Master 1 Informatique
Éléments de statistique inférentielle

Faicel Chamroukhi
Maître de Conférences
UTLN, LSIS UMR CNRS 7296

email: chamroukhi@univ-tln.fr
web: chamroukhi.univ-tln.fr

2014/2015

Plan I

1. Estimation par intervalle

Estimation par intervalle

- Estimation par intervalle
- Intervalle de confiance
- Calcul d'un intervalle de confiance
- Loi normale : Intervalle de confiance sur μ
- Loi normale : Intervalle de confiance sur σ^2

Estimation par intervalle I

Nous allons maintenant voir une autre approche d'estimation des paramètres **l'estimation par intervalle**.

Lorsqu'on est intéressé non seulement par l'estimation en elle-même mais aussi par le niveau de confiance et la marge d'erreur, on effectue une estimation par intervalle

L'estimation par intervalle fournit, à partir d'un échantillon d'une population, non seulement des valeurs des paramètres à estimer, mais un intervalle de valeurs centré sur la valeur numérique estimée du paramètre inconnu avec un *niveau de confiance* donné.

Ce niveau de confiance représente la probabilité que le vrai paramètre se trouve dans l'intervalle que l'on donne comme estimation.

Estimation par intervalle II

L'intervalle est appelé **intervalle de confiance** (IC) le niveau de confiance est aussi appelé **précision** ou **cœfficient de confiance**.

Intervalle de confiance I

Définition : **Intervalle de confiance**

Soit $(X_1, \ldots, X_n)$ un échantillon de variables aléatoires issues d'une population de densité $f(x;\theta)$, θ étant le (vecteur) paramètre à estimer. Supposons aussi que $T_1(X_1, \ldots, X_n)$ et $T_2(X_1, \ldots, X_n)$ deux statistiques sur l'échantillon telle que $T_1 < T_2$. L'intervalle $[T_1, T_2]$ est dit *intervalle de confiance à* $100(1-\alpha)\%$ pour θ si

$$\mathbb{P}(T_1 < \theta < T_2) = 1 - \alpha \cdot \tag{1}$$

α représente le *risque* que le vrai paramètre θ ne soit pas dans cet intervalle et $1-\alpha$ s'appelle niveau ou (cœfficient) de confiance. Une estimation par intervalle de confiance sera donc d'autant meilleure que l'intervalle sera petit pour un cœfficient de confiance grand (proche de 1) ou de manière équivalente pour un risque α proche de zéro.

Les valeurs généralement prises pour $1-\alpha$ sont 0.90, 0.95, 0.99, and 0.999.

Intervalle de confiance II

Les limites de l'intervalle T_1 et T_2 sont appelés respectivement la *limite inférieure de confiance* et *limite supérieure de confiance*.

⚠Remarques

- L'intervalle de confiance est fonction de l'estimation du paramètre θ
- L'intervalle de confiance est également fonction de α. A taille d'échantillon *n* fixée, lorsqu'on augmente le niveau de confiance $1-\alpha$, la largeur de l'Intervalle de Confiance (IC)
- Pour un niveau de confiance $1-\alpha$ fixé, lorsqu'on augmente la taille de l'échantillon *n*, la largeur de l'IC diminue.

Intervalle de confiance III

Soit a et b les bornes d'un intervalle de confiance $\mathrm{IC}_{1-\alpha}(\theta)$ pour le paramètre θ on a On a :

$$\mathbb{P}(a < \theta < b) = 1 - \alpha \Longrightarrow \mathbb{P}(\theta < a) + \mathbb{P}(\theta > b) = \alpha \tag{2}$$

En posant $\alpha = \alpha_1 + \alpha_2$, il existe donc une infinité de choix possibles pour α_1 et α_2, et donc de choix pour a et b et donc de l'IC. Pour l'instant, nous ne considérons que le cas d'un intervalle de confiance bilatéral symétrique, où on a les mêmes risques $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$

Notons que, connaissant la loi de l'estimateur, il est possible de donner un intervalle de confiance. Ici nous considérons les intervalles de confiance les plus classiques.

Loi normale : Intervalle de confiance pour μ avec σ connu I

On a vu que $\bar{X}$ est le meilleur estimateur de μ et que

$$U = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \quad \sim \quad \mathcal{N}(0, 1) \tag{3}$$

En prenant des risques symétriques ($\frac{\alpha}{2}$), pour un risque fixé, on peut donc lire dans les tables de probabilités de la loi normale centrée réduite, le **quantile** $\mathbb{P}(U \leq \frac{\alpha}{2})$

⚠ Remarque : Comme le risque est symétrique ici, on a donc

$$\mathbb{P}(U \geq \frac{\alpha}{2}) = \alpha - \mathbb{P}(U \leq \frac{\alpha}{2}) = \frac{\alpha}{2} \tag{4}$$

La notion de quantile est définie de la façon suivante :

Loi normale : Intervalle de confiance pour μ avec σ connu II

Definition

pour une variable aléatoire continue X, le quantile q_α d'ordre α de la loi de X est telle que

$$\mathbb{P}(U \leq q_\alpha) = \alpha \quad (5)$$

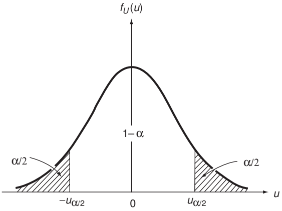

Figure : Loi normale centrée réduite et IC à $[100(1-\alpha)]\%$

Loi normale : Intervalle de confiance pour μ avec σ connu III

⚠ Remarque : la notation généralement utilisée pour les quantile est : u_α pour la loi normale, t_α pour la loi de Student à n degrés de liberté, χ^n_α pour la loi χ^2_n, etc

Le risque étant symétrique, d'après (4) on a

$$\mathbb{P}(-u_{\frac{\alpha}{2}} \leq U \leq u_{\frac{\alpha}{2}}) = 1 - \alpha \tag{6}$$

et d'après (3) on obtient

$$\mathbb{P}\left(\bar{X} - u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \tag{7}$$

d'où l'intervalle de confiance sur μ :

$$\mathrm{IC}_{1-\alpha}(\mu) = \left[\bar{X} - u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\,,\, \bar{X} + u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] \tag{8}$$

Loi normale : Intervalle de confiance pour μ avec σ connu IV

exemple pour $\alpha = 0.05$

$$\mathrm{IC}_{0.95}(\mu) = \left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}\,,\ \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right] \tag{9}$$

Loi normale : Intervalle de confiance pour μ avec σ inconnu I

Pour calculer l'IC sur μ on a vu que la statistique à utiliser est $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$. Or, comme la variance σ^2 est inconnue, on utilise a sa place son meilleur estimateur : la variance empirique corrigée $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_i)^2$

La statistique à utiliser est donc

$$T = \frac{\sqrt{n}(\bar{X}-\mu)}{S} \tag{10}$$

On sait que $Z = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

la statistique (10) pour calculer l'IC s'écrit dont

$$T = \frac{U}{\sqrt{\frac{Z}{n-1}}} \sim \text{ Loi de Student à n-1 degrés de liberté} \tag{11}$$

Loi normale : Intervalle de confiance pour μ avec σ inconnu II

Soit $t_{n-1,\frac{\alpha}{2}} = \mathbb{P}(T \leq \frac{\alpha}{2})$ le quantile d'ordre $\alpha/2$ de la loi de Student à $n-1$ degrés de liberté.

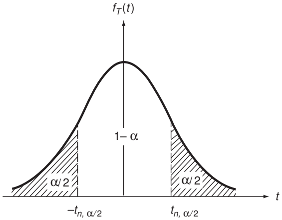

Figure : Loi de Student à $n-1$ degrés de liberté et IC à $[100(1-\alpha)]\%$

Loi normale : Intervalle de confiance pour μ avec σ inconnu III

L'intervalle de confiance est donc donné par

$$\mathbb{P}(-t_{n-1,\frac{\alpha}{2}} \leq T \leq t_{n-1,\frac{\alpha}{2}}) = 1 - \alpha \tag{12}$$

On obtient donc l'intervalle de confiance comme précédemment

$$\mathrm{IC}_{1-\alpha}(\mu) = \left[\bar{X} - t_{n-1,\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\, , \, \bar{X} + t_{n-1,\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right] \tag{13}$$

où $t_{n-1,\frac{\alpha}{2}}$ est le quantile d'ordre $\alpha/2$ de la loi de Student à $n-1$ degrés de liberté

⚠ Remarque : Si la loi de X n'est pas normale, on sait d'après le théorème central limite que lorsque la taille d'échantillon est grande, $\bar{X}$ suit une loi normale, et donc les résultats précédents sont applicables.

Loi normale : Intervalle de confiance pour σ lorsque μ connu I

On sait que la variance observée $V^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ constitue le meilleur estimateur de σ^2 lorsque μ est connue.

D'autre part :

$$D = \frac{n}{\sigma^2}V^2 \sim \chi^2_n \tag{14}$$

Soit $\chi^2_{n,\frac{\alpha}{2}} = \mathbb{P}(D \leq \frac{\alpha}{2})$ le quantile d'ordre $\frac{\alpha}{2}$ de la loi de χ^2 à n degrés de liberté.

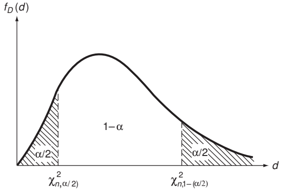

Figure : Loi de χ^2 à n degrés de liberté et IC à $[100(1-\alpha)]\%$

Loi normale : Intervalle de confiance pour σ lorsque μ connu II

l'$\text{IC}_{1-\alpha}(\sigma^2)$ est donc donné par

$$\mathbb{P}\left(\chi^2_{n,\frac{\alpha}{2}} \leq D = \frac{n}{\sigma^2}V^2 \leq \chi^2_{n,1-\frac{\alpha}{2}}\right) = 1-\alpha$$

donc

$$\mathbb{P}\left(\frac{nV^2}{\chi^2_{n,1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{nV^2}{\chi^2_{n,\frac{\alpha}{2}}}\right) = 1-\alpha$$

finalement on obtient

$$\text{IC}_{1-\alpha}(\sigma^2) = \left[\frac{nV^2}{\chi^2_{n,1-\frac{\alpha}{2}}}, \leq \frac{nV^2}{\chi^2_{n,\frac{\alpha}{2}}}\right] \tag{15}$$

Loi normale : Intervalle de confiance pour σ lorsque μ inconnu I

lorsque μ est inconnue, on sait que que la variance empirique corrigée $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ constitue le meilleur estimateur de σ^2

On sait également que :

$$D = \frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1} \tag{16}$$

l'$\mathrm{IC}_{1-\alpha}(\sigma^2)$ est donc donné par

$$\mathbb{P}\left(\chi^2_{n-1,\frac{\alpha}{2}} \leq D = \frac{n-1}{\sigma^2} S^2 \leq \chi^2_{n-1,1-\frac{\alpha}{2}}\right) = 1-\alpha$$

donc

Loi normale : Intervalle de confiance pour σ lorsque μ inconnu II

$$\mathrm{IC}_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right] \tag{17}$$

⚠ Remarque : Ces intervalles de confiance sur la variance ne sont valables que pour une loi normale. Contrairement au cas de la moyenne, ces résultats ne peuvent être étendus aux cas d'autres lois