Résumé du cours Estimation par intervalles de confiance

Objet:

On cherche des intervalles de confiance pour $\theta(\theta$ inconnu) à un niveau des signification 1-α avec α fixé. pour le paramètre d'intêret θ qui peut être soit une moyenne μ une variance σ^2 ou une proportionπ notée f une fréquence.

Le standard est $\alpha = 5\%$. souvent des intervalle symétriques.(-T,T)

chercher un intervalle de confiance à un niveau 1-α noté $IC_{1-\alpha}$ c'est chercher (t_1, t_2)tel que $P(\theta \in (T_1, T_2)) = 1 - \alpha$ avec $T_1 = t_1$ et $T_2 = t_2$

Dans le cas symétrique chercher $P(\theta \in (-T, T)) = 1 - \alpha$

A) Cas Gaussien ou Population infinie de loi quelconque Echantillon exhaustif ou l'échantillon non exhaustif

On applique les formules du cours avec les notation

n: Taille de l'échantillon tiré de la population mére

$z_{\alpha/2}$ est le quantile d'ordre$1 - \alpha/2$ noté $q_{1-\alpha/2}$ de la loi Normale N(0,1) comme valeur numérique

$q_{1-\alpha/2}$ est telle que $\Phi(q_{1-\alpha/2}) = 1 - \alpha/2$

Ainsi pour la loi Normale N(0,1) $\alpha = 5\%$ on a: $\Phi(q_{1-\alpha/2}) = 0.9750$ d'où $q_{1-\alpha/2} = 1.96$ le plus souvent on le rapproché à 2

Remarquer une propriété sur les quantiles:

par symétrie de la loi Normale on a

$$q_{1-\alpha/2} = -q_{\alpha/2}$$

car $\Phi(-q_{\alpha/2}) = 1 - \Phi(q_{\alpha/2})$

Estimation de la moyenne μ dans le cas Gaussien $N(\mu, \sigma^2)$ avec σ connu

Rappel:

On sait que $\bar{X}_n$ est le meilleur estimateur de μ (il est sans biais, convergent et efficace)

Sans biais: On a le Biais= $E(\bar{X}_n) - \mu = 0$

Convergent: $\bar{X}_n \to \mu$ convergence en proba $\forall\epsilon > 0$ $p(|\bar{X}_n - \mu| < \epsilon) \to 1$

efficace: Le risque $R(\bar{X}_n) \to 0$ avec $R(\bar{X}_n) = E([(\bar{X}_n - \mu)^2])$c'est EQM ou MSE. si

X_i l'échantillon suivent la loi $N(\mu, \sigma^2)$ alors $\bar{X}_n$ **suit la loi** $N(\mu, \sigma^2/n)$ce qui implique que $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ sui

$-z_{\alpha/2} \le \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le z_{\alpha/2} \Leftrightarrow \mu - z_{\alpha/2}\sigma/\sqrt{n} \le \bar{X}_n \le \mu + z_{\alpha/2}\sigma/\sqrt{n} \Leftrightarrow \bar{X}_n - z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \bar{X}_n + z_{\alpha/2}\sigma/\sqrt{n}$

Alors:

L'intervalle de Probabilité pour μ

$(\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \ \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$

d'où L'intervalle de confiance pour μ :

(Estimation) (aprés l'observation, c'est des mesures)

Ecriture formelle:

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Ecriture correcte

$$IC_{1-\alpha} = (\bar{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \ \bar{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

Exemple:

Appliquation Numérique:

pour $n = 121$ et $\sigma = 0.5$ et $\alpha = 5\%$ avec $\bar{x} = 4$

$IC = (3.9091, 4.0909)$

Ainsi la moyenne de la population μ se trouve avec une probabilité au moins égale à 0.95 dans cet intervalle.

Exercice:

Soit un 7-échantillons(7 mesures exprimés en grammes) de loi $N(\mu, \sigma^2)$ avec var=$\sigma^2 = 49$

121, 113, 135, 121, 126, 115, 123.

(i) Calculer $\bar{X}_n$

(ii) Donner l'expression de $IC_{1-\alpha}$

(iii) Appliquer avec un niveau de confiance 1-α = 98% et un niveau 95%

Solution

(i) $\bar{X}_n = \bar{x}_n = 122$

(ii) $IC_{1-\alpha} = (\bar{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \ \bar{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$

(iiiOn a 1-$\alpha = 98\% \Rightarrow \alpha = 2\% \Rightarrow 1 - \alpha/2 = 99\% = 0.99$

comme $z_{\alpha/2}$ est le quantile d'ordre$1 - \alpha/2$

La table donne: $q_{0.99} = 2.33$

$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 2.33\frac{7}{\sqrt{7}} = 6.1646$

$IC_{1-\alpha} = (122 - 6.1646, 122 + 6.1646) = (115.84, \quad 128.16)$

Pour $1 - \alpha = 95\% \Rightarrow \alpha = 5\% \Rightarrow 1 - \alpha/2 = 97.5\% = 0.975$

La table donne: $q_{0.975} = 1.96$

$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.96\frac{7}{\sqrt{7}} = 5.1857$

$IC_{1-\alpha} = (122 - 5.1857, 122 + 5.1857) = (116.81, \quad 127.19)$

2) Estimation de la moyenne μ dans le cas Gaussien $N(\mu, \sigma^2)$ avec σ inconnu

Soit T la statistique écrite de deux maniéres

$$T_n = \frac{\bar{X}_n - \mu}{S_n}.\sqrt{n-1} = \frac{\bar{X}_n - \mu}{\acute{S}_n}.\sqrt{n}$$

Remarque:

La deuxième écriture est la plus utilisée car elle fait intervenir

$\acute{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ variance corrigée c'est un estimateur sans biais de σ^2

par contre $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ (variance de l'échantillon) est un estimateur biaisé de σ^2

T suit une **loi de Student à n-1 degrés de liberté**

L'intervalle de **Probabilité:** $-t_{\alpha/2} < \frac{\bar{X}_n - \mu}{S_n}.\sqrt{n-1} < t_{\alpha/2}$ ou $-t_{\alpha/2} < \frac{\bar{X}_n - \mu}{\acute{S}_n}.\sqrt{n} < t_{\alpha/2}$

d'où l'intervalle de **confiance** pour μ

$$\bar{x}_n \pm t_{\alpha/2}\frac{s_n}{\sqrt{n-1}}$$

avec $s^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$

où

$$\bar{x}_n \pm t_{\alpha/2}\frac{\acute{s}_n}{\sqrt{n}}$$

avec $s_n'^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$

comme la loi utilisé est la loi de Student elle est symétrique, on prend des intervalles symétrique

Exemple: Donner un IC à 98% de μ avec variance inconnue pour l'échantillon suivant de taille n=7 il sagit de 7 poids d'une certaine piéce mécanique exprimés en grammes .

121, 113, 135, 121, 126, 115, 123. l'échantillon est de loi $N(\mu, \sigma^2)$ de variance inconnue

On ulilise la statistique $\frac{\bar{X}_n - \mu}{\acute{S}_n}.\sqrt{n}$ suivant la loi de Student de degré de liberté$T_{n-1} = T_6$

$IC_{1-\alpha}$ avec 1-$\alpha = 98\% \Rightarrow \alpha = 2\%$

1-$\alpha/2 = 99\%$ La table de Student nous indique $t_{6,0.99} = 3.143$

TC=$(\bar{x}_n - t_{\alpha/2}\frac{\acute{s}_n}{\sqrt{n}},\ \bar{x}_n + t_{\alpha/2}\frac{\acute{s}_n}{\sqrt{n}}) = (113, 131)$

$(\bar{x}_n = 122$ et $s_n'^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2 = 7.28$

Remarque: Le Théorème centrale limite nous permet de généraliser les formules pour n'importe quelle loi quand n est grand.

Par exemple pour la **loi Binomiale** on a le théorème de Demoivre qui donne les condition de Normalité

$$n \geq 30, \quad np \geq 5 \quad \text{et} \quad n(1-p) \geq 5$$

3) Estimation de la variance σ^2 d'une loi normale avec μ connu

Soit la statistique

$$T_n = \frac{1}{n}\sum_{i=1}^{n}(X_{i-}\mu)^2$$

$\mathbf{T}_n$ est le meilleur estimateur de σ^2

et

$$\frac{nT_n}{\sigma^2}$$

comme v.a.r suit **la loi du khi-deux à $v = n$ degré de liberté** comme somme de n carrés de N(0,1) indépendantes

Soient k_1, k_2 les bornes de l'intervalle de probabilité d'un khi-deux à n degré de liberté

p$(k_1 < \frac{nT}{\sigma^2} < k_2) = 1 - \alpha$

Remarquer k_2 est la borne supérieure

Remarque on a : $k_1 < \frac{nT}{\sigma^2} < k_2 \Leftrightarrow \frac{1}{k_2} < \frac{\sigma^2}{nT} < \frac{1}{k_1}$

d'où l'intervalle de confiance est

$$\frac{nt}{k_2} < \sigma^2 < \frac{nt}{k_1}$$

$IC_{1-\alpha}$=$(\frac{nt}{k_2}, \frac{nt}{k_1})$

Avec

Remarque

La loi du khi est de d°n .On a gagne un degré par rapport à l'autre cas où μ est inconnu (d° $= n - 1$)

Exemple: Le même

121, 113, 135, 121, 126, 115, 123. l'échantillon est de loi $\mathrm{N}(\mu, \sigma^2)$ avec μ connu

IC=( )

Calcul:

On a $\bar{x}_n = 128$ avec $\mu =$ connu on a $T_n = \frac{1}{n}\sum_{i=1}^{n}(X_{i-}\mu)^2 = t$

n=7 On applique la formule $IC_{1-\alpha} = (\frac{nt}{k_2}, \frac{nt}{k_1})$

$k_2 = q_{1-\alpha/2} = q_{0.9750}$ du khi deux à 7 d° =

$k_1 = q_{\alpha/2} = q_{0.025}$ du khi deux à 7 d° =

4) Estimation de la variance σ^2 d'une loi normale avec μ inconnu

La statistique

$$\frac{nS_n^2}{\sigma^2}$$

ou une autre écriture $\frac{(n-1)S_n'^2}{\sigma^2}$ suit la loi du Khi deux à (n-1) degrés de liberté

car $nS^2 = \sum (X_i - \bar{X})^2 = (n-1)' S_n'^2$

Soient a,b les bornes de l'intervalle de probabilité d'un khi-deux à n-1 degré de liberté (la loi n'est pas symétrique)

Remarquer a est la borne supérieur

$$p(b < \frac{nS^2}{\sigma^2} < a) = 1 - \alpha$$

On a alors l'intervalle

$$\frac{ns^2}{a} < \sigma^2 < \frac{ns^2}{b}$$

Remarque on a : $b < \frac{nS^2}{\sigma^2} < a \Leftrightarrow \frac{1}{b} > \frac{\sigma^2}{nS^2} > \frac{1}{a}$

$$IC_{1-\alpha} = (\frac{ns_n^2}{a}, \frac{ns_n^2}{b}) = (\frac{(n-1)\acute{s}_n^2}{a}, \frac{(n-1)\acute{s}_n^2}{b})$$

avec

$$a = q_{1-\alpha/2} \text{ et } b = q_{\alpha/2}$$

Voir graphe (trés important)

Exemple: Trouver un IC à 95% de σ^2 ($\alpha = 5\%$)

Avec les mêmes données 121, 113, 135, 121, 126, 115, 123

IC=(4.69, 16.03)

Justification:

On a $\bar{x}_n = 122$ et $s_n'^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2 = 7.28$

n=7 On applique la deuxième écriture $IC_{1-\alpha} = (\frac{(n-1)s_n^2}{a}, \frac{(n-1)s_n^2}{b})$

a=$q_{1-\alpha/2}$ =$q_{0.9750}$ du khi deux à 6=7-1 d° = 14.449

$b = q_{\alpha/2}$ =$q_{0.025}$ du khi deux à 6 =7-1d° = 1.23

Application:

$n = 30, s^2 = 12, 1-\alpha = 0.90$

$8.46 < \sigma^2 < 20.33 \Rightarrow 2.91 < \sigma < 4.51$

Ces formules sont valables uniquement pour le cas Gaussien

La loi du khi- deux n'est pas symétrique.

B) Cas d'une population finie de taille $N_{\mathcal{P}}$de loi quelconque

Estimation d'une proportion p

tirage avec remise:

La fréquence est un bon estimateur de p

$\mu(F) = p$

var(F)=$\frac{p(1-p)}{n}$

tirage sans remise (facteur d'exhaustivité) facteur decorrection $\frac{N_{\mathcal{P}} - n}{N_{\mathcal{P}} - 1}$

$\mu(F) = p$

var(F)=$\frac{p(1-p)}{n}\frac{N_{\mathcal{P}} - n}{N_{\mathcal{P}} - 1}$

Estimation de la moyenne et de l'écartd'une population finie

On distingue deux cas

Echantillon non exhaustif (Tirage avec remise)

$\bar{x}$ est un estimateur sans biais de la moyenne et $\text{var}(\bar{x}) = \frac{\sigma^2}{n}$

$\frac{ns^2}{n-1} = \frac{1}{n-1}\sum(X_{i-}\mu)^2$ est un estimateur sans biais de σ^2

Echantillon exhaustif (Tirage sans remise)(loi hypergéométrique

$\bar{X}$ est un bon estimateur il est sans biais de la moyenne μ et il est convergent

et $\text{var}_c(\bar{X}\ ) = \frac{\sigma^2}{n}\frac{N_{\mathcal{P}} - n}{N_{\mathcal{P}} - 1}$ $\quad$ var_c variance corrigée

$\frac{N-1}{N}\frac{n}{n-1}S^2$ est un estimateur sans biais de σ^2

avec $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_{i-\mu})^2$

Intervalle de confiance pour le paramétre d'une loi Binomilale quand n est grand($n \geq 30$ et $np \geq 5$ $n(1-p) \geq 5$)

Résultat du cours:

on a nF suit la loi $N(\mu = np, \sigma^2 = np(1-p))$

donc F suit la loi de $N(\mu = p, \sigma^2 = \sqrt{\frac{p(1-p)}{n}})$

L'intervalle de probabilité symétrique est

$$p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < F < p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

où $z_{\alpha/2} = q_{1-\alpha/2}$ quantile d'ordre 1-$\alpha/2$

cas **p** (proportion de la population) **inconnue**

f fréquence de l'échantillon connue

Intervalle de **confiance**

$$f - z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} < p < f + z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}$$

Remarque:

Pour l'Intervalle de fluctuation (p inconnu) On fait un test $p=p_0$

pour p **connu** on désigne par intervalle de **fluctuation**

pour p **inconnu** on désigne par intervalle de **confiance**