Fiche TD Statistique

1 Estimation ponctuelle

Exercice 1.

Soient $X_1, \ldots, X_n$ i.i.d. de loi de Poisson de moyenne λ. Déterminez le MLE pour le paramètre λ.

Solution :

5. Pour $X \sim \text{Poisson}(\lambda)$, et $x \in \mathbb{N}$

$$\mathbb{P}(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Donc pour $X_1, \ldots, X_n$ i.i.d $\sim$ Poisson (λ), (il s'git d'une loi discrète ici) pour un échantillor $(x_1, \cdots, x_n)$ donné, la vraisemblance s'écrit

$$\begin{aligned} L(\lambda; x_1, \cdots, x_n) &= \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) \\ &= \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \\ &= \frac{e^{-n\lambda}}{x_1! \cdots x_n!} \lambda^{x_1 + \cdots + x_n} \end{aligned}$$

La log-vraisemblance s'écrit

$$\begin{aligned} \ell(\lambda; x_1, \cdots, x_n) &= \log L(\lambda; x_1, \cdots, x_n) \\ &= -n\lambda - \log(x_1! \cdots x_n!) + (x_1 + \cdots + x_n) \log(\lambda) \end{aligned}$$

On cherche a maximiser $\ell(\lambda; x_1, \ldots, x_n)$ pour $\lambda > \mathbb{C}$. On remarque que $\ell \to -\infty$ quand $\lambda \longrightarrow 0$, et $\ell \to -\infty$ quand $\lambda \longrightarrow +\infty$. En plus,

$$\frac{\partial \ell}{\partial \lambda} = 0 \to -n + (x_1 + \cdots + x_n) \frac{1}{\lambda} = 0$$

On obtient l'estimation de maximum de vraisemblance

$$\hat{\lambda}^{\mathrm{MV}} = \frac{1}{n}(x_1 + \cdots + x_{\mathrm{n}})$$

$$\hat{\lambda}^{\mathrm{MV}} = \frac{1}{\mathrm{n}}(X_1 + \cdots + X_{\mathrm{n}})$$

et nous retrouvons donc

$$\hat{\lambda}^{\mathrm{MV}} = \bar{X}_{\mathrm{n}}.$$

Exercice 2.

Soient $X_1, \ldots, X_n$ i.i.d. $\text{Bin}(1, p)$. Déterminez le MLE pour le paramètre p.

Solution :

L'ensemble des valeurs possibles est $\{0,1\}$. Le paramètre inconnu est p. Si $(x_1,\ldots,x_n) \in \{0,1\}^n$ est un échantillon, la vraisemblance vaut :

$$L(x_1,\ldots,x_n,p) = p^{\sum x_i}(1-p)^{n-\sum x_i} .$$

Son logarithme est :

$$\log(L(x_1,\ldots,x_n,p)) = (\sum x_i)\log p + (n - \sum x_i)\log(1-p) .$$

La dérivée par rapport à p est :

$$\frac{\partial \log(L(x_1,\ldots,x_n,p))}{\partial p} = (\sum x_i)\frac{1}{p} - (n-\sum x_i)\frac{1}{1-p} .$$

Elle s'annule pour :

$$\widehat{p} = \frac{\sum x_i}{n} .$$

La dérivée seconde est :

$$\frac{\partial^2 \log(L(x_1,\ldots,x_n,p))}{\partial p^2} = -(\sum x_i)\frac{1}{p^2} - (n-\sum x_i)\frac{1}{(1-p)^2} .$$

Elle est strictement négative, la valeur $\widehat{p}$ est bien un maximum. Si $(X_1,\ldots,X_n)$ est un échantillon de la loi de Binomiale (Bernoulli) de paramètre p, l'estimateur du maximum de vraisemblance de p est :

$$\frac{\sum X_i}{n} ,$$

à savoir la fréquence empirique.

Exercice 3.

Soit $(X_1,\ldots,X_n)$ un échantillon aléatoire simple issu d'une population de densité

$$f_\theta(x) = \begin{cases} \frac{1}{\theta}e^{-\frac{1}{\theta}(x-\gamma)} \text{ si } x > \gamma \\ 0 \text{ sinon} \end{cases}$$

où $\theta > 0$. Déterminez une estimation des paramètres θ et γ par la méthode du MLE.

Solution :

$$\begin{aligned} L_{\theta,\gamma}(\mathbf{X}) &= \prod_{i=1}^{n} \frac{1}{\theta}\exp\left(-\theta^{-1}(X_i-\gamma)\right)\mathbb{I}_{[\gamma,+\infty[}(X_i) \\ &= \frac{1}{\theta^n}\exp\left(-\theta^{-1}\sum(X_i-\gamma)\right)\mathbb{I}_{[\gamma,+\infty}\left[X_{(1)}\right) . \end{aligned}$$

En se limitant à $X_{(1)} > \gamma$,

$$\begin{aligned} \log L_{\theta,\gamma}(\underline{\mathbf{X}}) &= -n\log\theta - \frac{1}{\theta}\sum_i (X_i-\gamma) \\ \partial_\theta \log L_{\theta,\gamma}(\underline{\mathbf{X}}) &= -n\frac{1}{\theta} + \frac{1}{\theta^2}\sum_i (X_i-\gamma) \\ \partial_\theta \log L_{\theta,\gamma}(\underline{\mathbf{X}}) = 0 &\Leftrightarrow -n\theta + \sum_i X_i - n\gamma = 0 \\ &\Leftrightarrow \theta = \bar{X} - \gamma \\ \partial_\gamma \log L_{\theta,\gamma}(\underline{\mathbf{X}}) &= \frac{n}{\theta}. \end{aligned}$$

Cette dernière quantité n'est jamais nulle. Souhaitant maximiser la vraisemblance, on remarque qu'à θ fixé, la vraisemblance est une fonction croissante de γ. Quand γ prend sa valeur maximale, la vraisemblance sera maximale. Or, $\gamma \leq X_{(1)}$.

On trouve alors

$$\hat{\gamma} = X_{(1)} = X_{\min} \text{ et } \hat{\theta} = \bar{X} - X_{(1)}$$

Exercice 4.

Soit $X_1, \ldots, X_n$ un échantillon aléatoire simple issu d'une population de densité

$$f_\theta(x) = \begin{cases} \frac{\theta}{1-\theta} x^{\frac{2\theta-1}{1-\theta}} & \text{si } 0 < x < 1 \\ 0 & \text{sinon} \end{cases}$$

où $1/2 < \theta < 1$. Déterminez le MLE pour θ.

Solution :

Immédiatement, en supposant $X_1 \ldots, X_n$ dans le support de f_θ

$$\begin{aligned} L_\theta(\underline{\mathbf{X}}) &= \prod_{i=1}^n f_\theta(X_i) \\ &= \prod_{i=1}^n \frac{\theta}{1-\theta} X_i^{\frac{2\theta-1}{1-\theta}} \mathbb{I}_{0<X<1}(X_i) \\ &= \left(\frac{\theta}{1-\theta}\right)^n \left(\prod_{i=1}^n X_i\right)^{\frac{2\theta-1}{1-\theta}} \prod_{i=1}^n \mathbb{I}_{0<X<1}(X_i) \\ \partial_\theta \log L_\theta(\underline{\mathbf{X}}) = 0 \Leftrightarrow \theta &= \frac{1}{1 - \frac{1}{n}\sum_i \log(X_i)} \end{aligned}$$

Exercice 5.

Les éléments d'une population possèdent un caractère X qui suit une loi de densité

$$f_\theta(x) = \frac{\theta+1}{2}(1-|x|)^\theta, x \in (-1,1)$$

où on suppose le paramètre $\theta > -1$. On en extrait un échantillon simple $X_1, \ldots, X_n$. Déterminez l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ.

Solution :

Comme d'habitude, en supposant que les indicatrices sont vérifiées (le support ne dépendant pas de θ),

$$\begin{aligned} L_\theta(\underline{\mathbf{X}}) &= \prod_{i=1}^n f_\theta(X_i) \\ &= \left(\frac{\theta+1}{2}\right)^n \left(\prod_{i=1}^n (1-|X_i|)\right)^\theta \mathbb{I}_{\{X \geq -1\}}\left(X_{(1)}\right) \mathbb{I}_{\{X \leq 1\}}\left(X_{(n)}\right) \\ \log L_\theta(\underline{\mathbf{X}}) &= n \log\left(\frac{\theta+1}{2}\right) + \theta \sum_{i=1}^n \log(1-|X_i|) \\ \partial_\theta \log L_\theta(\underline{\mathbf{X}}) &= \frac{n}{\theta+1} + \sum_{i=1}^n \log(1-|X_i|) \\ \partial_\theta \log L_\theta(\underline{\mathbf{X}}) = 0 \Leftrightarrow \theta &= -\frac{n}{\sum \ln(1-|X_i|)} - 1 \end{aligned}$$

Exercice 6.

Les éléments d'une population possèdent un caractère X qui suit une loi de densité

$$f_\theta(x) = \frac{\sqrt{\theta}}{\sqrt{2\pi}} e^{-\theta x^2/2}$$

où $\theta > 0$. Pour étudier le paramètre θ, on a effectué une suite de n expériences indépendantes qui ont donné les réalisations $x_1, \dots, x_n$ de n v.a. $X_1, \dots, X_n$ i.i.d. de même loi que X.

1. Déterminez un estimateur $\hat{\theta}$ du paramètre θ par la méthode du maximum de vraisemblance.
2. $\hat{\theta}$ est-il exhaustif ?
3. Calculez la moyenne et la variance de $\hat{\theta}$. Déduisez-en un estimateur $\hat{\theta}_1$ de θ non biaisé. Quelle est la variance de $\hat{\theta}_1$? Est-il convergent ?

Exercice 7.

Les éléments d'une population possèdent un caractère X qui suit une loi de densité

$$f_\theta(x) = \frac{2}{\sqrt{\pi}\theta^{3/2}} x^2 e^{-x^2/\theta}$$

où $\theta > 0$. Une suite de n expériences indépendantes a donné les valeurs $x_1, \dots, x_n$.

1. Déterminez un estimateur $\hat{\theta}$ du paramètre θ par la méthode du maximum de vraisemblance.
2. Examinez les qualités suivantes de $\hat{\theta}$: efficacité, biais, convergence, exhaustivité.

Exercice 8.

Les éléments d'une population possèdent un caractère X qui suit une loi de probabilité dont la densité est

$$f_{a,\theta}(x) = \begin{cases} \theta e^{-\theta(x-a)} & \text{si } x \geq a \\ 0 & \text{sinon} \end{cases}$$

où $\theta, a > 0$. Une suite de n expériences indépendantes a donné les valeurs $x_1, \dots, x_n$.

Inférence sur a, en supposons que θ est connu. Proposez un estimateur $\hat{a}$ de a par la méthode du maximum de vraisemblance.

2 Tests statistique

Exercice 9.

On veut savoir si la résistance moyenne de composants produits dans une usine est 400Ω. On considère que la distribution des résistances est normale, et on mesure pour 16 composants les valeurs $392, 396, 386, 389, 388, 387, 403, 397, 401, 391, 400, 402, 394, 406, 406, 400$.

(a) Donner les estimations ponctuelles des moyenne et variance.

(b) Peut-on considérer, au seuil de signification $\alpha = 5\%$, que le lot respecte la norme de 400 ? Même question avec un seuil de $\alpha = 1\%$.

Solution :

(a) On trouve $\bar{x} = 396.125, s = 6.742$ et $s^2 = 45.45$.

b) Si l'on fait l'hypothèse H_0 : "le lot respecte la norme de $400\Omega''$, alors dans 95% des cas la moyenne sur un échantillon d'effectif 16 se trouve dans l'intervalle

$$[400 - t * 6.742/4, 400 + t * 6.742/4],$$

t étant lu dans la table de la loi de Student à 15 degrés de liberté : $t = 2.1314$.

Ainsi l'intervalle de confiance 95% pour la résistance est $[396.40, 403.59]$, et on peut donc, au risque 5%, rejeter l'hypothèse.

Au seuil $\alpha = 1\%$, on a dans l'hypothèse H_0 un intervalle de confiance pour la moyenne $[400 - t * 6.742/4, 400 + t * 6.742/4]$, avec $t = 2.9467$. Ainsi, l'intervalle est $[395.03, 404.97]$. Au risque 1%, on ne rejette pas H_0.

Exercice 10.

Un fabricant se vante de proposer des tubes à essai d'une durée de vie supérieure à $2000h$ de chauffage. A l'aide d'un échantillon de 100 tubes testés, on estime la durée de vie moyenne à 1975h, avec un écart-type de $130h$. Peut-on affirmer, au risque 5%, que le fabriquant ment ?

Solution :

Il s'agit ici d'un test unilatéral... H_0 est l'hypothèse : "la durée de vie moyenne vérifie $\mu \geq 2000''$. On peut supposer, l'effectif de l'échantillon étudié étant grand, que $\sqrt{n}\frac{\bar{X}-\mu}{s}$ suit une loi normale centrée réduite. Si H_0 est vérifiée, on cherche t tel que

$$\mathbb{P}(\mu - ts/\sqrt{n} \leq \bar{X}) = 0.95,$$

soit

$$\mathbb{P}\left(-t \leq \sqrt{n}\frac{\bar{X}-\mu}{s}\right) = 0.95,\ \text{et donc } 1 - F(-t) = F(t) = 0.95 : t = 1.64$$

Ainsi, dans I' hypothèse H_0, la durée de vie moyenne d'un échantillon d'effectif 100se trouve, dans 95% des cas, dans l'intervalle $[2000 - 1.64 * 130/10, +\infty[= [1978.68, +\infty[$. La mesure de 1975 h sur l'échantillon n' étant pas dans cet intervalle,

H_0 doit être rejetée : il est probable que le fabriquant mente.

Exercice 11.

Un fabricant annonce que la masse d'un composant de l'un de ses produits est de $75mg$. Les mesures pour le vérifier étant coûteuses, trois seulement sont réalisées, dont les résultats sont $70, 72$ et $74mg$. Peut-on, au risque de 5% de se tromper, dénoncer la publicité du fabriquant ?

Solution :

Notons X la variable aléatoire correspondante. (on doit supposer que la loi de X est une loi normale pour pouvoir appliquer les méthodes du cours). On note $\mu = E(X)$: il s'agit donc icl d'effectuer un test bilatéral de l'hypothèse $H_0 : \mu = 75$. On obtient sur un échantillon de 3 mesures : $n = 3, \bar{x} = 72, \sigma'^2 = 8/3$ et $s^2 = 8/2 = 4$ donc l'estimation ponctuelle de l'écart-type est $s = 2$. On sait que $\sqrt{n}\frac{\bar{X}-\mu}{s}$ suit une loi de Student à 2 degrés de liberté, donc si $\alpha = 0.05$,

$$t(\alpha) = 4.3027$$

Ainsi, la moyenne des durées de vie mesurées sur un échantillon d'effectif 3 sera, dans 95% des cas, dans l'intervalle $[75 - 4.3027 \times s/\sqrt{3}, 75 + 4.3027 \times s/\sqrt{3}] = [70.03, 79.97]$. La valeur moyenne 72 mesurée sur l'échantillon étant bien dans cet intervalle, on n' a pas de raisons, au vu de ces mesures, de rejeter H_0.

Exercice 12.

Un laboratoire pharmaceutique désire étudier les effets secondaires potentiels d'un médicament sur le taux de cholestérol des patients. Cent volontaires sains sont donc choisis pour tester le médicament.

(a) Avant l'expérience, le taux de cholestérol moyen de ces volontaires est de $2.02 \pm 0.2g/l$. Le taux de cholestérol moyen dans la population étant de 2 g/l, vérifier que cet échantillon est représentatif au risque 5

(b) Après un mois de traitement, seuls 97 volontaires reviennent faire un test. Leur taux moyen de cholestérol est passé à $2.09g/l$ avec un écart-type d'échantillon de $0.25g/l$.

La différence est-elle significative au risque 5% ? Au risque 1% ?

Solution :

(a) Soit X_1 la variable aléatoire qui mesure le taux de cholestérol d'un individu ; $E(X_1) = \mu_1 = 2$ $\overline{X_1}$ est le taux moyen mesuré sur un échantillon de taille $n_1 = 100$ Alors n_1 étant plus grand que 30, on peut considérer que $\sqrt{n_1}\frac{\overline{X_1}-2}{s_1}$ suit une loi normale, avec $s_1 = 0.2$ estimation ponctuelle de l'écart-type de X_1. Ainsi, dans 95% des cas le taux moyen observé sur un échantillon sera compris dans $[2 - 1.96 \times 0.2/10, 2 + 1.96 \times 0.2/10] = [1.961, 2.039]$.

Le taux de cholestérol moyen des volontaires étant bien dans cet intervalle, on peut considérer que cet échantillon est représentatif.

b) Soit X_2 la variable aléatoire mesurant le taux de cholestérol d'un individu après un mois de traitement ; son espérance μ_2 est inconnue. $\overline{X_2}$ est le taux moyen d'un échantillon de taille $n_2 = 97$.

On fait l'hypothèse H_0 : "est les taux de cholestérol moyens sont les mêmes avant et après traitement".

Alors $\mu_1 = \mu_2$, et on peut considérer que

$$\frac{\overline{X_1} - \overline{X_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \sim N(0,1)$$

(avec $s_1 = 0.2, s_2 = 0.25$), et par conséquent on détermine l'intervalle de confiance au risque 5% de

$$\overline{X_1} - \overline{X_2} : \left[-1.96\sqrt{s_1^2/n_1 + s_2^2/n_2}, 1.96\sqrt{s_1^2/n_1 + s_2^2/n_2}\right] = [-0.063, 0.063]$$

Comme la différence entre les taux moyens mesurés $2.02-2.09 = 0.07$ n'est pas dans cet intervalle, elle est significative, et on rejette H_0 donc on considère, au risque 5% de se tromper, que le médicament a un effet.

En revanche, l'intervalle de confiance au risque 1% est

$$\left[-2.57\sqrt{s_1^2/n_1 + s_2^2/n_2}, 2.57\sqrt{s_1^2/n_1 + s_2^2/n_2}\right] = [-0.083, 0.083],$$

intervalle qui contient la valeur $2.02 - 2.09 = 0.07$, donc la différence n'est pas significative au risque de 1%.

Exercice 13.

Pour étudier un nouvel alliage métallique, on a soumis un échantillon aléatoire de 16 tiges aux essais pour obtenir les résistances suivantes en kg/cm^2 :
$1895, 1920, 1886, 1890, 1864, 1880, 1875, 1915, 1850, 1927, 1910, 1912, 1886, 1903, 1854, 1880.$

On suppose la résistance distribuée normalement.

(a) Estimer par intervalle avec un niveau de confiance de 95%, la résistance moyenne à la rupture.

(b) Avant l'introduction de ce nouvel alliage la résistance moyenne à la rupture des tiges était de $1840 kg/cm^2$. Que peut-on conclure des essais effectués avec le nouvel alliage ?

Exercice 14.

Les habitants d'une région aéroportuaire se plaignent que le bruit des avions dépasse la limite autorisée de 80 décibels en moyenne imposée par la législation. On admet que l'intensité du bruit causé par les avions est une variable aléatoire X de loi gaussienne d'espérance μ et de variance 64 .

On mesure un échantillon journalier de $\mathrm{n} = 16$ variables aléatoires indépendantes $\mathrm{X}_1, \ldots, \mathrm{X}_\mathrm{n}$ de l'intensité du bruit, et on effectue le test statistique suivant.

$$\begin{cases} \mathrm{H}_0 : \mu = \mu_0 = 80 \text{ décibels} \\ \mathrm{H}_1 : \mu = \mu_1 = 85 \text{ décibels} \end{cases}$$

1. Expliciter les risques de première et deuxième espèces. De quel point de vue est fait ce test ? Celui des habitant ou celui des responsables de I'aéroport ?
2. Quelle variable de décision faut-il choisir et quelle est sa loi ?
3. Calculer le seuil de la région critique pour un risque $\alpha = 5\%$.
4. Calculer la puissance du test.
5. Enoncer les règles de décision avec les probabilités d'erreur.
6. La moyenne calculée sur l'échantillon est $\bar{x} = 83$ décibels. Les habitants ont-ils raison de se plaindre ? Le test d'hypothèses ainsi établi leur est-il favorable ou défavorable ?
7. Combien faudrait-il faire de relevés journaliers, pour que le risque de deuxième espèce soit de 5% ?
8. Quelle serait alors le seuil de décision ?

Exercice 15.

Sur un échantillon de 900 naissances, on constate qu'il y a 470 garçons. Un généticien décide d'utiliser ces données pour effectuer le test suivant relatif aux proportions p et 1 -p de naissances respectivement masculines et féminines :

$$\mathrm{H}_0 : \mathrm{p} = 0.5$$
$$\mathrm{H}_1 : \mathrm{p} = 0.55$$

1) Construire un test pour ces hypothèses avec un risque $\alpha = 5\%$. Peut-on être satisfait du test ? Si non comment peut-on I'améliorer ?

2) Ce généticien effectue une nouvelle étude sur un échantillon de même taille. II souhaite cette fois tester les hypothèses :

$$\mathrm{H}_0 : \mathrm{p} = 0.5$$
$$\mathrm{H}_1 : \mathrm{p} \neq 0.5$$

Solution :

1)

HO : p = 0.5
H1 : p = 0.55

Variable de décision La fréquence empirique $\mathrm{F_n}$ est un estimateur de p et avec un échantillon de taille 900, on peut considérer grâce au TCL que F_n suit une loi normale $N(p, p(1-p)/n)$ où p = 0.5 sous I'hypothèse $\mathrm{H_0}$ et p = 0.55 sous I'hypothèse $\mathrm{H_1}$.

La région critique W est la région d'acceptation de H_1 d'où $W=\{F_n \geq C\}$. (c) Calcul du seuil On sait que $\alpha=P(W \mid H_0 \text{ vraie })=P(F_n \geq C \mid H_0 \text{ vraie })$. Supposons $\mathrm{H_0}$ vraie alors $\mathrm{F_n}$ suit une $\mathrm{N}\left(0.5, {\sigma_0}^2\right)$ où ${\sigma_0}^2=0.5 * 0.5/\mathrm{n}$, d'où $\alpha=P(F_n \geq C)=P\left(\frac{F_n-0.5}{\sqrt{0.5 * 0.5}} \sqrt{n} \geq \frac{C-0.5}{0.5} \sqrt{n}\right) \Leftrightarrow P(Z \geq C')=0.05$ où Z suit une loi $N(0,1)$

$$\Rightarrow C'=1.64 \Rightarrow C=0.5+1.64 * \frac{0.5}{\sqrt{n}}$$
$$\Rightarrow C \approx 0.5+1.64^* 0.5/30 \approx 0.53$$

(d) Calcul de la puissance On sait que $1-\beta=P(W \mid H_1 \text{ vraie })=P(F_n \geq C \mid H_0 \text{ vraie })$. Supposons $\mathrm{H_1}$ vraie alors $\mathrm{F_n}$ suit une $\mathrm{N}\left(0.55, \sigma_1^2\right)$ où $\sigma^2=0.45^* 0.55/\mathrm{n}, \mathrm{d}'$ où

$$1-\beta=P(F_n \geq C)=P\left(\frac{F_n-0.55}{\sqrt{0.55 * 0.45}} \sqrt{n} \geq \frac{C-0.55}{\sqrt{0.55 * 0.45}} \sqrt{n}\right)=P(Z \geq-1.21)=P(Z<1.21)=0.89$$

$\Rightarrow B=0.11$

(d) Règles de décision Si $f_n \geq 0.53$ alors on accepte H_1, i.e on considère qu'il y a plus de garçons que de filles avec 5% de risque de se tromper. Si $f_n<0.53$ alors on garde H_0, i.e on considère qu'il y a autant de filles que de garçons avec 11% de risque se tromper. L'échantillon considéré indique $f_n=470/900=0.52$ donc le généticien conclut qu'il y autant de garçons que de filles avec 11% de risque de se tromper. (e) Taille échantillon Le risque de $2^{\text{ème}}$ espèce n'est pas acceptable. Pour le réduire nous allons jouer sur la taille de I'échantillon. L'erreur de première espèce donne une première équation

$$C=0.5+1.64^* \frac{0.5}{\sqrt{n}}$$

Et si on impose une erreur de seconde espèce de 0.05 alors on obtient une deuxième équation

$$1-\beta=P(F_n \geq C)=P\left(\frac{F_n-0.55}{\sqrt{0.55^* 0.45}} \sqrt{n} \geq \frac{C-0.55}{\sqrt{0.55^* 0.45}} \sqrt{n}\right)$$
$$\Leftrightarrow 0.95=P\left(Z \geq C^2\right) \Leftrightarrow 0.95=P\left(Z \leq-C^t\right) \Rightarrow-C'=1.64$$
$$C=0.55-1.64^* \frac{\sqrt{0.55^* 0.45}}{\sqrt{n}}$$

d'ou

$$C=0.55-1.64 * \frac{\sqrt{0.55^* 0.45}}{\sqrt{n}}=0.5+1.64 * \frac{0.5}{\sqrt{n}}$$
$$\Leftrightarrow 0.55-0.5=\frac{1.64}{\sqrt{n}}(0.5+\sqrt{0.55 * 0.45})$$
$$\Leftrightarrow \sqrt{n}=32.72 \Rightarrow n=1070$$

Si on souhaite diminuer I'erreur de $2^{\text{ème}}$ espèce, il faut tester un échantillon de taille au moins 1070 . 2)

HO : p = 0.5
H1 : p ≠ 0.5

Le test est maintenant bilatéral. Même estimateur, même lois. (a) Région critique La région critique W est la région d'acceptation de H_1 d'où $W=\{\mathrm{F}<\mathrm{C}_1 \text{ ou} \mathrm{F}>\mathrm{C}_2\}$

et West la région d'acceptation de H_0 d'où $\overline{\mathrm{W}}=\{\mathrm{C}_1 \leq \mathrm{F} \leq \mathrm{C}_2\}$ (cf. dessin). (b) Calcul des seuils On sait que $\alpha=P(W \mid H_0$ vraie $)=P(F<C_1$ ou $F>C_2 \mid H_0$ vraie). Afin de simplifier le calcul de probabilité, on passe à $1-\alpha=P(\bar{W} \mid H_0$ vraie $)=P(C_1 \leq F \leq C_2 \mid H_0$ vraie). Supposons H_0 vraie alors F suit une $\mathrm{N}\left(0.5, \sigma_0{}^2\right)$ où $\sigma_0{}^2=0.5^* 0.5 / \mathrm{n}, \mathrm{d}'$ où

$$1-\alpha=\mathrm{P}\left(\mathrm{C}_1 \leq \mathrm{F} \leq \mathrm{C}_2\right)=\mathrm{P}\left(\frac{\mathrm{C}_1-0.5}{\sqrt{0.5^* 0.5}} \sqrt{\mathrm{n}} \leq \frac{\mathrm{F}-0.5}{\sqrt{0.5^* 0.5}} \sqrt{\mathrm{n}} \leq \frac{\mathrm{C}_2-0.5}{\sqrt{0.5^* 0.5}} \sqrt{\mathrm{n}}\right)$$

$\Leftrightarrow P\left(C_1 \leq Z \leq C_2'\right)=0.05$ On suppose que le risque est symétrique, or la loi normale centrée est aussì symétrique par rapport à 0, on a donc $\mathrm{C}_1'=-\mathrm{C}_2'$. D'où $\mathrm{P}\left(\mathrm{C}_1' \leq \mathrm{Z} \leq \mathrm{C}_2'\right)=0.05 \Leftrightarrow$

$\mathrm{P}\left(-\mathrm{C}_2' \leq \mathrm{Z} \leq \mathrm{C}_2'\right)=0.05 \Leftrightarrow \mathrm{F}\left(\mathrm{C}'_2\right)-\mathrm{F}\left(-\mathrm{C}'_2\right)=0.05$ où F est la fonction de répartition de z $\Leftrightarrow \quad F\left(C'2\right)-\left[1-F\left(C'2\right)\right]=0.05 \quad \Leftrightarrow \quad F\left(C'2\right)=1.95 / 2=0.975 \Rightarrow C'_2=1.96 \Rightarrow C_2=0.5+1.96^* \sqrt{0.5^* 0.5} / \sqrt{n}=0.53$ et $C_1=0.5-1.96^* \sqrt{0.5^* 0.5} / \sqrt{n}=0.47$

(c) Règles de décision

- Si $f_n<0.47$ ou $f_n>0.53$, on accepte H_1, i.e on considère que la côte du président a changé avec 5% de chance de se tromper.
- Dans le cas contraire, on accepte H_0, i.e on considère que sa côte est stable, mais on ne connait pas le risque encourus car on ne connait pas la loi sous H_1.

Exercice 16.

On s'interroge sur la comparaison des tailles moyennes des garçons et des filles de 6 ans dans une population, pour cela on a pris comme échantillon, jugé représentatif de cette tranche d'âge, et on a observé : 16 garçons : moyenne 126.5 cm, écart-type estimé 12.9 cm 15 filles : moyenne 136.9 cm, écart-type estimé 11.9 cm. On admet que la distribution des tailles dans chacune des sous-populations (garcons, filles) suit une loi gaussienne.

1. Donner des intervalles de confiance à 95% pour les tailles moyennes des garcons et des filles.
2. Donner un intervalle de confiance à 95% pour l'écart-type de la taille des garçons. Meme question pour les filles.
3. Donner un intervalle de confiance à 95% pour le rapport des deux écart-types. Peut-on conclure que les variances des deux populations sont différentes ? (On donne, pour la loi de Fisher-Snedecor à (15,14) degrés de libertés, les quantiles $f_{0.975}=2.95$ et $\left.f_{0.025}=\frac{1}{f_{0.975}}=0.339\right)$

Solution :

Corrigé : Soit $X=\left(X_1, \ldots, X_m\right)$ un m -échantillon de loi normale $\mathcal{N}\left(\mu, \sigma^2\right)$ représentant la taille des garçons mesurés et $X'=\left(X_1', \ldots, X_n'\right) m$ -échantillon de loi normale $\mathcal{N}\left(\mu', \sigma'^2\right)$ représentant la tailles des filles mesurées avec X et X' indépendants. (Ici, $m=16$ et $n=15$). Soit $\bar{X}_m=\frac{\sum_{i=1}^m X_i}{m}$ et $\bar{X}_n'=\frac{\sum_{i=1}^n X_i'}{n}$, les moyennes empiriques associées. Soit $S_m^2=\frac{\sum_{i=1}^m\left(X_i-\bar{X}_m\right)^2}{m-1}$ et $S_n'^2=\frac{\sum_{i=1}^n\left(X_i'-\bar{X}_n'\right)^2}{n-1}$, les variances estimées associés.

1. Comme σ est inconnu, pour déterminer un intervalle de confiance pour μ, on va utiliser une statistique de Student. (Si σ était connu, on utiliserait directement la loi normale).

   Propriété du cours sur les lois normales : Les statistiques $\frac{\sqrt{m}\left(\bar{X}_m-\mu\right)}{\sigma}$ et $\frac{(m-1) S_m^2}{\sigma^2}$ sont indépendantes et de lois respectives $\mathcal{N}(0,1)$ et $\chi^2(m-1)$ On en déduit que l'estimateur :

$$T_m = \frac{\frac{\sqrt{m}\left(\bar{X}_m - \mu\right)}{\sigma}}{\sqrt{\frac{(m-1)S_m^2}{\sigma^2}/(m-1)}} = \frac{\sqrt{m}\left(\bar{X}_m - \mu\right)}{S_m} \sim \mathcal{T}(m-1)$$

suit une loi de student de paramètre $(m-1)$. Remaryue : Lorsque σ est connue, on utilise $\frac{\sqrt{m}(X_m-\mu)}{\sigma} \sim \mathcal{N}(0,1)$. Si σ est inconnu, "ion par son écart-type estúné S_m pour obtenir la statistique $T_m \sim \mathcal{T}(m-1)$.

On a donc $\mathbf{P}\left(|T_m| \leq t_{1-\alpha/2}^{m-1}\right) = 1-\alpha$. D'où l'intervalle de confiance de niveau $1-\alpha$ pour μ :

$$\bar{X}_m - \frac{t_{1-\alpha/2}^{m-1} S_m}{\sqrt{m}} \leq \mu \leq \bar{X}_m + \frac{t_{1-\alpha/2}^{m-1} S_m}{\sqrt{m}}$$

De même, on obtient l'intervalle de confiance pour μ' :

$$\bar{X}'_n - \frac{t_{1-\alpha/2}^{n-1} S'_n}{\sqrt{n}} \leq \mu' \leq \bar{X}'_n + \frac{t_{1-\alpha/2}^{n-1} S'_n}{\sqrt{n}}$$

Application Numérique : On lit sur la table de statistique pour $\alpha = 0.05$: $t_{0.975}^{15} = 2.1314$ et $\frac{\text{Application Numérique : On lit sur la table de statistique pour}}{t_{0.975}\text{=2.1448 d'où les intervalles de confiance de niveau 95\%}}$

$$\mu \in [119.63, 133.37], \mu' \in [130.31, 143.49].$$

2. Pour déterminer un intervalle de confiance pour σ, on utilise l'estimateur S_m^2 et une loi du χ^2. On sait que $\frac{(m-1)S_m^2}{\sigma^2} \sim \chi^2(m-1)$ d'ou

$$\mathbf{P}\left(k_{\alpha/2}^{m-1} \leq \frac{(m-1)S_m^2}{\sigma^2} \leq k_{1-\alpha/2}^{m-1}\right) = 1-\alpha$$

En isolant σ. on obtient lintervalle de confiance de niveay $1-\alpha$

$$S_m\sqrt{\frac{m-1}{k_{1-\alpha/2}^{m-1}}} \leq \sigma \leq S_m\sqrt{\frac{m-1}{k_{\alpha/2}^{m-1}}}$$

et de même, on obtint l'intervalle de confiance pour σ' :

$$S'_n\sqrt{\frac{n-1}{k_{1-\alpha/2}^{n-1}}} \leq \sigma' \leq S'_n\sqrt{\frac{n-1}{k_{\alpha/2}^{n-1}}}$$

$\frac{\text{Application Numérique : On lit sur la table de statistique } k_{0.025}^{15}\text{=6.2621.}k_{0.975}^{15}\text{=27.4884}}{k_{0.025}^{14}\text{=5.6287 et }k_{0.975}^{14}\text{=26.1186 d'ou les intervalles de confiance de niveau 95\%:}}$

$$\sigma \in [9.53, 19.97], \sigma' \in [8.712, 18.77]$$

3. Pour déterminer un intervalle de confiance de σ'/σ, on utilise l'estimateur S'_n/S_m et une loi de Fisher-Snedecor. On sait que $\frac{(m-1)S_m^2}{\sigma^2}$ et $\frac{(n-1)S_n'^2}{\sigma'^2}$ sont indépendantes et de lois respectives $\chi^2(m-1)$ et $\chi^2(n-1)$. On en déduit que la statistique :

$$F_{m,n} = \frac{\frac{(m-1)S_m^2}{\sigma^2}/(m-1)}{\frac{(n-1)S_n'^2}{\sigma'^2}/(n-1)} = \frac{\sigma^2 S_m^2}{\sigma^2 S_n'^2} \sim \mathcal{F}(m-1, n-1),$$

suit une loi de Fisher-Snederor a $(m-1, n-1)$ degrís de libertés. Donc

$$\mathbb{P}\left(F_{m,n} \in \left[f_{\alpha/2}^{m-1,n-1}, f_{1-\alpha/2}^{m-1,n-1}\right]\right) = 1-\alpha.$$

D'où l'intervalle de confiance pour σ'/σ de niveau $1-\alpha$:

$$\frac{S'_n}{S_m}\sqrt{f_{\alpha/2}^{m-1,n-1}} \leq \sigma'/\sigma \leq \frac{S'_n}{S_m}\sqrt{f_{1-\alpha/2}^{m-1,n-1}}$$

Application Numérigue : On nous donne $f_{0.025}^{15,14} = 0.339$ et $f_{0.975}^{15,14} = 2.95$ d'où l'intervalle de confiance de niveau 95%

$$\sigma'/\sigma \in [0.537, 1.58]$$

Conclusion : la valeur $\sigma'/\sigma = 1$ appartient a lintervalle de confiance de niveau 95% ci-dessus done on ne peut pas rejeter l'hypothèse $\sigma' = \sigma$.

3 Intervalle de Confiance

Exercice 17. Des essais en laboratoire sur 20 lampes miniatures donnent les durées de vie suivantes, en heures :

$$451, 412, 412, 375, 407, 454, 375, 393, 355, 364, 414, 413, 345, 432, 392, 329, 439, 381, 451, 413.$$

On suppose la durée de vie distribuée normalement. Estimer par un intervalle de confiance 95% la durée de vie moyenne.

Solution :

corrigé succint : Les estimations ponctuelles de l'espéance, de l'écart-type et de la variance sont respectivement

$$\overline{x} = 400.35 \, s = 36.01 \; s^2 = 1297.$$

Alors $\sqrt{n}\frac{\bar{X}-\mu}{s}$ suit une loi de Student à 19 degrés de liberté, et donc l'espérance de $\bar{X}$ sera dans 95% des cas dans l'intervalle $[\bar{x} - t(0.95)s/\sqrt{n}, \bar{x} + t(0.95)s/\sqrt{n}]$, et on lit à partir de la table de la loi de Student à 19 degrés de liberté, $t(0.95) = 2.093$, donc l'intervalle cherché est :

$$\mu \in [383.5\, ,\, 417.2]$$

(complément hors-programme : estimation de l'écart-type par intervalle de confiance : X suit une loi normale, donc $(n-1)s^2/\sigma^2$ suit une loi du χ^2 à $n-1$ degré de liberté, donc σ^2 est avec une probabilité de 95% dans l'intervalle

$$\left[(n-1)s^2/c_1^2, (n-1)s^2/c_2^2\right]$$

pour $c_1^2 \simeq 32.8523$ et $c_2^2 \simeq 8.9065$ (lus dans la table du χ^2 à 19 degrés de liberté), soit $\sigma^2 \in [750.11, 2766.86]$, et donc l'écart-type a 95% de chances de vérifier $\sigma \in [27.39, 52.6]$

Exercice 18.

Une machine fabrique des billes métalliques dont le poids, mesuré en grammes, suit une loi normale. Nous prélevons au hasard 10 billes. Leurs poids sont

$$19,6; 20; 20,2; 20,1; 20; 19,9; 20; 20,3; 20,1; 19,8.$$

1. Quel est l'intervalle de confiance à 95% du poids des billes métalliques fabriquées ?
2. En réalité, l'écart-type σ de la population est connu et égal à $0,2$. Quel est l'intervalle de confiance à 95% du poids des billes métalliques fabriquées ?

Solution :

1. On calcule la moyenne $\widehat{\mu}$ de l'échantillon :

$$\widehat{\mu} = 20$$

Calculons la variance corrigée puis l'écart-type corrigé de l'échantillon à partir de la moyenne de l'échantillon

$$s_c^2 = \frac{10}{9}\left(\frac{19,6^2 + 20^2 + \cdots + 19,8^2}{10} - 20^2\right) = 0,04$$

puis

$$s_c = \sqrt{0,04} = 0,2$$

Dans la table de la loi de Student, pour 9ddl, on trouve

$$\mathbb{P}[|T| > 2,26] = 0,05 \text{ ou } \quad \mathbb{P}[|T| < 2,26] = 0,95.$$

L'intervalle de confiance pour le poids moyen est donc

$$\left[20 - 2,26 \times \tfrac{0,2}{\sqrt{10}}; 20 + 2,26 \times \tfrac{0,2}{\sqrt{10}}\right] \simeq [19,86; 20,14].$$

2. Si l'écart-type de la population est connu, on utilise la loi normale :

$$\mathrm{P}[|U| > 1,96] = 0,05 \text{ ou } \mathbb{P}[|U| < 1,96] = 0,95.$$

L'intervalle de confiance pour le poids moyen est donc :

$$\left[20 - 1,96 \times \tfrac{0,2}{\sqrt{10}}; 20 + 1,96 \times \tfrac{0,2}{\sqrt{10}}\right] \simeq [19,88; 20,12].$$

Exercice 19.

Voulant évaluer rapidement les résultats obtenus par ses 200 étudiants ingénieurs lors d'un partiel, un professeur décide de corriger quelques copies tirées au hasard. Il admet par ailleurs que les notes de ses étudiants suivent une loi normale de variance 4 .

1. Le professeur corrige un échantillon de 7 copies et trouve une moyenne de 11 . Quel est l'intervalle de confiance à 95% de la moyenne des 200 copies ?
2. Combien de copies le professeur doit-il corriger s'il veut situer la moyenne générale de ses étudiants dans un intervalle de confiance d'amplitude 2, avec un risque de 5%?
3. En trouvant une moyenne égale à 11, combien de copies le professeur devraitil corriger pour pouvoir dire, avec un risque de 1%, que la moyenne de tous les étudiants est supérieure à 10?

Solution :

La moyenne des notes

1. L'intervalle de confiance de la moyenne des 200 copies est :

$$\left[11 - 1,96 \times \tfrac{2}{\sqrt{7}}; 11 + 1,96 \times \tfrac{2}{\sqrt{7}}\right]$$
$$\simeq [9,52; 12,48].$$
1

2. Si l'amplitude de l'intervalle de confiance est égale à 2, on doit avoir

$$1,96 \times \frac{2}{\sqrt{n}} = 1$$

ce qui donne

$$n \simeq 15,4$$

En corrigeant 16 copies, l'enseignant peut situer la moyenne de ses étudiants.

3. Il faut que l'intervalle de confiance à 99% soit égal à $[10; 12]$. On doit donc avoir :

$$2,575 \times \frac{2}{\sqrt{n}} = 1$$

ce qui donne

$$n \simeq 26,5$$

Si l'enseignant corrige 27 copies et qu'il trouve une moyenne égale à 11, il peut dire que la moyenne de ses étudiants est supérieure à 10 , avec un risque d'erreur de 1%.

Exercice 20.

Une entreprise fabrique un certain type de composants électroniques dont la durée de vie X, exprimée en heures, est une variable aléatoire. Des mesures effectuées sur un échantillon aléatoire de taille 50 ont donné les résultats suivants :

$$\sum_{i=1}^{50} x_i = 60000; \quad \sum_{i=1}^{50} x_i^2 = 74 \times 10^6$$

1. Donner une estimation ponctuelle de la durée de vie moyenne des composants.
2. Donner une estimation ponctuelle de l'écart-type de cette durée de vie.
3. Donner l'intervalle de confiance à 95%, puis à 99% de cette durée de vie moyenne.
4. Quelle aurait du être la taille de l'échantillon pour que l'intervalle de confiance à 95% de la durée de vie moyenne des composants ait une amplitude de 60 heures ?

Solution :

1. La moyenne μ de la population est estimée par la moyenne de l'échantillon

$$\widehat{\mu} = \frac{60000}{50} = 1200$$

2. L'écart-type σ de la population est estimé à partir de l'écart-type s_c de l'échantillon :

$$s^2 = \tfrac{74 \times 10^6}{50} - 1200^2 = 40000$$
$$s_c^2 = s^2 \times \tfrac{50}{49} = 40816$$

D'où

$$s_c \simeq 202$$

3. La variance de la population étant estimée, on utilise la loi de Student. On trouve dans la table pour 49 ddl :

$$\mathbb{P}[|T| > 2,01] = 0,05 \text{ ou } \mathbb{P}[|T| < 2,01] = 0,95.$$

L'intervalle de confiance à 95% de la moyenne est :

$$\simeq [1143; 1257].$$

On trouve dans la table pour 49ddl :

$$\mathbb{P}[|T| > 2,68] = 0,01 \text{ ou } \mathbb{P}[|T| < 2,68] = 0,99$$

L'intervalle de confiance à 99% de la moyenne est :

$$\left[1200 - 2,68 \times \frac{202}{\sqrt{50}}; 1200 + 2,68 \times \frac{202}{\sqrt{50}}\right]$$
$$\simeq [1123; 1277].$$

4. Puisque l'on souhaite avoir une amplitude de 60 heures, la taille de l'échantillon est nécessairement supérieure à 50 et nous sommes dans les conditions d'utilisation de la loi normale. On doit avoir :

$$1,96 \times \frac{202}{\sqrt{n}} = 30$$

ce qui donne

$$n \simeq 175$$

Exercice 21.

À la veille d'une consultation électorale, nous effectuons un sondage.

1. Dans un échantillon représentatif de 1000 personnes, 500 personnes déclarent vouloir voter pour X, 250 pour Y et 50 pour Z. Donner les intervalles de confiance à 95% et 99% du pourcentage de personnes ayant l'intention de voter X, Y ou Z.
2. Nous évaluons le pourcentage de personnes ayant l'intention de voter pour un quatrième candidat, H, à 17% ? Combien faut-il interroger de personnes pour obtenir un intervalle de confiance à 95% du pourcentage de personnes ayant l'intention de voter H, avec une précision de 1%?

Solution :

1. Avec 1000 personnes, on peut déterminer un intervalle de confiance. L'intervalle de confiance à 95% de la proportion de personnes ayant l'intention de voter pour Monsieur Dupont est :

$$\left[0,5 - 1,96 \times \sqrt{\frac{0,5 \times 0,5}{1000}}; 0,5 + 1,96 \times \sqrt{\frac{0,5 \times 0,5}{1000}}\right]$$
$$\simeq [0,469; 0,531].$$

L'intervalle de confiance a 95% de la proportion de personnes ayant l'intention de voter pour Monsieur Durand est :

$$\left[0,25 - 1,96 \times \sqrt{\frac{0,25 \times 0,75}{1000}}; 0,25 + 1,96 \times \sqrt{\frac{0,25 \times 0,75}{1000}}\right]$$
$$\simeq [0,223; 0,277].$$

L'intervalle de confiance à 95% de la proportion de personnes ayant l'intention de voter pour Monsieur Duroc est

$$\left[0,05 - 1,96 \times \sqrt{\tfrac{0,05\times0,95}{1000}}; 0,05 + 1,96 \times \sqrt{\tfrac{0,05\times0,95}{1000}}\right]$$
$$\simeq [0,036; 0,064].$$

L'intervalle de confiance à 99% de la proportion de personnes ayant l'intention de voter pour Monsieur Dupont est

$$\left[0,5 - 2,575 \times \sqrt{\tfrac{0,5\times0,5}{1000}}; 0,5 + 2,575 \times \sqrt{\tfrac{0,5\times0,5}{1000}}\right]$$
$$\simeq [0,459; 0,541].$$

L'intervalle de confiance à 99% de la proportion de personnes ayant l'intention de voter pour Monsieur Durand est :

$$\left[0,25 - 2,575 \times \sqrt{\tfrac{0,25\times0,75}{1000}}; 0,25 + 2,575 \times \sqrt{\tfrac{0,25\times0,75}{1000}}\right]$$
$$\simeq [0,215; 0,285].$$

L'intervalle de confiance a 99% de la proportion de personnes ayant l'intention de voter pour Monsieur Duroc est :

$$\left[0,05 - 2,575 \times \sqrt{\tfrac{0,05\times0,95}{1000}}; 0,05 + 2,575 \times \sqrt{\tfrac{0,05\times0,95}{1000}}\right]$$
$$\simeq [0,032; 0,068].$$

2. Pour un échantillon de taille n (on suppose $n > 1000$), l'intervalle de confiance à 95% du pourcentage de personness ayant l'intention de voter Duval est

$$\left[0,17 - 1,96 \times \sqrt{\frac{0,17 \times 0,83}{n}}; 0,17 + 1,96 \times \sqrt{\frac{0,17 \times 0,83}{n}}\right]$$

Puisque l'on veut une précision de 1%, cet intervalle de confiance doit être l'intervalle $[0,16; 0,18]$ Et on doit avoir

$$1,96 \times \sqrt{\frac{0,17 \times 0,83}{n}} = 0,01$$

ce qui donne

$$n \simeq 5420.$$

Exercice 22.

On veut étudier la proportion p de gens qui vont au cinéma chaque mois. On prend donc un échantillon de taille $n = 100$. Soit N le nombre de personnes dans l'échantillon qui vont au cinéma mensuellement.

1. Quelle est la loi de N ? Par quelle loi peut-on l'approcher et pourquoi ? En déduire une approximation de la loi de $F = N/n$.
2. On observe une proportion f de gens qui vont chaque mois au cinéma. Donner la forme d'un intervalle de confiance pour p, de niveau de confiance $1 - \alpha$.

3. Applications numériques : $f = 0,1, 1-\alpha = 90\%, 95\%, 98\%$.

Solution :

1. On suppose que les personnes ont bien été interrogées indépendamment. Ainsi, on a un schéma de Bernoulli : une personne interrogée va au cinéma chaque mois $\longrightarrow$ SUCCES, sinon, ECHEC. Et donc N suit une loi binomiale $\mathcal{B}(n = 100, p)$

$$\mathbb{P}[X = k] = \begin{pmatrix} 100 \\ k \end{pmatrix} p^k (1-p)^{100-k}, \quad k = 0, \ldots, 100$$

   Comme $n \geq 20$, si $np > 5$ et $n(1-p) > 5$ (à vérifier lors de l'application numérique), on peut approcher cette loi par la loi normale $\mathcal{N}(np, \sqrt{np(1-p)})$, et donc F suit approximativement la loi $\mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

2. IC $\left[f - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}, f + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right]$ où $\mathbb{P}\left[Z \geq z_{\alpha/2}\right] = \alpha/2, Z$ de loi normale centrée réduite, $1-\alpha$ est le niveau de confiance.

3. $f = 0.1$
   $-1-\alpha = 90\%, z_{\alpha/2} = 1.645$, IC [0.05,0.15]
   $-1-\alpha = 95\%, z_{\alpha/2} = 1.96$, IC [0.04,0.16]
   $-1-\alpha = 98\%, z_{\alpha/2} = 2.326$, IC [0.03,0.17].

Exercice 23.

On suppose que le poids d'un nouveau né est une variable normale d'écart-type égal à $0,5$ kg. Le poids moyen des 49 enfants nés au mois de Décembre 2019 dans l'hôpital de la ville a été de $3,6$ kg.

1. Déterminer un intervalle de confiance à 95% pour le poids moyen d'un nouveau né dans cet hôpital.
2. Quel serait le niveau de confiance d'un intervalle de longueur 0,1 kg centré en 3,6 pour ce poids moyen ?

Solution :

1. IC de niveau de confiance 95% pour le poids moyen ($z_{\alpha/2} = 1.96$) :

$$\left[\bar{x} - 1.96\frac{\sigma}{7}, \bar{x} + z_{\alpha/2}\frac{\sigma}{7}\right] = [3.46, 3.74]$$

2.

$$\begin{aligned} \mathbb{P}[\bar{X} - 0.05 \leq m \leq \bar{X} + 0.05] &= \mathbb{P}\left[\frac{-0.05}{\sigma/7} \leq \frac{\bar{X}-m}{\sigma/\sqrt{n}} \leq \frac{0.05}{\sigma/7}\right] \\ &= 2F\left(\frac{0.05}{0.5/7}\right) = 2F(0.7) - 1 = 2 * 0.758 - 1 = 0.516. \end{aligned}$$

   Le niveau de confiance est donc 0.516.