MTH2302: Estimation

Wissem Maazoun

École Polytechnique de Montréal

Département de Mathématiques et de génie industriel

Introduction

Introduction

L'inférence statistique comprend l'estimation et les tests d'hypothèses (chap. 11). L'estimation d'un paramètre consiste à déterminer une statistique qui "reflète" autant que possible la vraie valeur (inconnue) de ce paramètre. Deux méthodes d'estimations sont possibles : l'estimation ponctuelle et l'estimation par intervalle de confiance.

Estimation ponctuelle

Définition

Soit X une v.a. dont la distribution dépend d'un paramètre θ (ex. Exp(λ), Poisson(λ), etc.), et soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire de X. Un estimateur ponctuel de θ est une statistique $\hat{\theta}$ de la forme $\hat{\theta} = h(X_1, \ldots, X_n)$, vérifiant certaines propriétés.

Exemples

- Pour estimer une moyenne $\mu = E(X)$ d'une v.a. X, on peut utiliser $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$.
- Pour estimer une variance $\sigma^2 = V(X)$ d'une v.a. X, on peut utiliser la variance de l'échantillon $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$.

Propriétés d'un estimateur

Critère 1 : Le biais

Le biais d'un estimateur $\hat{\theta}$ d'un paramètre θ est défini par

$$Biais(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

On dit que l'estimateur $\hat{\theta}$ est sans biais (ou non biaisé) pour θ si $Biais(\hat{\theta}) = 0$, c-à-d si $E(\hat{\theta}) = \theta$.

Exemples

- $E(\bar{X}) = \mu$, donc $\bar{X}$ est un estimateur sans biais de μ.
- $E(\hat{p}) = p$, donc $\hat{p}$ est un estimateur sans biais de p la probabilité de succès.
- $E(S^2) = \sigma^2$, donc S^2 est un estimateur sans biais de σ^2.

Propriétés d'un estimateur

Critère 2 : L'erreur quadratique moyenne (EQM)

On mesure la précision globale d'un estimateur par son erreur quadratique moyenne. Soit $\hat{\theta}$ un estimateur d'un paramètre θ. On définit l'erreur quadratique moyenne de $\hat{\theta}$ par

$$EQM(\hat{\theta}) = E(\hat{\theta} - \theta)^2.$$

On a toujours $EQM(\hat{\theta}) = V(\hat{\theta}) + [Biais(\hat{\theta})]^2$.
Lorsque $\hat{\theta}$ est sans biais pour θ, on a
$Biais(\hat{\theta}) = 0$ et $EQM(\hat{\theta}) = V(\hat{\theta})$.
Le critère de l'erreur quadratique moyenne permet de comparer les estimateurs d'un même paramètre. Ainsi , si $\hat{\theta}_1$ et $\hat{\theta}_2$ sont deux estimateurs d'un paramètre θ, le plus efficace des deux est celui ayant la plus petite erreur quadratique moyenne.
$\hat{\theta}_1$ est plus efficace que $\hat{\theta}_2$ si $EQM(\hat{\theta}_1) < EQM(\hat{\theta}_2)$.

Propriétés d'un estimateur

Exemple

Soit $X_1, X_2, \ldots, X_5$ un échantillon aléatoire d'une v.a. X telle que $E(X) = \mu$ et $V(X) = \sigma^2$. Pour estimer μ, on considère

$$\hat{\mu}_1 = \frac{X_1 + \ldots + X_5}{5} \text{ et } \hat{\mu}_2 = \frac{2X_1 - X_2 + X_4}{2}.$$

1. $\hat{\mu}_1$ et $\hat{\mu}_2$ sont-ils sans biais ?
2. Quel est le meilleur des deux ? Justifiez.

Propriétés d'un estimateur

Critère 3 : La convergence

Un estimateur $\hat{\theta}_n$ de θ, calculé sur un échantillon de taille n est dit convergent si lorsque $n \to \infty$, la probabilité que $\hat{\theta}_n$ soit proche de θ devient grande.

$$\forall \epsilon > 0, \lim_{n \to \infty} P(|\hat{\theta}_n - \theta| < \epsilon) = 1.$$

Une condition suffisante pour la convergence de $\hat{\theta}_n$ est que son erreur quadratique moyenne tende vers zéro lorsque $n \to \infty$.

Exemple

Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire d'une v.a. X telle que $E(X) = \mu$ et $V(X) = \sigma^2$. Pour estimer μ, on considère $\hat{\mu} = \bar{X}$. On a $E(\bar{X}) = \mu$ et $V(\bar{X}) = \frac{\sigma^2}{n} \to 0$ lorsque $n \to \infty$. $\bar{X}$ est donc un estimateur convergent de μ.

Estimation ponctuelle

Méthode des moments

Nous remarquons que la plupart des lois vues aux chapitres 5, 6 et 7, sont entièrement définies par un ou deux paramètres. Ces paramètres sont généalement liés aux deux premiers moments de la v.a. On peut citer par exemple :

Loi	Paramètres
Bernouilli	1^{er} moment
Géométrique	1^{er} moment
Poisson	1^{er} moment
Normale	Les deux premiers moments
Lognormale	Les deux premiers moments du logarithme
Exponentielle	Inverse du 1^{er} moment
Beta	2 paramètres fonctions des 2 premiers moments

Méthode des moments

Utilisation

Soit X est une v.a. continue ayant comme fonction de densité $f(x, \theta_1, \theta_2)$, ou une v.a discrète dont la fonction de masse est $p(x, \theta_1, \theta_2)$ et que cette v.a est définie à partir de ces deux premiers moments qui nous sont inconnus. Si $X_1, X_2, \ldots, X_n$ est un échantillon aléatoire de taille n des valeurs de la v.a. X, on peut alors définir les deux premiers moments de cet échantillon par rapport à l'origine par :

$$m_i' = \frac{1}{n}\sum_{j=1}^{n} X_j^i, \qquad i = 1, 2.$$

Ainsi, on peut estimer θ_i par résolution du système $E[X^i] = m_i', i = 1, 2.$

Wissem Maazoun MTH2302: Estimation 9 / 31

Estimation ponctuelle

Méthode du maximum de la vraisemblance

La méthode de maximum de la vraisemblance est l'une des meilleures méthodes d'estimation ponctuelle. Soit X une v.a. ayant comme fonction de probabilité $f(x, \theta)$ où θ représente un paramètre inconnu. Si $x_1, x_2, \ldots, x_n$ est un échantillon de taille n des valeurs de la v.a. X, la fonction de vraisemblance de cet échantillon est alors :

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta).$$

$\hat{\theta}$ l'estimateur du maximum de la vraisemblance de θ consiste à la valeur de θ telle que $L(\theta)$ est maximale. Il s'agit essentiellement de la valeur de θ qui maximise la probabilité d'obtenir les résultats fournis par l'échantillon.

Méthode du maximum de la vraisemblance

Exemples

1. Déterminer l'estimateur du paramètre de la loi de Poisson.
2. Déterminer l'estimateur du paramètre de la loi exponentielles.
3. Déterminer l'estimateur des paramètres de la loi normale.

Estimation par intervalle de confiance

Généralités

La méthode d'estimation d'un paramètre θ par intervalle de confiance consiste à déterminer deux statistiques $L(X_1, \ldots, X_n)$ et $U(X_1, \ldots, X_n)$ vérifiant l'égalité $P(L \leq \theta \leq U) = 1 - \alpha$. On l'appelle le niveau (ou le coefficient) de confiance. On le choisit généralement proche de 1. Généralement les bornes L et U s'obtiennent en posant $P(\theta < L) = \alpha/2$ et $P(\theta > U) = \alpha/2$.
On obtient ainsi des intervalles de petite longueur (et généralement symétrique).
La méthode générale pour obtenir L et U consiste à considérer un estimateur $\hat{\theta}$ de θ et une fonction $H(\hat{\theta}, \theta)$ dont la distribution ne dépend d'aucun paramètre inconnu (fonction pivotale). On peut alors trouver deux nombres h_1 et h_2 vérifiant
$P(h_1 \leq H(\hat{\theta}, \theta) \leq h_2) = 1 - \alpha$.
La forme exacte d'un intervalle de confiance dépend du paramètre à estimer et de la distribution de la variable considérée.

Estimation par intervalle de confiance

Intervalle de confiance pour une moyenne

Soit X une v.a de moyenne μ et de variance σ^2 et soit $X_1, \ldots, X_n$ un échantillon aléatoire de taille n de X. Pour un niveau de confiance $1-\alpha$ donné, les intervalles de confiance pour μ sont résumés dans le tableau suivant différentes situations (tableau I.C. P262).

Une moyenne μ		
σ^2 est connue et $X \sim N(\mu, \sigma^2)$, ou n grand	$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,\ 1)$	$\mu \in \overline{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
σ^2 est inconnue et $X \sim N(\mu, \sigma^2)$	$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim T_{n-1}$	$\mu \in \overline{X} \pm t_{\alpha/2;n-1} \cdot \frac{S}{\sqrt{n}}$
σ^2 est inconnue et n est très grand	$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim N(0,\ 1)$	$\mu \in \overline{X} \pm z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$

Estimation par intervalle de confiance

Remarques

- Les intervalles donnés dans le tableau sont dits bilatéraux et symétriques. Pour un niveau $1-\alpha$ fixé, les intervalles du tableau sont les plus courts. De plus, lorsque n augmente, la longueur de l'intervalle diminue et la précision augmente.
- Les intervalles de confiance unilatéraux de niveau $1-\alpha$ se déduisent de ceux du tableau en considérant une seule borne de l'intervalle et en remplaçant $\alpha/2$ par α.
- Taille d'échantillon nécessaire : L'erreur d'estimation dans un intervalle de confiance pour μ est de $|\bar{X}-\mu|$. Lorsque la variance σ^2 est connue, on peut déterminer une taille d'échantilon nécessaire garantissant une erreur d'estimation d'au plus E, pour un niveau de confiance $1-\alpha$ donné. Lorsque E est fixé, on doit avoir $|\bar{X}-\mu| \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq E \Rightarrow n \geq \left(z_{\alpha/2}\frac{\sigma}{E}\right)^2$.

Estimation par intervalle de confiance

Remarques (suite)

- Dans le cas où la variance σ^2 est inconnue, on remplace σ par l'écart type s de l'échantillon. Cette procédure est valide si X suit une loi normale (on a alors une loi T de Student à $\nu = n-1$ d.d.l), ou si n est grand on a alors une loi normale.

Exemple

Le poids X des contenants d'un certain produit est une v.a de moyenne μ et d'écart type $0,3$ grammes. Un échantillon aléatoire de 100 contenants est choisi et donne une moyenne de $49,7$ grammes. Construire un I.C. pour μ de niveau

1. $1-\alpha = 0,95$
2. $1-\alpha = 0,99$
3. Quelle taille d'échantillon doit-on prélever pour que l'erreur d'estimation soit d'au plus 0,04 au niveau de confiance $1-\alpha = 0,90$.

Wissem Maazoun MTH2302: Estimation 15 / 31

Estimation par intervalle de confiance

Exemple

Un certain procédé est utilisé pour la production en série de pièces dont le poids X peut être approché par une distribution normale $N(\mu, \sigma^2)$. Un échantillon de 5 pièces prises au hasard dans la production du procédé est examiné et les poids suivants sont observés :

$$45,7 \quad 46,0 \quad 46,8 \quad 45,2 \quad 46,3 \quad \text{(en grammes)}$$

1. Donner une estimation ponctuelle de μ.
2. Donner un I.C. pour μ au niveau de confiance 90%.
3. Donner un I.C. unilatéral (avec borne supérieure) pour μ au niveau de confiance 90%.

Estimation par intervalle de confiance

Intervalle de confiance pour une variance σ^2

Soit X une v.a de loi $N(\mu, \sigma^2)$, avec σ^2 inconnue, et soit $X_1, \ldots, X_n$ un échantillon aléatoire de X. Pour un niveau de confiance $1 - \alpha$ fixé, l'I.C bilatéral de σ^2 est

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2;n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2n-1}}$$

.

Une variance σ^2		
$X \sim N(\mu, \sigma^2)$ et μ , σ^2 sont inconnues	$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$	$\frac{(n-1)S^2}{\chi^2_{\alpha/2;n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2;n-1}}$
Cas approximatif n est grand ($n \geq 40$)	$\frac{S-\sigma}{\sigma/\sqrt{2n}} \sim N(0,\ 1)$	$\frac{S}{1+\frac{z_{\alpha/2}}{\sqrt{2n}}} \leq \sigma \leq \frac{S}{1-\frac{z_{\alpha/2}}{\sqrt{2n}}}$

Estimation par intervalle de confiance

Remarques

Les I.C unilatéraux de niveau $1 - \alpha$ pour σ^2 sont

- $\frac{(n-1)S^2}{\chi^2_{\alpha;n-1}} \leq \sigma^2$
- $\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha;n-1}}$.
- Lorsque X n'est pas distribuée selon une loi normale, on peut utiliser l'approximation suivante si la taille d'échantillon est grande $(n \geq 40)$: $S \sim N(\sigma, \frac{\sigma^2}{2n})$.
  On en déduit que pour un niveau de confiance $1 - \alpha$ fixé, un intervalle de confiance pour σ est

$$\frac{S}{1+\frac{z_{\alpha/2}}{\sqrt{2n}}} \leq \sigma \leq \frac{S}{1-\frac{z_{\alpha/2}}{\sqrt{2n}}}.$$

Wissem Maazoun MTH2302: Estimation 18 / 31

Estimation par intervalle de confiance

Intervalle de confiance pour une proportion p

Soit p la proportion de "succès" dans une population et soit $\hat{p}$ la proportion de "succès" dans un échantillon aléatoire de taille n.
Pour un niveau $1-\alpha$ fixé, un I.C (approximatif) pour p lorsque n est grand est :

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Une proportion p		
$X \sim$ Bernoulli(p) et n est très grand	$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,\ 1)$	$p \in \hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Estimation par intervalle de confiance

Remarques

1. Les I.C unilatéraux de niveau $1-\alpha$ pour p se déduisent de celui du tableau en considérant une seule borne de l'intervalle et en remplaçant $\alpha/2$ par α.
   - $\hat{p} - z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p$
   - $p \le \hat{p} + z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
2. L'erreur d'estimation dans un I.C pour p est de $|\hat{p}-p|$. On peut déterminer une taille d'échantillon nécessaire garantissant une erreur d'estimation d'au plus E, pour un niveau de confiance $1-\alpha$ donné. On doit avoir
   $|\hat{p}-p| \le z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le E \Longrightarrow n \ge \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p)$
   1. On dispose d'une estimation antérieure $\hat{p}_0$ de p, on prend $n \ge \left(\frac{z_{\alpha/2}}{E}\right)^2 \hat{p}_0(1-\hat{p}_0)$.
   2. On ne dispose d'aucune information sur p, on prend $p = 0,5$, et $n \ge \left(\frac{z_{\alpha/2}}{E}\right)^2 \frac{1}{4}$.

Estimation par intervalle de confiance (deux échantillons)

I.C pour une différence de deux moyennes $\mu_1 - \mu_2$

Soit X_1 et X_2 deux v.a (populations) de moyennes μ_1 et μ_2 et de variances σ_1^2 et σ_2^2 respectivement ; et soient $X_{11}, \dots, X_{1n_1}$ et $X_{21}, \dots, X_{2n_2}$ deux échantillons indépendants provenants de X_1 et X_2 respectivement.
Notons $\bar{X}_1$ et S_1^2 la moyenne et la variance de l'échantillon 1, et $\bar{X}_2$ et S_2^2 la moyenne et la variance de l'échantillon 2.
Pour un niveau de confiance $1-\alpha$ donné, les intervalles de confiance pour la différence $\mu_1 - \mu_2$ sont résumés dans le tableau suivant différentes situations.

Wissem Maazoun MTH2302: Estimation 21 / 31

Estimation par intervalle de confiance (deux échantillons)

Situation	Loi utilisée	Intervalle
Une différence $\mu_1 - \mu_2$ σ_1^2 et σ_2^2 connues et $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$ ou alors n_1 et n_2 grands	$\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$(\overline{X}_1 - \overline{X}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
σ_1^2 et σ_2^2 inconnues avec $\sigma_1^2 = \sigma_2^2 = \sigma^2$ et $X_i \sim N(\mu_i, \sigma^2)$, $i = 1, 2$	$\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T_{n_1+n_2-2}$	$(\overline{X}_1 - \overline{X}_2) \pm t_{\alpha/2;n_1+n_2-2} \cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ où $S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$
σ_1^2 et σ_2^2 inconnues avec $\sigma_1^2 \neq \sigma_2^2$ et $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$	$\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim T_\nu$	$(\overline{X}_1 - \overline{X}_2) \pm t_{\alpha/2;\nu} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ où $\nu = \frac{(S_1^2/n_1+S_2^2/n_2)^2}{\frac{(S_1^2/n_1)^2}{n_1+1}+\frac{(S_2^2/n_2)^2}{n_2+1}} - 2$
σ_1^2 et σ_2^2 inconnues et n_1, n_2 grands $(n_1 \geq 30,\ n_2 \geq 30)$	$\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1)$	$(\overline{X}_1 - \overline{X}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$
Observations couplées $D = X_1 - X_2 \sim$ Normale $D_j = X_{1j} - X_{2j}$, $j = 1, \ldots, n$	$\dfrac{\overline{D} - (\mu_1 - \mu_2)}{S_D/\sqrt{n}} \sim T_{n-1}$	$\overline{D} \pm t_{\alpha/2;n-1} \cdot \frac{S_D}{\sqrt{n}}$ où $\overline{D} = \frac{1}{n}\sum_{j=1}^{n} D_j, S_D = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(D_j - \overline{D})^2}$

Estimation par intervalle de confiance (deux échantillons)

Remarques

1. Les intervalles donnés dans le tableau sont dits bilatéraux et symétriques. Pour un niveau de confiance $1 - \alpha$ fixé, les intervalles du tableau sont plus courts. Les intervalles unilatéraux de niveau $1 - \alpha$ se déduisent de ceux du tableau en considérant une seule borne de l'intervalle et en remplaçant $\alpha/2$ par α.
2. L'erreur d'estimation dans un intervalle de confiance pour la différence $\mu_1 - \mu_2$ est de $|\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)|$. Lorsque les variances σ_1^2 et σ_2^2 sont connues, on peut déterminer une taille d'échantillon nécessaire n (la même pour les deux échantillons) garantissant une erreur d'estimation d'au plus E, pour un niveau de confiance $1 - \alpha$ donné. En effet, si $n_1 = n_2 = n$, on a $|\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)| \leq \frac{z_{\alpha/2}}{\sqrt{n}} \times \sqrt{\sigma_1^2 + \sigma_2^2} \leq E \Longrightarrow n \geq \left(\frac{z_{\alpha/2}}{E} \times \sqrt{\sigma_1^2 + \sigma_2^2}\right)^2$.

Estimation par intervalle de confiance (deux échantillons)

Exemple

Une analyse statistique descriptive des données sur des mesures de diamètre (en cm) des pièces produites par deux chaines de production (A et B) a donné les résultats suivants :

Chaîne	A	B
nombre d'observations	51	51
moyenne : $\bar{x}$	$1,375147$	$1,374982$
écart-type : s	$0,000167$	$0,000172$

Soit X_1 (X_2) le diamètre des pièces produites par la chaîne A (B). $E(X_1) = \mu_1, E(X_2) = \mu_2, V(X1) = \sigma_1^2, V(X2) = \sigma_2^2$.

1. Donner un intervalle de confiance pour $\mu_1 - \mu_2$ au niveau de confiance 95%.
2. Donner un intervalle de confiance unilatéral (avec borne inférieure) pour $\mu_1 - \mu_2$ au niveau de confiance 95%.

Wissem Maazoun MTH2302: Estimation 24 / 31

Estimation par intervalle de confiance (deux échantillons)

I.C pour un rapport de deux variances σ_1^2/σ_2^2

Soit X_1 et X_2 deux v.a telles que $X_1 \sim N(\mu_1, \sigma_1^2)$ et $X_2 \sim N(\mu_2, \sigma_2^2)$, et soit $X_{11}, \ldots, X_{1n_1}$ et $X_{21}, \ldots, X_{2n_2}$ deux échantillons indépendants provenant de X_1 et X_2 respectivement. Pour un niveau de confiance $1-\alpha$ donné, l'intervalle de confiance bilatéral pour le rapport σ_1^2/σ_2^2 est

$$\frac{S_1^2}{S_2^2} F_{1-\alpha/2;n_2-1,n_1-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} F_{\alpha/2;n_2-1,n_1-1}.$$

Estimation par intervalle de confiance (deux échantillons)

Exemple

On dispose de deux méthodes (A et B) pour la fabrication d'un produit. On s'intéresse au temps nécessaire pour fabriquer une unité du produit. Une analyse statistique descriptive a donné les résultats préliminaires suivants

Méthode	A	B
Taille d'échantillon	12	16
Écart-type s	0,85	0,98

Soit X_1 (X_2) le le temps requis par le procédé A (B).
$X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$.

1. Donner un intervalle de confiance pour le rapport $\frac{\sigma_1^2}{\sigma_2^2}$ au niveau de confiance $1-\alpha = 0,90$.
2. Donner un intervalle de confiance unilatéral (avec borne inférieure) pour le rapport $\frac{\sigma_1^2}{\sigma_2^2}$ au niveau de confiance $1-\alpha = 0,90$.

Autres problèmes d'estimation par intervalle

Les intervalles de prévision

Soit $X_1, \ldots, X_n$ un échantillon aléatoire d'une v.a X de loi normale $N(\mu, \sigma^2)$. On cherche parfois à prédire la future observation, disons X_{n+1} de X. On estime ponctuellement une telle observation par $\bar{X}$. Pour un niveau de confiance $1-\alpha$ fixé, l'intervalle de confiance pour X_{n+1} est

$$X_{n+1} \in \bar{X} \pm t_{\alpha/2;n-1}\sqrt{S^2(1+\frac{1}{n})}.$$

Les intervalles unilatéraux de niveau 1_α sont :

- $X_{n+1} \geq \bar{X} - t_{\alpha;n-1}\sqrt{S^2(1+\frac{1}{n})}$, cas unilatéral avec borne inférieure ;
- $X_{n+1} \leq \bar{X} + t_{\alpha;n-1}\sqrt{S^2(1+\frac{1}{n})}$, cas unilatéral avec borne supérieure.

Wissem Maazoun MTH2302: Estimation 27 / 31

Autres problèmes d'estimation par intervalle

Exemple 10.27

On a mesuré au cours de 10 vols l'accélération maximale (en g) d'un avion de ligne assurant une certaine liaison. Voici les résultats obtenus :

$$1,15 \quad 1,23 \quad 1,56 \quad 1,69 \quad 1,71 \quad 1,83 \quad 1,83 \quad 1,85 \quad 1,90 \quad 1,91$$

On a $\bar{X} = 1,666, s = 0,273$. Donner un intervalle de prédiction de l'accélération maximale que l'avion subira lors de son prochain vol.

Autres problèmes d'estimation par intervalle

Les intervalles de tolérance

Un intervalle de tolérance (statistique) d'une v.a X est un intervalle calculé à partir des données d'un échantillon aléatoire $X_1, \ldots, X_n$ de X, qui contient (au moins) un pourcentage q (taux de couverture) des valeurs de X, avec une grande probabilité $1 - \alpha$ (coefficient de confiance). On distingue deux méthodes selon que X suit une loi normale ou non.

Les intervalles de tolérance

Cas où X suit une loi normale

Lorsque $X \sim N(\mu, \sigma^2)$, il est démontré que l'intervalle de tolérance, avec un taux de couverture de q et un niveau de confiance $1 - \alpha$ est de la forme

$$\bar{X} \pm k \times S.$$

$\bar{X}$ représente la moyenne de l'échantillon et S l'écart type. Les valeurs de k sont données dans la table **XIII** page 502 en fonction de $q, 1 - \alpha$ et n.

Remarque

Les intervalles unilatéraux de tolérance sont de la forme $\bar{X} - k \times S$ ou $\bar{X} + k \times S$. Les valeurs de k sont alors lues dans la table **XIII** page 501 en fonction de $q, 1 - \alpha$ et n.

Wissem Maazoun MTH2302: Estimation 30 / 31

Les intervalles de tolérance

Cas où X suit une loi quelconque

Lorsque X est distribuée selon une loi quelconque mais de type continu, on considère

- $X_{min} = X_{(1)} = \min\{X_1, \dots, X_n\}$
- $X_{max} = X_{(n)} = \max\{X_1, \dots, X_n\}$

Il est alors démontré que $[X_{min}, X_{max}]$ constitue un intervalle de tolérance pour la v.a X. Le taux de couverture $q = P$ et le niveau de confiance $1-\alpha$ de cet intervalle sont tels que
$1-\alpha = 1 - nP^{n-1} + (n-1)P^n$.
On a donc deux possibilité

1. Déterminer $1-\alpha$ lorsque P et n sont connus ;
2. Déterminer n approximativement, lorsque $1-\alpha$ et P sont connus, par $n = \frac{1}{2} + \frac{1+P}{1-P}\frac{\chi_4^2(\alpha)}{4}$.