{1/n} \cdot \theta \leq M_n\}) \\ &= \mathbb{P}(0.05^{1/n} \cdot \theta \leq M_n \leq \theta).\end{aligned}$$

Dans la série précédente on a trouvé la fonction de répartition de M_n,

$$F_{M_n}(x) = P(M_n \leq x) = \left(\frac{x}{\theta}\right)^n \qquad \text{pour } x \in [0, \theta].$$

Cela nous dit que

$$\mathbb{P}(0.05^{1/n} \cdot \theta \leq M_n \leq \theta) = \left(\frac{\theta}{\theta}\right)^n - \left(\frac{0.05^{1/n} \cdot \theta}{\theta}\right)^n = 1 - 0.05 = 0.95,$$

ce qui est la confiance désirée.

(b) On voit que la longueur $D_n = (0.05^{-1/n} - 1) \cdot M_n$. L'espérance de D_n est donnée par

$$\mathbb{E}[D_n] = \mathbb{E}[(0.05^{-1/n} - 1) \cdot M_n] = (0.05^{-1/n} - 1)\mathbb{E}[M_n] = (0.05^{-1/n} - 1)\frac{n}{n+1} \cdot \theta,$$

où l'espérance de M_n a été calculée dans la série précédente.

$$\lim_{n\to\infty} \mathbb{E}[D_n] = \lim_{n\to\infty} \left[(0.05^{-1/n} - 1)\frac{n}{n+1} \cdot \theta\right] = 0.$$

L'espérance de la longueur de I_n converge vers zéro lorsque $n \to \infty$, donc la précision de l'estimation croît en fonction de n.

(c) L'intervalle est $(298158, 346336.26)$.

Exercice 2. (a) Si $X_1, \ldots, X_n$ sont iid $\mathrm{N}(\mu, \sigma^2)$, alors

$$Z_n = \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \sim \mathrm{N}(0, 1).$$

On sait que

$$\mathbb{P}(-z < Z_n < z) = 2\,\Phi(z) - 1,$$

pour chaque constante $z > 0$. Donc si on choisit $z(1-\alpha/2) = \Phi^{-1}(1-\alpha/2)$ (le $(1-\alpha/2)-$quantile de la loi $N(0,1)$), on obtient que

$$\begin{aligned}1-\alpha &= \mathbb{P}(-z(1-\alpha/2) < Z_n < z(1-\alpha/2)) = \\ &= \mathbb{P}\left(-z(1-\alpha/2) < \sqrt{n}\,\frac{\bar{X}_n-\mu}{\sigma} < z(1-\alpha/2)\right) = \\ &= \mathbb{P}\left(\bar{X}_n - \frac{z(1-\alpha/2)}{\sqrt{n}}\,\sigma < \mu < \bar{X}_n + \frac{z(1-\alpha/2)}{\sqrt{n}}\,\sigma\right).\end{aligned}$$

L'intervalle

$$\left(\bar{X}_n - \frac{z(1-\alpha/2)}{\sqrt{n}}\,\sigma,\ \bar{X}_n + \frac{z(1-\alpha/2)}{\sqrt{n}}\,\sigma\right)$$

couvre donc la vraie valeur de μ avec la probabilité $1-\alpha$.

Dans notre cas, $\sigma^2 = 3.5$, $n = 12$, $\bar{x}_{12} = 13.31$, et $\alpha = 0.05$. On peut trouver dans le tableau de la fonction de répartition de la loi normale que $\Phi^{-1}(0.975) = 1.96$. L'intervalle cherché est donc $(12.25,\ 14.37)$.

(b) Si $X_1, \ldots, X_n$ sont iid $\mathrm{N}(\mu, \sigma^2)$, alors

$$T_{n-1} = \sqrt{n}\,\frac{\bar{X}_n - \mu}{S_n} \sim \mathrm{t}_{n-1},$$

où $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$, et t_ν est la loi de Student avec ν degrés de liberté. La loi de Student est symétrique autour de zéro de la même manière que la loi $\mathrm{N}(0,1)$. On a donc

$$\mathbb{P}(-t_{n-1}(1-\alpha/2) < T_{n-1} < t_{n-1}(1-\alpha/2)) = 1-\alpha,$$

où $t_\nu(1-\alpha/2)$ est $(1-\alpha)-$quantile de la loi t_ν. De la même manière que dans la partie (a) on obtient l'intervalle de confiance sous la forme

$$\left(\bar{X}_n - \frac{t_{n-1}(1-\alpha/2)}{\sqrt{n}}\,S_n,\ \bar{X}_n + \frac{t_{n-1}(1-\alpha/2)}{\sqrt{n}}\,S_n\right).$$

Dans notre situation on a $n = 12$, $\alpha = 0.05$, $s^2 = 3.69$, et on peut trouver dans le tableau que $t_{11}(0.975) = 2.201$. L'intervalle cherché est $(12.09,\ 14.53)$. On note que cet intervalle est plus long que celui de la partie (a), ceci vient du fait que nous avons maintenant deux paramètres à estimer, l'incertitude est donc plus grande que dans le cas où un seul paramètre est à estimer.

(c) On trouve que $t_{11}(0.95) = 1.796$, et que l'intervalle cherché est $(12.31, 14.31)$. Cet intervalle est plus petit que celui calculé dans la partie (b), car son seuil de confiance est plus petit. Plus on veut être confiant qu'un intervalle couvre la vraie valeur de μ, plus cet intervalle doit être grand (et vice-versa).

(d) Récolter plus de données. Plus on a de données, plus notre incertitude décroit.

(e) Nous savons que

$$\begin{aligned}1-\alpha = \mathbb{P}(T_{n-1} > -t_{n-1}(1-\alpha)) &= \mathbb{P}\left(\sqrt{n}\,\frac{\bar{X}_n-\mu}{S_n} > -t_{n-1}(1-\alpha)\right) = \\ &= \mathbb{P}\left(\mu < \bar{X}_n + \frac{t_{n-1}(1-\alpha)}{\sqrt{n}}\,S_n\right).\end{aligned}$$

L'intervalle

$$\left(-\infty,\ \bar{X}_n + \frac{t_{n-1}(1-\alpha)}{\sqrt{n}}\, S_n\right)$$

est donc l'intervalle de confiance unilatéral à gauche à $(1-\alpha)\,\%$.

Dans notre cas on a $(-\infty, 14.31)$. On voit bien que la borne supérieure est plus petite que celle de l'intervalle dans la partie (b). Cela est du au fait que toute la possibilité d'erreur passe maintenant du côté droit de l'intervalle.

(f) Nous savons que

$$1-\alpha = \mathbb{P}(T_{n-1} < t_{n-1}(1-\alpha)) = \mathbb{P}\left(\sqrt{n}\,\frac{\bar{X}_n - \mu}{S_n} < t_{n-1}(1-\alpha)\right) = \\ = \mathbb{P}\left(\mu > \bar{X}_n - \frac{t_{n-1}(1-\alpha)}{\sqrt{n}}\, S_n\right).$$

L'intervalle

$$\left(\bar{X}_n - \frac{t_{n-1}(1-\alpha)}{\sqrt{n}}\, S_n,\ \infty\right)$$

est donc l'intervalle unilatéral à droite de confiance à $(1-\alpha)\,\%$.

Dans notre cas on a $(12.31, \infty)$. On voit bien que la borne inférieure est plus grande que celle de l'intervalle dans la partie (b). Cela est du au fait que toute la possibilité d'erreur passe maintenant du côté gauche de l'intervalle.

Exercice 3. (a) On compte le nombre de succès ("succès" = pièce défectueuse) parmi 100 essais. Donc $Y \sim \mathcal{B}(100, p)$, où p est le vrai pourcentage de pièces défectueuses dans un paquet. On observe une réalisation de l'échantillon $Y_1, \ldots, Y_{18} \sim \mathcal{B}(100, p)$.

(b) La fonction de vraisemblance est

$$V(p) = \prod_{i=1}^{18} P(Y_i = y_i) = \prod_{i=1}^{18} \binom{100}{y_i} p^{y_i} (1-p)^{100-y_i} = \left(\prod_{i=1}^{18} \binom{100}{y_i}\right) p^{\sum_{i=1}^{18} y_i} (1-p)^{1\,800 - \sum_{i=1}^{18} y_i}.$$

Pour maximiser cette fonction, on peut maximiser son logarithme,

$$L(p) = \log V(p) = \log\left(\prod_{i=1}^{18} \binom{100}{y_i}\right) + \sum_{i=1}^{18} y_i \log p + \left(1\,800 - \sum_{i=1}^{n} y_i\right) \log(1-p).$$

On dérive, on égalise à zéro,

$$\frac{d}{dp} L(p) = \sum_{i=1}^{18} y_i/p - \left(1\,800 - \sum_{i=1}^{18} y_i\right)/(1-p) = 0,$$

et on trouve l'unique solution

$$\hat{p} = \frac{1}{1\,800} \sum_{i=1}^{18} y_i = \bar{y}/100.$$

On peut vérifier que $\frac{d^2}{dp^2} L(p) < 0$ pour tout $p \in (0,1)$, donc il s'agit d'un maximum.

(c) On a

$$\hat{p} = \bar{y}/100 = 0.00833.$$

Ici on ne peut pas donner d'intervalle de confiance exact, vu que les observations ne sont pas issues d'une loi normale. Cependant, par le théorème 5 du cours, on sait que l'intervalle avec limites $\hat{p} \pm z(1-\alpha/2)J(\hat{p})^{-1/2}$ est un intervalle de confiance approximatif de niveau α pour p. On trouve que pour $\alpha = 0.05$, $z(1-\alpha/2) = 1.96$. De plus

$$J(\hat{p}) = -\frac{d^2}{dp^2}L(\hat{p}) = \frac{18\bar{y}}{\hat{p}^2} + \frac{18\,(100-\bar{y})}{(1-\hat{p})^2},$$

et en insérant les valeurs numériques, $J(0.00833)^{-1/2} = 0.00214$. L'intervalle approximatif à 95 % est donc $(0.00413, 0.01253)$.