Retour sur le maximum de vraisemblance

- Soient $x_1, \dots, x_n$ des données supposées être une réalisation d'un échantillon aléatoire $X_1, \dots, X_n \overset{\text{iid}}{\sim} f(x;\theta)$ (densité / masse). La **vraisemblance** pour θ est la fonction

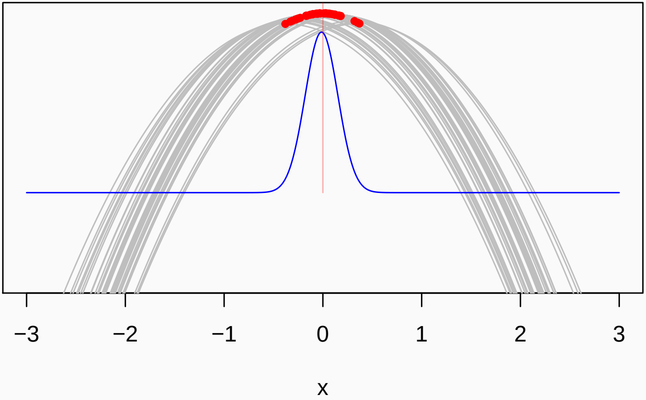
$$L(\theta) = f(x_1;\theta) \times f(x_2;\theta) \times \cdots \times f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

- $\widehat{\theta}_{\mathrm{ML}}$ satisfait

$$L(\widehat{\theta}_{\mathrm{ML}}) \geq L(\theta) \quad \text{pour tout } \theta$$

- **Interprétation** : dans le cas discret, on maximise $\mathrm{Pr}_\theta(X_1 = x_1, \dots, X_n = x_n)$
- Il est plus facile de maximiser $\ell(\theta) := \log L(\theta)$, souvent en résolvant $d\ell(\theta)/d\theta = 0$, et vérifiant qu'il s'agit bien d'un maximum

La vraisemblance est une fonction aléatoire



Fonctions de vraisemblances correspondantes à 50 échantillons de taille $n = 40$. La vraie valeur est $\theta = 0$ et les 50 maxima sont en rouge.

Distribution asymptotique/approximative de $\widehat{\theta}_{\mathrm{ML}}$

- Dans des modèles "réguliers" (normal, exponentiel, Poisson, ...) on a $\ell'(\widehat{\theta}_{\mathrm{ML}}) = 0$
- Comme $\ell(\theta)$ est une somme de variables aléatoires iid, le théorème central limite s'applique
- Grâce à la méthode delta $\widehat{\theta}_{\mathrm{ML}} \overset{\mathrm{app}}{\sim} \mathcal{N}(\theta, 1/I_n(\theta))$ avec **l'information de Fisher**

$$I_n(\theta) = -\mathbb{E}_\theta(\ell''(\theta)) = \mathbb{E}_\theta(J_n(\theta)), \qquad J_n(\theta) = -\ell''(\theta)$$

  $1/I_n(\theta)$ est la **variance asymptotique** de $\widehat{\theta}_{\mathrm{ML}}$, $I_n(\theta)$ est la courbure
- **L'information observée** est $J_n(\widehat{\theta}_{\mathrm{ML}})$
- **Attention !** Certains modèles, tel que $U[0, \theta]$, ne sont pas réguliers

Exemple $x_1, \ldots, x_n$ réalisations d'une loi $\exp(\lambda)$ avec densité $\lambda e^{-\lambda x} I(x \geq 0)$, $\lambda > 0$

Exemple

Exemple $x_1, \ldots, x_n$ réalisations d'une loi $\exp(\lambda)$ avec densité $\lambda e^{-\lambda x} I(x \geq 0)$, $\lambda > 0$

3.2 Estimation par intervalle

Les intervalles de confiance

Un élément clé de la statistique est de donner une idée de l'incertitude d'un constat

Soit θ un paramètre inconnu, et soit $\tilde{\theta}_n = 1$ une estimation de θ basée sur $y_1, \dots, y_n$:

- si $n = 10^5$ on est beaucoup plus sûr que $\theta \approx \tilde{\theta}_n$ que si $n = 10$
- pour exprimer ceci on aimerait donner un intervalle qui serait plus large quand $n = 10$ que quand $n = 10^5$, pour expliciter l'incertitude liée à $\tilde{\theta}_n$

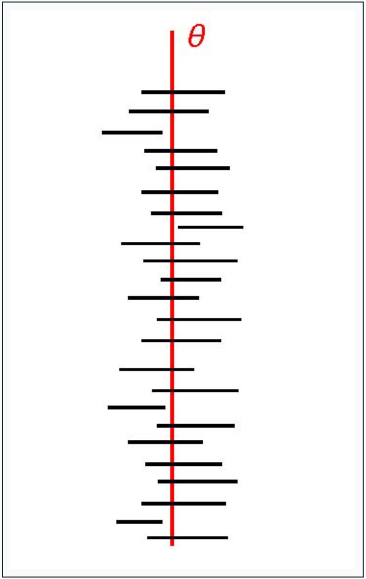
Définition: Soient $Y \equiv Y_1, \dots, Y_n$ des données issues d'une loi F de paramètre $\theta \in \mathbb{R}$. Un **intervalle de confiance** (IC, 'confidence interval' en anglais) (L_n, U_n) pour θ est une statistique sous forme d'intervalle qui contient θ avec une probabilité spécifiée $1 - \alpha$

$$\mathrm{Pr}_\theta(L_n \leq \theta \leq U_n) = 1 - \alpha \qquad \forall \theta$$

- $1 - \alpha \in (0, 1)$ est le **niveau**, souvent $\alpha \in \{0.05, 0.01, 0.1\}$
- Les bornes $L_n = L_n(Y), U_n = U_n(Y)$ sont des statistiques et non pas des inconnus
- Un IC **bilatéral**, de la forme (L_n, U_n), est le plus souvent utilisé
- Un IC **unilatéral à droite** est $(-\infty, U_n]$ tel que $\mathrm{Pr}_\theta(U_n \geq \theta) = 1 - \alpha$
- Un IC **unilatéral à gauche** est $[L_n, \infty)$ tel que $\mathrm{Pr}_\theta(L_n \leq \theta) = 1 - \alpha$

Interprétation d'un intervalle de confiance

- (L_n, U_n) est un intervalle aléatoire qui contient θ avec probabilité $(1 - \alpha)$
- Si on répète l'expérience avec d'autres données, on aura un autre intervalle de confiance
- Nous ne savons pas si notre IC contient θ, mais cette procédure nous fournit une garantie statistique que cet événement a une probabilité $(1 - \alpha)$



Construire un IC dans le cas normal

- Soient $Y_1, \ldots, Y_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, 1)$, et

$$\overline{Y}_n = \frac{1}{n} \sum_{j=1}^{n} Y_j$$

la moyenne de l'échantillon. On a $\overline{Y}_n \sim \mathcal{N}(\mu, 1/n)$ et donc

$$Z_n = n^{1/2}(\overline{Y}_n - \mu) \sim \mathcal{N}(0,1)$$

a une distribution qui ne dépend pas de μ. On obtient

$$\Pr_\mu \left(\overline{Y}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}} \leq \mu \leq \overline{Y}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}} \right) = 1 - \alpha$$

avec $z_\beta = \Phi^{-1}(\beta)$ les quantiles de la loi $\mathcal{N}(0,1)$

- Si $Y_1, \ldots, Y_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, et **on connaît** σ^2, alors

$$\frac{Z_n}{\sigma} = \frac{n^{1/2}(\overline{Y}_n - \mu)}{\sigma} \sim \mathcal{N}(0,1)$$

et l'intervalle de confiance pour μ est

$$\left[\overline{Y}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}} \sigma, \overline{Y}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}} \sigma \right]$$

- La largeur $\downarrow$ avec n, $\uparrow$ avec σ et avec le niveau $(1-\alpha)$, car $\Pr_\theta(L_n \leq \theta \leq U_n) \uparrow$

Exemple

Exemple On suppose que la résistance Y d'un certain type d'équipements électriques est distribuée selon une loi normale avec $\sigma = 0.12$ ohm.

Un échantillon de taille $n = 64$ a donné comme moyenne la valeur $\bar{y}_n = 5.34$ ohm.

Trouver un intervalle de confiance pour μ au niveau 95%.

Intervalle de Student

- Soient $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ avec σ^2 **inconnue**
- $\left[\overline{Y}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sigma, \overline{Y}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}\sigma\right]$ n'est plus un intervalle de confiance
- On suppose $n > 1$ et estime σ^2 par

$$S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \overline{Y}_n)^2$$

- On remplace σ par S_n : soit

$$T_n = \frac{Z_n}{S_n} = n^{1/2}\frac{\overline{Y}_n - \mu}{S_n} = \frac{Z_n/\sigma}{\sqrt{S_n^2/\sigma^2}}$$

- Nominateur et dénominateur indép, leurs lois ne dépendent pas de (μ, σ^2)
- T_n suit une **loi de Student avec $n-1$ degrés de liberté**, $T_n \sim t_{n-1}$, qui ne dépend de $\theta = (\mu, \sigma^2)$. C'est une loi symétrique, on dénote les quantiles par $t_{n-1,\beta}$
- L'intervalle de confiance qui en résulte est

$$\left[\overline{Y}_n - \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}}S_n, \overline{Y}_n + \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}}S_n\right]$$

- La largeur $\downarrow$ avec n, $\uparrow$ avec S_n et avec le niveau $(1-\alpha)$

Exemple

Exemple Pour déterminer le point de fusion μ d'un certain alliage, on a procédé à $n = 9$ observations qui ont donné une moyenne $\overline{y}_n = 1040^\circ$ avec $s_n = 16^\circ$. On suppose que les données suivent une loi normale.

Trouver un intervalle de confiance pour μ à niveau 95%.

Intervalles de confiance pour μ dans le cas normal : résumé

Soient $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ avec μ inconnu

- Pour σ **connue** $n^{1/2} \frac{\overline{Y}_n - \mu}{\sigma} \sim \mathcal{N}(0,1)$ et on a les IC de niveau $1-\alpha$

$$\left[\overline{Y}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sigma, \overline{Y}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}\sigma\right]$$

$$\left[\overline{Y}_n - \frac{z_{1-\alpha}}{\sqrt{n}}\sigma, \infty\right] \qquad \left[-\infty, \overline{Y}_n + \frac{z_{1-\alpha}}{\sqrt{n}}\sigma\right]$$

  avec $z_\beta = \Phi^{-1}(\beta)$ les quantiles de la loi $\mathcal{N}(0,1)$

- Pour σ **inconnue** (et $n > 1$) $n^{1/2} \frac{\overline{Y}_n - \mu}{S_n} \sim t_{n-1}$ et on a les IC de niveau $1-\alpha$

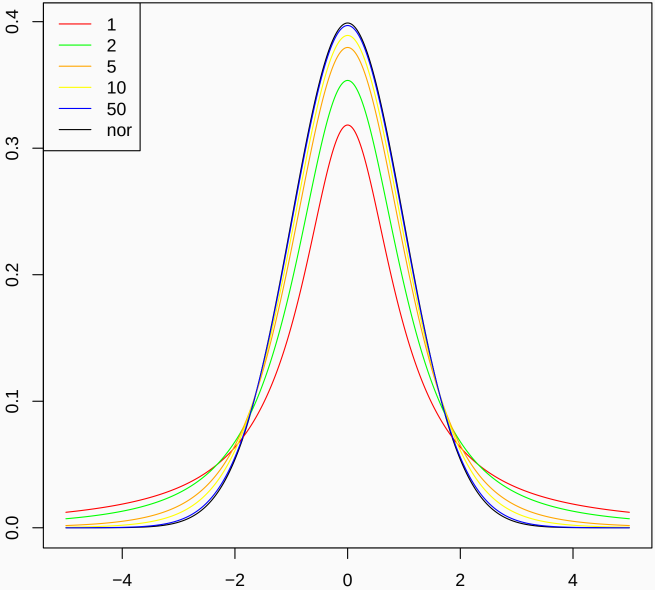
$$\left[\overline{Y}_n - \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}}S_n, \overline{Y}_n + \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}}S_n\right]$$

$$\left[\overline{Y}_n - \frac{t_{n-1,1-\alpha}}{\sqrt{n}}S_n, \infty\right] \quad \text{ou} \quad \left[-\infty, \overline{Y}_n + \frac{t_{n-1,1-\alpha}}{\sqrt{n}}S_n\right]$$

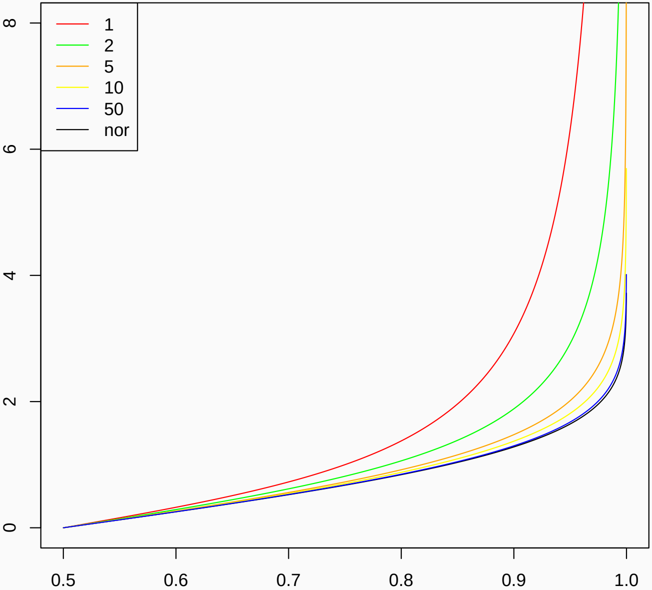
  avec t_{n-1} la **loi de Student avec $n-1$ degrés de liberté**, $t_{n-1,\beta}$ sont les quantiles de cette distribution, S_n définie à la diapositive 182.

t_{n-1} symétrique comme $\mathcal{N}(0,1)$, mais quantiles plus grands : $|t_{n-1,\beta}| > |z_\beta|$, $\beta \neq \frac{1}{2}$

Densités de t_k et $\mathcal{N}(0,1)$



Quantiles de t_k et $\mathcal{N}(0,1)$



Pivots

- Une fonction $T(Y_1, \ldots, Y_n, \theta)$ dont la loi est connue et ne dépend pas de θ s'appelle un **pivot**
- **Attention !** Ce n'est pas une statistique car c'est une fonction de θ
- Si $a \leq b$,

$$\alpha_1 = \Pr(T < a), \quad \alpha_2 = \Pr(T > b), \quad \alpha_1 + \alpha_2 = \alpha \in [0, 1]$$

(souvent $\alpha_1 = \alpha_2 = \alpha/2$) alors

$$\Pr_\theta(a \leq T \leq b) = \Pr_\theta(T \leq b) - \Pr_\theta(T < a) = (1 - \alpha_2) - \alpha_1 = 1 - \alpha$$

- Si on peut isoler θ, on peut trouver des variables aléatoires L et U telles que

$$\Pr_\theta(L \leq \theta \leq U) = 1 - \alpha$$

Exemples T_n (diapositive 182), Z_n (diapositive 179).

Pivot : cas uniforme

Exemple Soient $Y_1, \ldots, Y_n \sim U(0, \theta)$, $M_n = \max(Y_i)$. Alors $T(Y_1, \ldots, Y_n, \theta) = M_n/\theta$ est un pivot.

Intervalles de confiance approximatifs

- En pratique la plupart des intervalles de confiance se basent sur des approximations fournies par le théorème central limite, étant de la forme
$$(L_n, U_n) = (\widehat{\theta}_n - \sqrt{V_n} z_{1-\alpha/2}, \widehat{\theta}_n + \sqrt{V_n} z_{1-\alpha/2}),$$
où V_n est une estimateur de $\mathrm{var}_\theta(\widehat{\theta}_n)$ dont la racine s'appelle **erreur type (standard error)** de $\widehat{\theta}_n$. Sa réalisation $v_n^{1/2}$ est aussi appelée erreur type
- L'intervalle de confiance est approximatif dans le sens que
$$\mathrm{Pr}_\theta(L_n \leq \theta \leq U_n) \to 1 - \alpha, \qquad n \to \infty$$
- Dans des modèles réguliers, la variance asymptotique de $\widehat{\theta}_{\mathrm{ML}}$ (voir diapositive 174) est $1/I_n(\theta)$, estimée par $1/J_n(\widehat{\theta}_{\mathrm{ML}})$ et donc
$$(L_n, U_n) = \left(\widehat{\theta}_{\mathrm{ML}} - z_{1-\alpha/2}/\sqrt{J_n(\widehat{\theta}_{\mathrm{ML}})}, \widehat{\theta}_{\mathrm{ML}} + z_{1-\alpha/2}/\sqrt{J_n(\widehat{\theta}_{\mathrm{ML}})}\right)$$

Exemples

Exemple Soient $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \mathrm{Poiss}(\lambda)$, $\lambda > 0$ inconnu. Trouver une erreur type pour $\widehat{\lambda}_{\mathrm{ML}}$, et ainsi donner un intervalle de confiance approximatif de niveau 90% pour λ.

Exemples

Exemple Soient $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} U(0, \theta)$ avec θ inconnu et $\overline{Y}_n = n^{-1} \sum_{j=1}^{n} Y_j$. Utiliser le théorème central limite pour trouver un intervalle de confiance approximatif de niveau 95% pour θ.