Correction partielle de la Feuille de TD8

Tableau de valeurs de z_α telles que $F_{\mathcal{N}(0,1)}(z_\alpha) = 1 - \alpha$

$1-\alpha$	0.8	.9	0.95	0.96	0.97	0.975	0.98	0.99	0.995	0.9995
z_α	0.842	1.282	1.645	1.751	1.881	1.960	2.054	2.326	2.576	3.291

Exercice 8.1 On note $f_{\mathcal{N}(0,1)}$ la densité de la loi $\mathcal{N}(0,1)$, $F_{\mathcal{N}(0,1)}$ sa fonction de répartition. Montrer que les énoncés suivants sont équivalents :

1. $\displaystyle\int_{-z_{\frac{\alpha}{2}}}^{z_{\frac{\alpha}{2}}} f_{\mathcal{N}(0,1)}(x)dx = 1 - \alpha$
2. $F_{\mathcal{N}(0,1)}(z_{\frac{\alpha}{2}}) = 1 - \dfrac{\alpha}{2}$.

Quelle est la valeur de z_0, et de $z_{\frac{1}{2}}$?

Réponse. Soit $f = f_{\mathcal{N}(0,1)}$ la densité de la loi normale et $X \sim \mathcal{N}(0,1)$. L'étape clef de l'équivalence est l'égalité suivante venant de la symétrie de la loi de X qui a même loi que $-X$ (ex 1 du TD 6) :

$$P(|X| > z_{\alpha/2}) = P(X > z_{\alpha/2}) + P(X < -z_{\alpha/2}) = P(X > z_{\alpha/2}) + P(-X < -z_{\alpha/2}) = 2P(X > z_{\alpha/2})$$

On a équivalence entre les énoncés suivants :

$\displaystyle\int_{-z_{\frac{\alpha}{2}}}^{z_{\frac{\alpha}{2}}} f(x)dx = 1 - \alpha$

$\Leftrightarrow$ $P(|X| \leq z_{\alpha/2}) = 1 - \alpha$ (par formule de transfert)

$\Leftrightarrow$ $P(|X| > z_{\alpha/2}) = \alpha$ (en passant au complémentaire)

$\Leftrightarrow$ $2P(X > z_{\alpha/2}) = \alpha$ (par symétrie vue plus haut)

$\Leftrightarrow$ $P(X > z_{\alpha/2}) = \frac{\alpha}{2}$ (calcul)

$\Leftrightarrow$ $P(X \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}$ (en passant au complémentaire)

$\Leftrightarrow$ $F_{\mathcal{N}(0,1)}(z_{\frac{\alpha}{2}}) = 1 - \dfrac{\alpha}{2}$. (définition de la fonction de répartition)

Quelle est la valeur de z_0, et de $z_{\frac{1}{2}}$?

On veut $F_{\mathcal{N}(0,1)}(z_0) = 1 - 0 = 1$ soit $z_0 = +\infty$

On veut $\int_{-z_{\frac{1}{2}}}^{z_{\frac{1}{2}}} f(x)dx = 1 - 1 = 0$ donc vu $f(x) > 0$ il faut $-z_{\frac{1}{2}} = z_{\frac{1}{2}}$ soit $z_{\frac{1}{2}} = 0$.

Exercice 8.2 On considère un échantillon probabiliste de taille n avec loi commune $\mathcal{N}(\mu, \sigma)$. On veut estimer la moyenne μ à l'aide de l'estimateur $\bar{X}$ (moyenne empirique). Soit $d > 0$.

1. Quelle est l'espérance de $\bar{X}$? Quelle est la variance de $\bar{X}$? Quelle est la loi de $\bar{X}$?
2. Écrire $P(|\bar{X} - \mu| \leq d)$ en termes d'une variable normale centrée réduite.
3. Exprimer en fonction de σ et d le plus petit n t.q. $P(|\bar{X} - \mu| \leq d) \geq 1 - \alpha$ pour $\alpha = 5\%$ et $\alpha = 1\%$.

Rappel : Pour l'échantillon $(X_1, ..., X_n)$ **la moyenne empirique est** $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$.

1. Quelle est l'espérance de $\bar{X}$? Quelle est la variance de $\bar{X}$? Quelle est la loi de $\bar{X}$?

   Par linéarité de l'espérance, on a :

$$E(\bar{X}) = \frac{E(X_1) + \cdots + E(X_n)}{n} = \frac{n\mu}{n} = \mu$$

Par indépendance et homogénéité de la Variance, on a

$$Var(\bar{X}) = \frac{Var(X_1) + \cdots + Var(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Par le cours comme la loi mère est normale, $\bar{X}$ est aussi une loi normale, d'espérance μ et de variance $\frac{\sigma^2}{n}$ soit $X \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}.)$

2. Écrire $P(|\bar{X} - \mu| \leq d)$ en termes d'une variable normale centrée réduite.

   La variable centrée réduite associée à $\bar{X}$ est

$$Y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

   soit $\bar{X} - \mu = \sigma Y/\sqrt{n}$ On exprime la proba voulue en fonction de Y :

$$P(|\bar{X} - \mu| \leq d) = P(\sigma|Y|/\sqrt{n} \leq d) = P(|Y| \leq \frac{d\sqrt{n}}{\sigma})$$

3. Exprimer en fonction de σ et d le plus petit n t.q. $P(|\bar{X} - \mu| \leq d) \geq 1 - \alpha$ pour $\alpha = 5\%$ et $\alpha = 1\%$.

   Par l'exercice 1 la condition $P(|\bar{X} - \mu| \leq d) = P(|Y| \leq \frac{d\sqrt{n}}{\sigma}) \geq 1 - \alpha$ est équivalente à $\frac{d\sqrt{n}}{\sigma} \geq z_{\alpha/2}$.

   Pour $\alpha = 5\%$ on a $z_{\alpha/2} = 1.96$ donc il faut

$$\sqrt{n} \geq 1.96\frac{\sigma}{d}.$$

   Pour $\alpha = 1\%$ on a $z_{\alpha/2} = 2.58$ donc il faut

$$n \geq (2.58)^2\frac{\sigma^2}{d^2}.$$

Exercice 8.3 On aimerait estimer la quantité de phosphate (en mg/l) en moyenne dans l'eau d'un lac. Les études des années précédentes laissent supposer que le contenu en phosphate fluctue autour de sa moyenne suivant une loi normale avec un écart type $\sigma = 4$mg/l. Combien de tests d'eau faut-il faire pour être à 90% sûr que l'erreur de l'estimation ne dépasse pas 0.8mg/l ?

(cf TD)

Exercice 8.4 Déterminer un intervalle de confiance bilatéral à 90% pour la moyenne μ d'une population, si l'échantillon a une taille de $n = 63$, une moyenne empirique de 81.3 et une variance théorique de 33.64 (n est suffisament grand pour faire l'approximation par une loi normale). Que devient l'intervalle de confiance si c'est la variance empirique non-biaisée qui vaut 33.64 ?

Réponse. 1/ Cas variance connue
On applique la formule de l'intervalle de confiance pour un échantillon normal de variance connue (cas asymptotiquement presque normal). On prend $F_{\mathcal{N}(0,1)}(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$. Alors, l'intervalle

$$I = [\overline{x} - \frac{\sigma}{\sqrt{n}}z_{\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}}z_{\alpha/2}]$$

avec $\bar{x} = 81.3$, $\sigma = \sqrt{33.64}$, $n = 63$, donc $z_{0.05} = 1,645$.
Donc l'intervalle de confiance est (on prend l'arrondi en dessous pour la borne inférieure et au dessus pour la borne supérieure) :

$$\left[81.3 - \sqrt{\frac{33.64}{63}}1.645, 81.3 + \sqrt{\frac{33.64}{63}}1.645\right] \subset [80.09, 82.51]$$

2/Cas variance inconnue :

On applique la formule de l'intervalle de confiance pour un échantillon normal de variance inconnue (cas asymptotiquement presque normal). On prend $F_{T_{n-1}}(t(n-1)_{\alpha/2}) = 1 - \frac{\alpha}{2}$. Alors, l'intervalle

$$I = [\overline{x} - \frac{s}{\sqrt{n}} t(n-1)_{\alpha/2}, \overline{x} + \frac{s}{\sqrt{n}} z_{\alpha/2}]$$

avec $\bar{x} = 81.3$, $s = \sqrt{33.64}$, $n = 63$, donc $t(62)_{0.05} \simeq 1.671$.

Donc l'intervalle de confiance est (on prend l'arrondi en dessous pour la borne inférieure et au dessus pour la borne supérieure) :

$$\left[81.3 - \sqrt{\frac{33.64}{63}} 1.671, 81.3 + \sqrt{\frac{33.64}{63}} 1.671\right] \subset [80.07, 82.53]$$

Il y a peu de différence (l'approximation de la loi de Student par une loi normale semble aussi raisonnable dans ce cas).

<u>Exercice 8.5</u> Dans l'atmosphere, le poids en mg d'un gaz nocif pour un volume donné est modélisé par une va. $X \sim \mathcal{N}(m, \sigma^2)$. On effectue n prélèvements qui conduisent à un échantillon numérique $x = (x_1, \cdots, x_n)$ de moyenne empirique $\overline{x}$ et d'écart-type s (s^2 est la variance non biaisée). Déterminer, pour la moyenne m un intervalle de confiance I au niveau de confiance $1 - \alpha$ dans les cas suivants :

1. I bilatéral, $1 - \alpha = 0.9$, $n = 16$, $\sigma = 10$, $\overline{x} = 50$
2. I bilatéral, $1 - \alpha = 0.9$, $n = 16$, σ inconnu, $\overline{x} = 50$, $s = 10$.
3. I unilatéral à gauche, $1 - \alpha = 0.95$, $n = 49$, σ inconnu, $\overline{x} = 48$, $s = 9$
4. I unilatéral à droite, $1 - \alpha = 0.975$, $n = 49$, $\sigma = 10$, $\overline{x} = 48$

(cf TD)

<u>Exercice 8.6</u> Pour une étude sur la moyenne du revenu par semaine des serveurs et serveuses dans des restaurants d'une grande ville on a recueilli les données de 75 personnes. La moyenne empirique des données est $227, -$ et l'écart type est supposé connu et valant $15, -$. Déterminer un intervalle de confiance bilatéral pour la moyenne à 90%, et puis celui à 80%.

Réponse. On applique la formule (cas variance connue) avec $\bar{x} = 227, \sigma = 15, n = 75$, donc $z_{0.05} = 1,645$ dans le cas $\alpha = 0.1$ d'où l'intervalle de confiance de niveau 90% :

$$\left[227 - \frac{15}{\sqrt{75}} 1.645, 227 + \frac{15}{\sqrt{75}} 1.645\right] \simeq [224.1, 229.9].$$

Au niveau $\alpha = 80\%$, l'intervalle est plus étroit :

$$\left[227 - \frac{15}{\sqrt{75}} 1.28, 227 + \frac{15}{\sqrt{75}} 1.28\right] \simeq [224.7, 229.2].$$

<u>Exercice 8.7</u> Une compagnie pharmaceutique veut savoir si le procédé de fabrication qu'elle utilise fournit effectivement des comprimés dosés à 5mg de principe actif d'un medicament. L'écart type empirique est estimé à 0.07mg.

La quantité de principe actif est mesurée pour 100 comprimés issus d'un lot de fabrication. Le dosage moyen de principe actif est de 4.85 mg.

Peut-on dire avec une confiance de 95% que le processus donne le dosage prévu ?

Réponse. Méthode :On cherche un intervalle de confiance de niveau 95% et on regarde si la moyenne théorique s'y trouve.

On applique la formule (cas variance inconnue) avec $\bar{x} = 4.85$, $s = 0.07$, $n = 100$, donc $t_{0.025}(99) \simeq t_{0.025}(100) = 1,984$ (on prend la valeur disponible la plus proche dans la table) dans le cas $\alpha = 0.05$ d'où l'intervalle de confiance de niveau 95% :

$$\left[4.85 - \frac{0.07}{\sqrt{100}}1.984, 4.85 + \frac{0.07}{\sqrt{100}}1.984\right] \simeq [4.836, 4.864].$$

On remarque que $5 \notin [4.836, 4.864]$, donc on ne peut pas dire avec une confiance de 95% que le dosage correspond au dosage prévu de 5mg. (On dira plus tard qu'on rejette l'hypothèse nulle du test que la moyenne $\mu = 5$).