TD N°1 : Estimation

Sommaire

Rappels

CONVERGENCES

Convergence presque sûre

$(X_n)_{n\in N}$ une suite de VA converge presque sûrement vers la VA X si $P\left(\lim_{n\to+\infty} X_n = X\right)=1$. On le note $X_n \xrightarrow{ps} X$.

Convergence en probabilité

$(X_n)_{n\in N}$ une suite de VA converge presque sûrement vers la VA X si $\forall\varepsilon>0\, P\left(\left|X_n - X\right|>\varepsilon\right)=0$. On le note $X_n \xrightarrow{p} X$.

Convergence en loi

$(X_n)_{n\in N}$ une suite de VA converge presque sûrement vers la VA X si $\forall x>0 \lim_{n\to+\infty} F_{X_n}(x)=F_X(x)$. On le note $X_n \xrightarrow{L} X$.

Convergence en moyenne d'ordre r

$(X_n)_{n\in N}$ une suite de VA converge en moyenne d'ordre r vers la VA X si $\lim_{n\to+\infty} E\left(\left|X_n - X\right|^r\right)=0$.

Liens entre les convergences

$$\left.\begin{array}{c} \textit{convergence presque sûre} \\ \textit{ou} \\ \textit{convergence en moyenne d'ordre } r \end{array}\right\} \Rightarrow \textit{convergence en probabilité} \Rightarrow \textit{convergence en loi}$$

LOI DES GRANDS NOMBRES

Loi forte des grands nombres

Soit $(X_n)_{n\in N}$ une suite de VA indépendantes identiquement distribuées ayant une moyenne finie m alors $P\left(\lim_{n\to+\infty} \overline{X}_n = m\right)=1$ (convergence presque sûre).

Loi faible des grands nombres

Soit $(X_n)_{n\in N}$ une suite de VA indépendantes identiquement distribuées ayant une moyenne finie m et une variance σ^2 finie alors $\forall\varepsilon>0 \lim_{n\to+\infty} P\left(\left|\overline{X}_n - E(X)\right|\geq\varepsilon\right)=0$ (convergence en probabilité).

THEOREME CENTRAL LIMITE

Soit $(X_n)_{n\in N}$ une suite de VA indépendantes identiquement distribuées ayant une moyenne finie m et un écart-type fini σ alors $\frac{\overline{X}_n - m}{\sigma / \sqrt{n}} \xrightarrow{L} N(0,1)$ (convergence en loi). Il est équivalent de dire que $\sqrt{n}\left(\overline{X}_n - m\right) \xrightarrow{L} N(0,\sigma^2)$

PROPRIETES SUR LES VARIABLES GAUSSIENNES

- Soit $(X_n)_{n\in N}$ une suite de VA indépendantes identiquement distribuées de loi $N(m,\sigma^2)$ alors $\overline{X}_n \xrightarrow{L} N(m, \sigma^2/n)$.
- Soit $(X_n)_{n\in N}$ une suite de VA indépendantes identiquement distribuées de loi $N(m,\sigma^2)$ et $S^{*2} = 1/(n-1) * \sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$ la variance empirique corrigée alors $Y_n = \sqrt{n}\,\frac{\overline{X}_n - m}{S^*}$ suit une loi de Student à n-1 degré de liberté.

Exercices de base

EXERCICE 1

Enoncé

Suite à une étude réalisée dans une société, on a constaté que le montant des factures suit une loi d'espérance 5000€ avec un écart-type de 2000€.
Sur le 1er trimestre, 100 factures ont été émises. Quelle est la probabilité pour le montant moyen des factures dépasse 5100€ ?

Corrigé

Soit X_i le montant d'une facture. On sait que $E(X_i)$=5000€ et $\sigma(X_i)$=2000€. Soit $\overline{X}$ le montant moyen de n factures. On cherche

$$P(\overline{X} \geq 5100).$$

Si n est suffisamment grand (ici n=100 ok), alors on peut approcher la loi de $\overline{X}$ grâce au TCL,

$$\overline{X} \sim N(5000, \frac{2000^2}{n}).$$

D'où

$$P(\overline{X} \geq 5100) = P(\frac{\overline{X} - 5000}{2000}\sqrt{n} \geq \frac{5100 - 5000}{2000}\sqrt{n}) = P(Z \geq 0.5) = 0.3085$$

Il y a donc 31% de chance que les factures dépassent un montant de 5100€.

EXERCICE 2

ENONCE

Soit X le temps d'attente avant d'être servi au RU. Une série d'observations journalières donne l'échantillon suivant (en min)

10	11	15	8	9	12	13	11	10	7	19	5	10	6	12	15	16	5	6	10
9	10	14	7	8	11	12	10	9	12	15	7	9	5	11	14	15	8	9	9

On calcule la moyenne $\overline{x}$ =10.35 min et l'écart-type empirique s=3.3 min.

1) Quelle loi suit X ?
1) Comment déterminer le paramètre de cette loi à partir de l'échantillon ?
2) En déduire la probabilité qu'un étudiant attende plus de 15 min avant d'être servi.

Corrigé

1) Temps d'attente donc $X \rightsquigarrow Exp(\theta)$

2) D'après la loi des grands nombres si $E(X_i)$ et $var(X_i)$ existent

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n \to +\infty]{pr} E(X_i)$$

si on considère les observation comme des réalisations de v.a i.i.d de loi $Exp(\theta)$ alors

$$E(X_i) = \frac{1}{\theta} \quad \left(\text{et } var(X_i) = \frac{1}{\theta^2} \text{ existe}\right)$$

On peut donc envisager approcher θ par $\frac{1}{\overline{x}}$

3) $P(X > 15) = 1 - F_X(15) = e^{-15\theta} = e^{-\frac{15}{\overline{x}}} = 0.23$

EXERCICE 3

Enoncé

Un expérimentateur utilise un thermomètre sophistiqué mais présentant une incertitude de mesure. Afin d'estimer cette incertitude, il effectue une série de mesures sur de l'eau en ébullition. Soit T la température mesurée.

1) Quelle loi suit T ?
2) Comment déterminer le paramètre de cette loi à partir des mesures ?

Corrigé

1) L'eau en ébullition est à 100°C. On peut donc considérer que $T_i \leadsto \mathcal{N}(100, \sigma^2)$ car mesures

2) Si on considère les mesures comme les réalisations de v.a iid T_i de loi $\mathcal{N}(100, \sigma^2)$ alors

$$E(T_i^2) = \operatorname{var} T_i + (E(T_i))^2 = \sigma^2 + 100^2$$

$\operatorname{Var}(T_i^2)$ existe (moment d'ordre 4 se calcule par $M_T^{(4)}(0)$)

donc d'après la loi des grands nombres

$$\frac{1}{n}\sum_{i=1}^{n} T_i^2 \xrightarrow[n \to +\infty]{pr} E(T_i)^2 = \sigma^2 + 100^2$$

On peut donc approcher σ^2 par la moyenne des carrés des observations moins 100^2

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} t_i^2 - 100^2$$

EXERCICE 4 (CARTE DE CONTROLE)

Enoncé

On a construit une machine produisant des engrenages d'un diamètre de 60 mm avec un écart-type de 1.5 mm. Pour déterminer si la machine est en bon état de marche, on décide de prélever un échantillon de 9 engrenages toutes les 2 heures et, à partir de cet échantillon, de calculer la moyenne des diamètres.

1) Quel type de problème doit-on résoudre ?
2) Déterminer la loi de la variable aléatoire utilisée.
3) Etablir une règle de décision pour laquelle on puisse être certain avec un niveau de confiance de 95% que la qualité de la production est conforme à la normale.

Corrigé

1. Si la machine est en bonne état alors dans la population mère les engrenages ont un diamètre $\mu = 60$ mm avec un écart-type $\sigma = 1.5$ mm
On prélève un échantillon de 9 engrenages et on calcule $\bar{x}$ et s^2
On doit alors vérifier si ces valeurs sont acceptables, i.e si la machine fonctionne toujours bien. Pour cela, on doit établir un intervalle dans lequel le diamètre moyen a de grande chance de se trouver, i.e

$\bar{x}$ est une réalisation de $\bar{X}$ $\quad P[\bar{X} \in [a,b]] = 1-\alpha$ où α = risque toléré

Si $\bar{x} \in [a,b]$, on accepte que l'échantillon appartient à la population mère, sinon on ne l'accepte pas $\Rightarrow$ pb machine.

2. Soit X_i le diamètre d'un engrenage. On suppose que X_i suit une loi normale (c'est très souvent le cas lorsqu'il s'agit de mesures) donc ici $X_i \hookrightarrow \mathcal{N}(60, 2.25)$

3. On sait alors que

$$\frac{\bar{X}-\mu}{\sigma}\sqrt{n} \hookrightarrow \mathcal{N}(0,1)$$

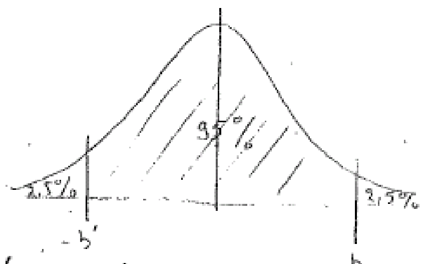
On cherche a et b tq $P(\bar{X} \in [a,b]) = 0.95$

$$P(a \leq \bar{X} \leq b) = P(a-\mu \leq \bar{X}-\mu \leq b-\mu)$$

$$= P\left(\frac{a-\mu}{\sigma}\sqrt{n} \leq \frac{\bar{X}-\mu}{\sigma}\sqrt{n} \leq \frac{b-\mu}{\sigma}\sqrt{n}\right)$$

$$= F(b') - F(a')$$ où F est la fonction de répartition de $\mathcal{N}(0,1)$

On suppose que le risque est symétrique, i.e $a' = -b'$



Donc $P(\bar{X} \in [a,b]) = 0.95 \Leftrightarrow F(b') - F(a') = 0.95$

$\Leftrightarrow F(b') - (1 - F(b')) = 0.95 \Leftrightarrow 2F(b') - 1 = 0.95 \Leftrightarrow F(b') = \frac{1.95}{2} = 0.97$

On obtient alors b' à l'aide des tables de $\mathcal{N}(0,1)$

On obtient $b' = 1.96$

d'où $b = \frac{\sigma}{\sqrt{n}} b' + \mu = \frac{1.5}{3} \times 1.96 + 60 = 60.98$

$a = \mu - \frac{\sigma}{\sqrt{n}} b' = 59.02$

$\Rightarrow P[59.02 \leq \bar{X} \leq 60.98] = 0.95$

A partir de l'échantillon, on calcule $\bar{x}$ et on regarde si $\bar{x} \in [59.02, 60.98]$

Exercice sur les intervalles de confiance

EXERCICE 5

Enoncé

On mesure la force de compression d'un ciment en moulant de petits cylindres et en mesurant la pression X (exprimée en kg/cm²) à partir de laquelle ils se cassent. On suppose que X suit une loi $N(\mu,\sigma^2)$.

1) Calculer un I.D.C symétrique avec un risque de 10% pour les paramètres inconnus pour un échantillon de taille n=9 où $\bar{x} = 19.72\,\text{kg/cm}^2$ et $s^{*2} = 0.69$.
2) Expliquer la différence entre le problème ci-dessus et l'exercice sur les cartes de contrôle.
3) Calculer un I.D.C dissymétrique pour μ avec un risque de 2.5% à gauche et 7.5% à droite pour un échantillon de taille n=9 où $\bar{x} = 19.72\,\text{kg/cm}^2$ et $s^{*2} = 0.69$. Comparer avec le résultat de la question 1.
4) On suppose maintenant que σ^2=0.69. Calculer un I.D.C symétrique pour μ avec un risque de 10% pour un échantillon de taille n=9 où $\bar{x} = 19.72\,\text{kg/cm}^2$. Comparer avec le résultat de la question 1.

Corrigé

1) IDC pour μ

On cherche l'intervalle [a,b] tel que

$$P(a \leq \mu \leq b) = 1-\alpha = 0.9$$

On utilise la moyenne $\bar{X}$ comme estimateur de μ. On sait que l'échantillon est gaussien avec σ² inconnu donc

$$Z = \sqrt{n}\frac{\bar{X}-\mu}{S} \sim t_{n-1}$$

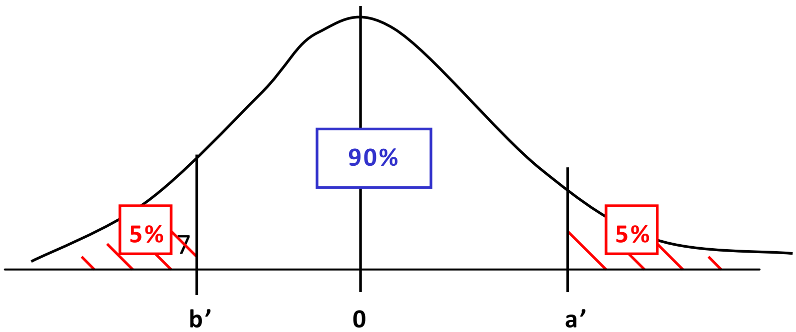
D'où

$$P(-b \leq -\mu \leq -a) = P(\sqrt{n}\frac{\bar{x}-b}{s} \leq \sqrt{n}\frac{\bar{X}-\mu}{S} \leq \sqrt{n}\frac{\bar{x}-a}{s}) = P(b' \leq Z \leq a') = 0.9$$

On a deux inconnues a' et b' et une seule équation. On suppose donc que le risque est symétrique. De plus la loi de Student est symétrique par rapport à l'axe Oy donc

$$P(b' \leq Z \leq a') = 0.9 \quad \Leftrightarrow \begin{cases} b' = -a' \\ P(Z \geq a') = 0.05 \end{cases}$$

$$\Rightarrow \begin{cases} b' = -a' \\ a' = 1.86 \end{cases} \text{(table } t_8\text{)}$$



$$\Rightarrow \begin{cases} a = \bar{x} - a'\dfrac{s}{\sqrt{n}} = \bar{x} - 1.86\dfrac{s}{3} \\ b = \bar{x} + a'\dfrac{s}{\sqrt{n}} = \bar{x} + 1.86\dfrac{s}{3} \end{cases}$$

Donc la valeur moyenne de la pression sur toute la population des cylindres a 90% de chance de se trouver entre 19.23kg/cm² et 20.23 kg/cm².

<u>IDC pour σ²</u>

On cherche l'intervalle [a,b] tel que

$$P(a \le \sigma^2 \le b) = 1 - \alpha = 0.9.$$

On utilise la variance empirique sans biais, S^{*2}, comme estimateur de σ². L'échantillon est gaussien donc

$$Z = (n-1)S^{*2}/\sigma^2 \sim \chi_{n-1}$$

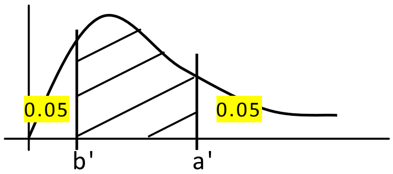
D'où

$$P(a \le \sigma^2 \le b) = P[\frac{ns^2}{b} \le \frac{nS^2}{\sigma^2} \le \frac{ns^2}{a}] = P(b' \le Z \le a') = 0.9$$

Là encore il y a deux inconnues pour une seule équation. On utilise donc le risque symétrique mais attention, la loi du chi-deux n'est pas symétrique par rapport à l'axe Oy, d'où

$$P(a \le \sigma^2 \le b) = 0.9 \Leftrightarrow \begin{cases} P(Z \ge a') = 0.05 \\ P(Z \ge b') = 0.95 \end{cases}$$

$$\Rightarrow \begin{cases} a' = 15.507 \\ b' = 2.733 \end{cases} \text{(table } \chi_8\text{)} \quad \Rightarrow \begin{cases} a = \dfrac{8s^2}{15.507} \\ b = \dfrac{8s^2}{2.733} \end{cases}$$



Donc l'écart-type de la pression sur toute la population des cylindres a 90% de chance de se trouver entre 0.60kg/cm² et 1.42kg/cm².

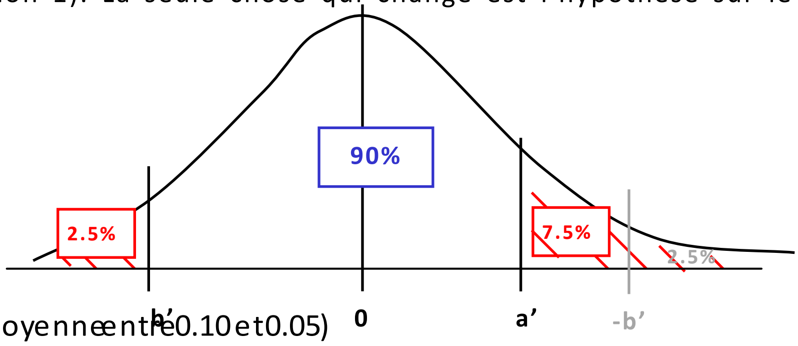
2) Dans cet exercice on cherche une estimation (par IDC) d'une valeur relative à toute la population. C'est-à-dire, on dispose d'un échantillon (un sous-ensemble de la population) et on essaie d'en tirer une généralité sur la population entière. Dans le cas des cartes de contrôle, on connaît les caractéristiques de la population, pas besoin de les estimer. En revanche, on cherche à savoir si un échantillon est conforme à ces caractéristiques.

3) Reprenons l'IDC pour μ de la question 1). La seule chose qui change est l'hypothèse sur le risque symétrique.

On a maintenant,

$$P(b' \le Z \le a') = 0.9 \Leftrightarrow \begin{cases} P(Z \ge b') = 0.975 \\ P(Z \ge a') = 0.075 \end{cases}$$

$$\Leftrightarrow \begin{cases} P(Z \ge -b') = 0.025 \\ P(Z \ge a') = 0.075 \end{cases}$$



$$\Rightarrow \text{(table } t_8\text{)} \begin{cases} -b' = 2.306 \\ a' = (1.397 + 1.860)/2 = 1.628 \quad \text{(moyenne entre 0.10 et 0.05)} \end{cases}$$

$$\Rightarrow \begin{cases} a = \bar{x} - 1.628\dfrac{s}{3} \\ b = \bar{x} + 2.306\dfrac{s}{3} \end{cases}$$

Donc la valeur moyenne de la pression sur toute la population des cylindres a 90% de chance de se trouver entre 19.17kg/cm² et 20.36 kg/cm². (à vérifier)

4) Reprenons l'IDC de μ de la question 1). On a une information supplémentaire, à savoir que la variance de la population est connue. On sait alors que

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

On obtient alors

$$P(b'\le Z\le a')=0.9 \Leftrightarrow \begin{cases} b'=-a' \\ P(Z\ge a')=0.05 \end{cases}$$

$$\Rightarrow \begin{cases} b'=-a' \\ a'=0.4801 \end{cases} \text{(table N(0,1))} \Rightarrow \begin{cases} a=\bar{x}-a'\frac{\sigma}{\sqrt{n}}=\bar{x}-0.4801\frac{\sigma}{3} \\ b=\bar{x}+a'\frac{\sigma}{\sqrt{n}}=\bar{x}+0.4801\frac{\sigma}{3} \end{cases}$$

Donc la valeur moyenne de la pression sur toute la population des cylindres a 90% de chance de se trouver entre 19.85kg/cm² et 19.58kg/cm².

Le fait d'avoir une information supplémentaire augmente la précision de l'estimation.

5) IDC pour µ

On cherche l'intervalle [a,b] tel que

$$P(a\le\mu\le b)=1-\alpha=0.9$$

On utilise la moyenne $\bar{X}$ comme estimateur de µ. On sait que l'échantillon est gaussien avec σ² inconnu donc

$$Z=\sqrt{n}\frac{\bar{X}-\mu}{S}\sim t_{n-1}$$

D'où

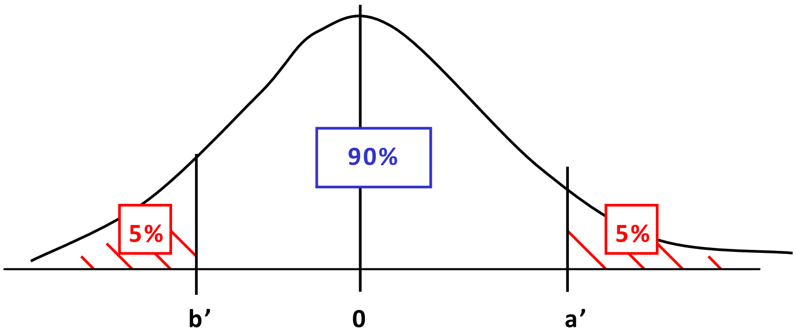
$$P(-b\le-\mu\le-a)=P(\sqrt{n}\frac{\bar{x}-b}{s}\le\sqrt{n}\frac{\bar{X}-\mu}{S}\le\sqrt{n}\frac{\bar{x}-a}{s})=P(b'\le Z\le a')=0.9$$

On a deux inconnues a' et b' et une seule équation. On suppose donc que le risque est symétrique. De plus la loi de Student est symétrique par rapport à l'axe Oy donc

$$P(b'\le Z\le a')=0.9 \Leftrightarrow \begin{cases} b'=-a' \\ P(Z\ge a')=0.05 \end{cases}$$

$$\Rightarrow \begin{cases} b'=-a' \\ a'=1.86 \end{cases} \text{(table } t_8\text{)}$$

$$\Rightarrow \begin{cases} a=\bar{x}-a'\frac{s}{\sqrt{n}}=\bar{x}-1.86\frac{s}{3} \\ b=\bar{x}+a'\frac{s}{\sqrt{n}}=\bar{x}+1.86\frac{s}{3} \end{cases}$$



Donc la valeur moyenne de la pression sur toute la population des cylindres a 90% de chance de se trouver entre 19.23kg/cm² et 20.23 kg/cm².

IDC pour σ²

On cherche l'intervalle [a,b] tel que

$$P(a\le\sigma^2\le b)=1-\alpha=0.9.$$

On utilise la variance empirique sans biais, S^{*2}, comme estimateur de σ². L'échantillon est gaussien donc

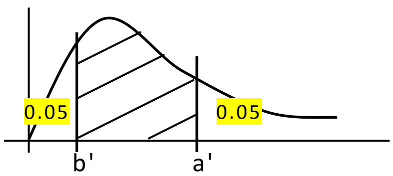
$$Z=(n-1)S^{*2}/\sigma^2\sim\chi_{n-1}$$

D'où

$$P(a\le\sigma^2\le b)=P[\frac{ns^2}{b}\le\frac{nS^2}{\sigma^2}\le\frac{ns^2}{a}]=P(b'\le Z\le a')=0.9$$

Là encore il y a deux inconnues pour une seule équation. On utilise donc le risque symétrique mais attention, la loi du chi-deux n'est pas symétrique par rapport à l'axe Oy, d'où

$$P(a\le\sigma^2\le b)=0.9 \Leftrightarrow \begin{cases} P(Z\ge a')=0.05 \\ P(Z\ge b')=0.95 \end{cases}$$

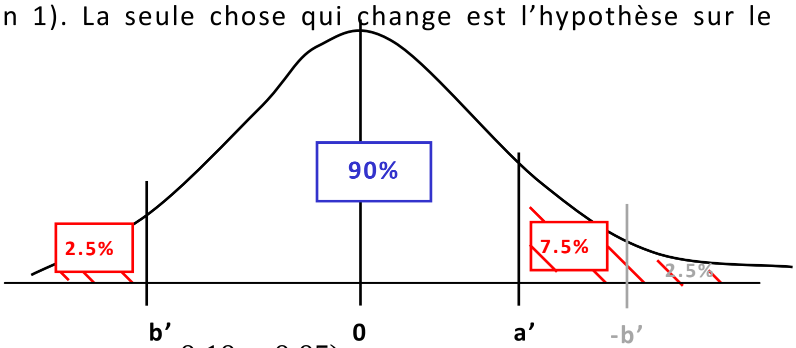


$$\Rightarrow \begin{cases} a'=15.507 \\ b'=2.733 \end{cases} \text{(table } \chi_8\text{)} \qquad \Rightarrow \begin{cases} a=\dfrac{8s^2}{15.507} \\ b=\dfrac{8s^2}{2.733} \end{cases}$$

Donc l'écart-type de la pression sur toute la population des cylindres a 90% de chance de se trouver entre 0.60kg/cm² et 1.42kg/cm².

6) Dans cet exercice on cherche une estimation (par IDC) d'une valeur relative à toute la population. C'est-à-dire, on dispose d'un échantillon (un sous-ensemble de la population) et on essaie d'en tirer une généralité sur la population entière. Dans le cas des cartes de contrôle, on connaît les caractéristiques de la population, pas besoin de les estimer. En revanche, on cherche à savoir si un échantillon est conforme à ces caractéristiques.
7) Reprenons l'IDC pour µ de la question 1). La seule chose qui change est l'hypothèse sur le risque symétrique.

On a maintenant,



$$P(b'\le Z\le a')=0.9 \quad \Leftrightarrow \begin{cases} P(Z\ge b')=0.975 \\ P(Z\ge a')=0.075 \end{cases}$$

$$\Leftrightarrow \begin{cases} P(Z\ge -b')=0.025 \\ P(Z\ge a')=0.075 \end{cases}$$

$$\Rightarrow \text{(table } t_8\text{)} \begin{cases} -b'=2.306 \\ a'=(1.397+1.860)/2=1.628 \quad (\textit{moyenne entre } 0.10 \textit{ et } 0.05) \end{cases}$$

$$\Rightarrow \begin{cases} a=\bar{x}-1.628\dfrac{s}{3} \\ b=\bar{x}+2.306\dfrac{s}{3} \end{cases}$$

Donc la valeur moyenne de la pression sur toute la population des cylindres a 90% de chance de se trouver entre 19.17kg/cm² et 20.36 kg/cm². (à vérifier)

8) Reprenons l'IDC de µ de la question 1). On a une information supplémentaire, à savoir que la variance de la population est connue. On sait alors que

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

On obtient alors

$$P(b'\le Z\le a')=0.9 \;\Leftrightarrow \begin{cases} b'=-a' \\ P(Z\ge a')=0.05 \end{cases}$$

$$\Rightarrow \begin{cases} b'=-a' \\ a'=0.4801 \end{cases} \text{(table N(0,1))} \;\Rightarrow \begin{cases} a=\bar{x}-a'\dfrac{\sigma}{\sqrt{n}}=\bar{x}-0.4801\dfrac{\sigma}{3} \\ b=\bar{x}+a'\dfrac{\sigma}{\sqrt{n}}=\bar{x}+0.4801\dfrac{\sigma}{3} \end{cases}$$

Donc la valeur moyenne de la pression sur toute la population des cylindres a 90% de chance de se trouver entre 19.85kg/cm² et 19.58kg/cm².

Le fait d'avoir une information supplémentaire augmente la précision de l'estimation.

EXERCICE 6

Enoncé

Le New York Times du 18 août 1985 publiait les résultats d'un sondage sur la punition corporelle des enfants. Sur les 604 parents interrogés, 63% étaient favorables aux punitions corporelles dans les écoles. Donner un intervalle de confiance de niveau 5% de la proportion de la population parentale favorable à ce genre de traitement?

Corrigé

Soit p la proportion de parents favorables. On cherche a et b tels que

$$P[a \le p \le b] = 0.95$$

Soit F_n la fréquence empirique estimateur usuel de p. Étant donné que l'échantillon est de grande taille, on peut considérer que

$$F_n \rightsquigarrow \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

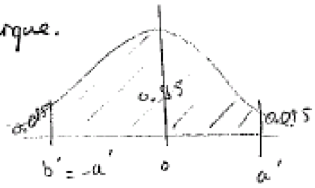
$$\Leftrightarrow Z = (F_n - p)\sqrt{\frac{n}{p(1-p)}} \rightsquigarrow \mathcal{N}(0,1)$$

D'où

$$P[a \le p \le b] = P[-b \le -p \le -a] = P\left[(f_n - b)\sqrt{\frac{n}{p(1-p)}} \le (F_n - p)\sqrt{\frac{n}{p(1-p)}} \le (f_n - a)\sqrt{\frac{n}{p(1-p)}}\right]$$

$$= P[b' \le Z \le a'] = 0.95$$

On suppose un risque symétrique.
De plus la loi $\mathcal{N}(0,1)$ est symétrique par rapport à Oy d'où



$$\Leftrightarrow \begin{cases} b' = -a' \\ P[Z \le a'] = 0.975 \end{cases}$$

$$\Leftrightarrow \begin{cases} a' = 1.96 \\ b' = -1.96 \end{cases} \Leftrightarrow \begin{cases} a = f_n - 1.96\sqrt{\frac{p(1-p)}{n}} \\ b = f_n + 1.96\sqrt{\frac{p(1-p)}{n}} \end{cases}$$

or quand n est grand ; on peut approcher a et b par

$$\begin{cases} a = f_n - 1.96\sqrt{\frac{f_n(1-f_n)}{n}} \\ b = f_n + 1.96\sqrt{\frac{f_n(1-f_n)}{n}} \end{cases}$$

Ici $f_n = 0.63$ d'où

$$\begin{cases} a = 0.63 - 0.038 \\ b = 0.63 + 0.038 \end{cases}$$

D'où $p(0,59 \le p \le 0.67) = 0.95$.

EXERCICE 7

Enoncé

La Monnaie de Paris dispose à Pessac en Gironde d'un établissement monétaire, qui produit en grande série des pièces de monnaie courante pour la France et de nombreux pays. Le poids d'une pièce de 2€ est une variable aléatoire d'espérance μ et d'écart-type de 10^{-1} grammes. On prélève pour un échantillon de 100 pièces et on trouve un poids moyen de 8.55 grammes.

1) Déterminer un intervalle dans lequel le poids d'une pièce de 2€ à 99% de chance de se trouver.
2) Combien faudrait-il prélever de pièces pour avoir une précision inférieure à 10^{-2} grammes ?

Corrigé

1) Soit μ le poids moyen des pièces de 2€. On cherche a et b tq

$$P[a \leq \mu \leq b] = 0.99$$

la moyenne $\overline{X}$ est un estimateur usuel de μ. L'échantillon est grand et l'écart-type est connu. On peut donc supposer que

$$\overline{X} \hookrightarrow \mathcal{N}\left(\mu, \frac{(10^{-1})^2}{n}\right)$$

$$\Leftrightarrow Z = (\overline{X} - \mu)\frac{\sqrt{n}}{10^{-1}} \hookrightarrow \mathcal{N}(0,1)$$

Donc

$$P[a \leq \mu \leq b] = P[-b \leq -\mu \leq -a] = P\left[(\bar{x} - b)\frac{\sqrt{n}}{10^{-1}} \leq (\overline{X} - \mu)\frac{\sqrt{n}}{10^{-1}} \leq (\bar{x} - a)\frac{\sqrt{n}}{10^{-1}}\right]$$

$$= P[b' \leq Z \leq a'] = 0.99$$

On suppose un risque symétrique
De plus la loi $\mathcal{N}(0,1)$ est symétrique par rapport à Oy d'où

$$\Leftrightarrow \begin{cases} b' = -a' \\ P[Z \leq a'] = 0.995 \end{cases}$$

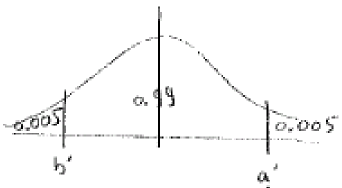


table $\Rightarrow \begin{cases} a' = 2.57 \\ b' = -2.57 \end{cases} \Rightarrow \begin{cases} a = \bar{x} - 2.57 \times \frac{10^{-1}}{\sqrt{n}} = 8.55 - 2.57 \times 10^{-2} = 8.52 \\ b = \bar{x} + 2.57 \times \frac{10^{-1}}{\sqrt{n}} = 8.55 + 2.57 \times 10^{-2} = 8.56 \end{cases}$

$$P[8.52 \leq \mu \leq 8.56]$$

2) On cherche n tq tq $b-a = 10^{-2}$

or d'après ce qui précède

$$b-a = 2 \times 2.57 \times \frac{10^{-1}}{\sqrt{n}} = 0.514 \times \frac{1}{\sqrt{n}} = 10^{-2}$$

$$\Leftrightarrow \quad \sqrt{n} = 51.4 \quad \Leftrightarrow \quad n \geq 2641,96$$

Il faut donc prélever au moins 2642 pièces.

EXERCICE 8

Enoncé

Un échantillon de 50 abonnés au journal Le Monde a révélé qu'un abonné passait en moyenne $\bar{x}=6.7$ heures par semaine à consulter internet. L'écart type de cet échantillon vaut $s^* = 5.8$ heures. Donner un intervalle de confiance de niveau 95% pour le temps moyen passé à consulter Internet dans la population des abonnés.

Corrigé

Soit X le temps passé sur internet par un abonné. L'échantillon n'est pas très grand et l'écart-type est inconnu. On doit donc supposer que X suit une loi normale

$$X \leadsto \mathcal{N}(\mu, \sigma^2)$$

avec σ^2 inconnu. On cherche a et b tq

$$P[a \leq \mu \leq b] = 0.95$$

$\bar{X}$ est un estimateur usuel de μ et

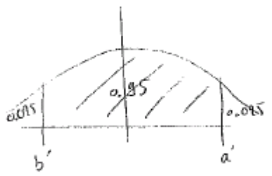
$$Z = \sqrt{n}\,\frac{\bar{X}-\mu}{S^*} \leadsto t_{n-1}$$

d'où

$$P[a \leq \mu \leq b] = P[-b \leq -\mu \leq -a] = P\left[\sqrt{n}\,\frac{(\bar{x}-b)}{s^*} \leq \sqrt{n}\left(\frac{\bar{X}-\mu}{s^*}\right) \leq \sqrt{n}\,\frac{(\bar{x}-a)}{s^*}\right]$$

$$= P[b' \leq Z \leq a'] = 0.95$$

On suppose un risque symétrique. De plus, la loi de Student est symétrique par rapport à Oy d'où



$$\begin{cases} b' = -a' \\ P[Z \geq a'] = 0.025 \end{cases}$$

t_{100}

$$\Rightarrow \begin{cases} b' = -2 \\ a' = 2 \end{cases} \Rightarrow \begin{cases} a = \bar{x} - \frac{2 \times s^*}{\sqrt{n}} = \bar{x} - 1.64 \\ b = \bar{x} + \frac{2 \times s^*}{\sqrt{n}} = \bar{x} + 1.64 \end{cases}$$

$$P[5.06 \leq \mu \leq 8.34] = 0.95$$

Exercices sur les qualités d'un estimateur

EXERCICE 9

Enoncé

Soit X une variable aléatoire de densité $f(x;\theta)=\begin{cases}\frac{2}{\theta}(1-\frac{x}{\theta}) & \text{si}\, 0\leq x\leq\theta \\ 0 & \text{sinon}\end{cases}$

1) Calculer E(X) et V(X).
2) On cherche à estimer θ à partir d'un échantillon $X_1,\ldots,X_n$ de même loi que X.

   On utilise l'estimateur $T_n=\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$.

   (a) Calculer le biais de cet estimateur. Pouvez-vous lui éliminer son biais ?
   (b) Etablir la convergence en probabilité de T_n (modifié).
   (c) Déterminer le risque quadratique puis la convergence en moyenne quadratique.
   (d) Déterminer la loi asymptotique de T_n.

Corrigé

1) $E(X)=\int x f(x;\theta)dx=\int_0^\theta x\frac{2}{\theta}(1-\frac{x}{\theta})dx=\frac{2}{\theta}\int_0^\theta x-\frac{x^2}{\theta}dx$

$=\frac{2}{\theta}\left[\frac{x^2}{2}-\frac{x^3}{3\theta}\right]_0^\theta=\frac{\theta}{3}$

$E(X^2)=\int x^2 f(x;\theta)dx=\int_0^\theta x^2\frac{2}{\theta}(1-\frac{x}{\theta})dx=\frac{2}{\theta}\int_0^\theta x^2-\frac{x^3}{\theta}dx$

$=\frac{2}{\theta}\left[\frac{x^3}{3}-\frac{x^4}{4\theta}\right]_0^\theta=\frac{\theta^2}{6}$

$\text{var}\,X=E(X^2)-(E(X))^2=\frac{\theta^2}{6}-\frac{\theta^2}{9}=\frac{\theta^2}{18}$

2) a) $E(T_n)=E(\overline{X})=E(X)=\frac{\theta}{3}$ donc T_n est biaisé pour que $E(T_n)=\theta$, il suffit de poser $T_n=3\overline{X}$

b) D'après la loi faible des grands nombres

$\overline{X}\xrightarrow{P}E(X)=\frac{\theta}{3}$ donc $3\overline{X}\xrightarrow{P}3\times\frac{\theta}{3}=\theta$

donc T_n cv en proba vers θ

c) $V(T_n)=9\,\text{var}\,\overline{X}=9\times\frac{1}{n}\text{var}\,X=\frac{9}{n}\times\frac{\theta^2}{18}=\frac{\theta^2}{2n}$

T_n est sans biais donc $E(T_n-\theta)^2=\text{var}\,T_n=\frac{\theta^2}{2n}\xrightarrow[n\to+\infty]{}0$

donc T_n cv en m.q vers θ

d) On sait d'après le TCL que pour n grand

$\overline{X}\xrightarrow{\mathcal{L}}\mathcal{N}(E(X),\frac{\text{var}\,X}{n})=\mathcal{N}(\frac{\theta}{3},\frac{\theta^2}{18n})$

donc $3\overline{X}\xrightarrow{\mathcal{L}}\mathcal{N}(\frac{\theta}{3}\times 3,\frac{\theta^2}{18n}\times 9)=\mathcal{N}(\theta,\frac{\theta^2}{2n})$

EXERCICE 10

Soit $X_1,\dots,X_n$ un échantillon de loi normale d'espérance θ et de variance $\theta(1-\theta)$ où $\theta\in]0,1[$ est un paramètre inconnu. On considère les estimateurs

$$T_1=\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \quad \text{et} \quad T_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2\ .$$

1) Montrer que les estimateurs T_1 et T_2 sont sans biais et convergents.
2) Quel estimateur a-t-on intérêt à choisir ?
3) Soit maintenant l'estimateur

$$T_3=\alpha T_1+(1-\alpha)T_2$$

où $\alpha\in]0,1[$.

(a) Déterminer le biais de cet estimateur
(b) Déterminer α pour que T_3 soit de variance minimale.
(c) Lequel des trois estimateurs est le meilleur ?

N.B. On note $V(X_i^2)=2\theta^2(1-\theta^2)$

Corrigé

1) $E(T_1)=E(\overline{X})=E(X)=\theta$ donc T_1 est sans biais

$E(T_2)=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i^2)=E(X^2)=\operatorname{var}X+\left(E(X)\right)^2=\theta(1-\theta)+\theta^2=\theta$

donc T_2 est sans biais

D'après la loi des grands nombres

$T_1=\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{P} E(X_i)=\theta$

$T_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2 \xrightarrow{P} E(X_i^2)=\theta$

2) $\operatorname{var}T_1=\operatorname{var}\overline{X}=\frac{\operatorname{var}X}{n}=\frac{\theta(1-\theta)}{n}=R_\theta(T_1)$ car estimateurs sans biais

$\operatorname{var}T_2=2\theta^2(1-\theta^2)=R_\theta(T_1)$

$=\frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}(X_i^2)$ car X_i^2 indpt $=\frac{1}{n}\operatorname{var}(X_i^2)$ ← donné à la fin de l'énoncé

$\frac{\operatorname{var}T_2}{\operatorname{var}T_1}=\frac{2\theta^2(1-\theta^2)}{\theta(1-\theta)}=2\theta(1+\theta)>1 \iff 2\theta^2+2\theta-1>0$

$\Delta=12 \qquad \theta_1=\frac{-2+\sqrt{12}}{4}\simeq 0{,}37 \qquad \theta_2=\frac{2+\sqrt{12}}{4}\simeq 1{,}37$

Donc si $\theta\in[\theta_1,\theta_2]$, on a $2\theta^2+2\theta-1<0$ donc $\operatorname{var}T_2<\operatorname{var}T_1$

i.e T_2 est meilleur que T_1, sinon T_1 est meilleur que T_2.

3) a) $E(T_3)=\alpha E(T_1)+(1-\alpha)E(T_2)=\alpha\theta+(1-\alpha)\theta=\theta$

donc T_3 est sans biais

b) $\operatorname{var}T_3=\alpha^2\operatorname{var}T_1+(1-\alpha)^2\operatorname{var}T_2+\alpha(1-\alpha)\operatorname{cov}(T_1,T_2)$

l'échantillon est gaussien, dont on sait que $\overline{X}$ et S^2 sont indépendantes

or $\begin{cases} S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=T_1-T_2 \\ \overline{X}=T_1 \end{cases}$ donc T_1 et T_2 indépendantes

d'où $\operatorname{var}T_3=\alpha^2\times\operatorname{var}T_1+(1-\alpha)^2\operatorname{var}T_2$

$$\rho(\alpha) = \alpha^2 \operatorname{var} T_1 + (1-2\alpha+\alpha^2)\operatorname{var} T_2$$

$$= \alpha^2 \left[\operatorname{var} T_1 + \operatorname{var} T_2\right] - 2\operatorname{var} T_2\,\alpha + \operatorname{var} T_2$$

$$\rho'(\alpha) = 2\alpha\left[\operatorname{var} T_1 + \operatorname{var} T_2\right] - 2\operatorname{var} T_2 = 0$$

$$\Rightarrow \bar{\alpha} = \frac{\operatorname{var} T_2}{\operatorname{var} T_1 + \operatorname{var} T_2} \in [0,1]$$

$$\rho(0) = -2\operatorname{var} T_2 < 0$$

α où	0		$\bar{\alpha}$		1
$\rho'(\alpha)$		$-$		$+$	
ρ		$\searrow$		$\nearrow$	

donc $\bar{\alpha}$ est un minimum pour ρ, i.e il minimise la variance de T_3

c) • $$\operatorname{var} T_3 = \rho(\bar{\alpha}) = \left(\frac{\operatorname{var} T_2}{\operatorname{var} T_1 + \operatorname{var} T_2}\right)^2 \operatorname{var} T_1 + \left(1 - \frac{\operatorname{var} T_2}{\operatorname{var} T_1 + \operatorname{var} T_2}\right)^2 \operatorname{var} T_2$$

$$= \frac{\operatorname{var}^2 T_2 \operatorname{var} T_1 + \operatorname{var}^2 T_1 \operatorname{var} T_2}{(\operatorname{var} T_1 + \operatorname{var} T_2)^2} = \frac{\operatorname{var} T_2 \operatorname{var} T_1}{\operatorname{var} T_1 + \operatorname{var} T_2}$$

$$\frac{\operatorname{var} T_3}{\operatorname{var} T_1} = \frac{\operatorname{var} T_2}{\operatorname{var} T_1 + \operatorname{var} T_2} < 1 \quad \text{donc } \operatorname{var} T_3 < \operatorname{var} T_1$$

$$\frac{\operatorname{var} T_3}{\operatorname{var} T_2} = \frac{\operatorname{var} T_1}{\operatorname{var} T_1 + \operatorname{var} T_2} < 1 \quad \text{donc } \operatorname{var} T_3 < \operatorname{var} T_2$$

Donc T_3 est un meilleur estimateur que T_2 et T_1