

## ECN 7060, cours 12

William McCausland

2020-11-25

# Introduction, estimation par intervalle

- ▶ Estimateur par intervalle  $[L(X), U(X)]$ , estimation par intervalle  $[L(x), U(x)]$ .
- ▶ Les propriétés fréquentistes concernent la probabilité de couvrage

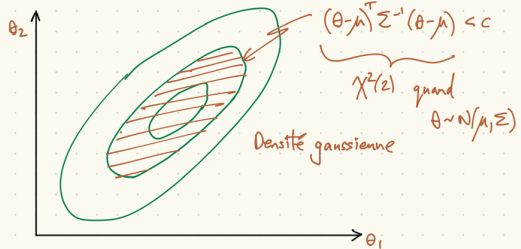
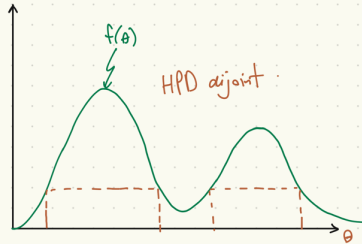
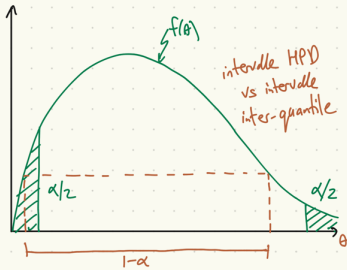
$$P_{\theta}[L(X) \leq \theta \leq U(X)].$$

- ▶ souvent une fonction de  $\theta$ , pas toujours (idéalement non)
  - ▶ restrictions sur le modèle pour obtenir cette non-dépendance
  - ▶ coefficient de confiance  $\inf_{\theta} P_{\theta}(L(X) \leq \theta \leq U(X))$ .
  - ▶ arbitrage : haute probabilité de couvrage v. intervalle court
- ▶ Les propriétés bayésiennes concernent la probabilité

$$P[L(x) \leq \theta \leq U(x)|x] \quad \text{ou} \quad P[l \leq \theta \leq u|x]$$

- ▶ Stratégie symétrique :  $L(x)$  et  $U(x)$  sont les quantiles  $\alpha/2$  et  $1 - \alpha/2$  de  $\theta$ ,  $U(x) - L(x)$  n'est pas forcément minimale
  - ▶ Intervalle de haute probabilité *a posteriori* :  $U(x) - L(x)$  minimale sous la contrainte  $P[L(x) \leq \theta \leq U(x)|x] = 1 - \alpha$ .

# Régions de haute probabilité et intervalles interquantile



# Estimation par ensemble

- ▶ Estimateur par ensemble  $C(X)$ , estimation par ensemble  $C(x)$ .
- ▶ Probabilité d'intérêt fréquentiste :  $P_\theta(\theta \in C(X))$ .
  - ▶  $\theta$  est fixe
  - ▶ la région  $C(X)$  est aléatoire
  - ▶ analyse *ex ante*
- ▶ Probabilité d'intérêt bayésienne :  $P(\theta \in C(x)|x)$ .
  - ▶  $x$  est fixe (l'échantillon observé)
  - ▶ l'élément  $\theta$  est aléatoire (la probabilité est conditionnelle)
  - ▶ analyse *ex post*

# Inversion d'une statistique test

- ▶ Pour chaque  $\theta_0$ , soit  $A(\theta_0)$  la région de non-rejet pour un test de niveau  $\alpha$  de l'hypothèse nulle  $H_0 : \theta = \theta_0$ .
- ▶ Alors  $A(\theta)$  vérifie  $P_\theta[X \notin A(\theta)] \leq \alpha$  et l'inégalité est sharp.
- ▶ Définiez, pour chaque  $x \in \mathcal{X}$ ,  $C(x) = \{\theta : x \in A(\theta)\}$ .
- ▶ Notez que  $x \in A(\theta) \Leftrightarrow \theta \in C(x)$ .
- ▶ Résultat :  $C(X)$  est une région de confiance avec coefficient de confiance  $(1 - \alpha)$ .
- ▶ Preuve :
  - ▶ Puisque le niveau du test est de  $\alpha$ ,

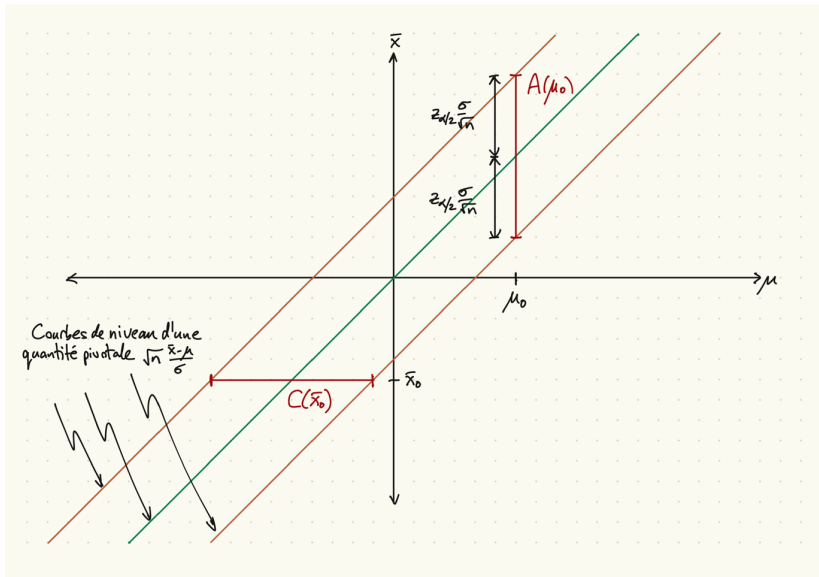
$$P_\theta[X \notin A(\theta)] \leq \alpha.$$

- ▶ Alors

$$P_\theta[\theta \in C(X)] = P_\theta[X \in A(\theta)] \geq (1 - \alpha).$$

- ▶ La première inégalité est sharp donc la deuxième l'est aussi.

## $C(\bar{x})$ et $A(\theta)$ pour un exemple gaussien



## Exemple gaussien, $\sigma^2$ connu

- ▶ Supposons que  $X_1, \dots, X_n \sim \text{iid } N(\mu, \sigma^2)$ ,  $\sigma^2$  connu.
- ▶ Statistique LRT pour  $H_0: \mu = \mu_0$  contre  $H_1: \mu \neq \mu_0$  :

$$\lambda(x) = \frac{\exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2 \right]}{\exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right]}$$

- ▶ Puisque  $\sum_{i=1}^n (x_i - \mu_0)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2$ ,

$$\lambda(x) = \exp[-n(\bar{x} - \mu_0)^2 / (2\sigma^2)].$$

- ▶ La loi de  $\bar{X}$  est connue :  $\bar{X} \sim N(\mu, \sigma^2/n)$
- ▶ Pour le test avec  $A(\mu_0) = \{x: |\bar{x} - \mu_0| \leq z_{\alpha/2}\sigma/\sqrt{n}\}$ , la probabilité de rejet quand  $\mu = \mu_0$  est de  $\alpha$ .
- ▶ Conditions équivalentes à  $x \in A(\mu_0)$  :

$$\begin{aligned} |\bar{x} - \mu_0| \leq z_{\alpha/2}\sigma/\sqrt{n} &\Leftrightarrow -z_{\alpha/2}\sigma/\sqrt{n} \leq \mu_0 - \bar{x} \leq z_{\alpha/2}\sigma/\sqrt{n} \\ &\Leftrightarrow \bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \leq \mu_0 \leq \bar{x} + z_{\alpha/2}\sigma/\sqrt{n} \end{aligned}$$

- ▶ Alors  $P[\bar{X} - z_{\alpha/2}\sigma/\sqrt{n} \leq \mu_0 \leq \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}] = 1 - \alpha$ .

# Quantités pivotales

- ▶ Une fonction  $Q(X, \theta)$  est pivotale si sa loi ne dépend pas de  $\theta$ .
- ▶ Interprétation bayésienne : sa loi ne dépend pas de  $f(\theta)$
- ▶ Famille  $f(x|\mu) = f_0(x - \mu)$  :  $Q(X, \theta) = \bar{X} - \mu$  est pivotale.
- ▶ Preuve :
  - ▶ Soit  $Z_i \sim f_0(z)$ . Sa distribution ne dépend pas de  $\mu$ .
  - ▶ Si  $X_i \sim f(x|\mu) = f_0(x - \mu)$ ,

$$(X_1, \dots, X_n) \sim (Z_1 + \mu, \dots, Z_n + \mu)$$

$$\bar{X} - \mu \sim \frac{1}{n} \sum_{i=1}^n (Z_i + \mu) - \mu = \bar{Z}$$

- ▶ La loi de  $\bar{Z}$  (et de  $Q(X, \theta) = \bar{X} - \mu$ ) ne dépend pas de  $\mu$ .
- ▶ Famille  $f(x|\sigma) = \frac{1}{\sigma} f_0(x/\sigma)$  :  $Q(X, \sigma^2) = \bar{X}/\sigma$  est pivotale.
- ▶ Famille  $f(x|\mu, \sigma^2) = \frac{1}{\sigma} f_0((x - \mu)/\sigma)$  :  $Q_1(X, \theta) = (\bar{X} - \mu)/\sigma$ ,  $Q_2(X, \theta) = (\bar{X} - \mu)/S$ ,  $Q_3(X, \theta) = S/\sigma$  sont pivotales.



## Utiliser une quantité pivotale pour construire un ensemble de confiance

- ▶  $C(X) = \{\theta : Q(X, \theta) \in \mathcal{A}\}$  est un estimateur par ensemble de  $\theta$  dont la probabilité  $P_\theta(C(X))$  ne dépend pas de  $\theta$ .
- ▶ Stratégie : trouver une quantité pivotale  $Q(X, \theta)$  et un ensemble  $\mathcal{A}$  avec de bonnes propriétés ( $C(X)$  petit,  $P_\theta(C(X))$  grand).

# Exemples gaussiens I

- ▶ Supposons que  $X_1, \dots, X_n \sim \text{iid } N(\mu, \sigma^2)$ .
- ▶ Quantités pivotales :
  - ▶  $Z = \sqrt{n}(\bar{X} - \mu)/\sigma \sim N(0, 1)$ ,
  - ▶  $T_{n-1} = \sqrt{n}(\bar{X} - \mu)/S \sim t(n-1)$ .
  - ▶  $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$ .
- ▶ Cas où  $\sigma^2$  est connu :

$$1 - \alpha = P_{\theta}(-z_{\alpha/2} \leq -Z \leq z_{\alpha/2}) = P_{\theta}(\mu \in C(X)).$$

où  $C(X)$  est l'estimateur par ensemble suivant

$$C(X) = \{\mu : \bar{X} - z_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}\}.$$

## Exemples gaussiens II

- Cas où  $\sigma^2$  n'est pas connu, intervalle pour  $\mu$  :

$$1 - \alpha = P_{\theta}(-t_{n-1,\alpha/2} \leq -T_{n-1} \leq t_{n-1,\alpha/2}) = P_{\theta}(C(X)),$$

où

$$C(X) = \{\mu: \bar{X} - t_{n-1,\alpha/2}S/\sqrt{n} \leq \mu \leq \bar{X} + t_{n-1,\alpha/2}S/\sqrt{n}\}.$$

- Cas où  $\sigma^2$  n'est pas connu, intervalle pour  $\sigma^2$  :

$$1 - \alpha = P_{\theta}(\chi_{n-1,1-\alpha/2} \leq (n-1)S^2/\sigma^2 \leq \chi_{n-1,\alpha/2}) = P_{\theta}(C(X)),$$

où

$$C(X) = \left\{ \sigma^2: \frac{(n-1)S^2}{\chi_{n-1,\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{n-1,1-\alpha/2}} \right\}.$$

## Un aparté

- ▶ Soit  $T$  une variable aléatoire avec fonction de répartition  $F$  inversible.
- ▶  $F(t)$  est une fonction,  $F(T)$  est une variable aléatoire avec une loi sur  $[0, 1]$ .
- ▶ Soit  $G$  la fonction de répartition de  $F(T)$ .
- ▶ Pour  $u \in [0, 1]$ ,

$$G(u) = P[F(T) \leq u] = P[T \leq F^{-1}(u)] = F[F^{-1}(u)] = u.$$

- ▶ Alors  $F(T) \sim U(0, 1)$ .

## Pivot de la fonction de répartition

- ▶ Soit  $T$  une statistique avec fonction de répartition  $F_T(t|\theta)$ .
- ▶ Si  $F_T$  est toujours inversible,  $F_T(T|\theta) \sim U(0, 1)$ , une loi qui ne dépend pas de  $\theta$ .
- ▶ Supposons que  $T$  est stochastiquement croissante en  $\theta$ .
- ▶ C'est à dire que  $F_T(t, \theta)$  est décroissante en  $\theta$ .
- ▶ Pour  $t$  donné, soit  $\theta_L(t)$  et  $\theta_U(t)$  les solutions de

$$F_T(t|\theta_U(t)) = \alpha_1 \quad F_T(\theta_L(t)) = 1 - \alpha_2.$$

- ▶ Pour tous  $t, \theta$ ,

$$\theta > \theta_U(t) \Leftrightarrow F_T(t, \theta) < \alpha_1$$

$$\theta < \theta_L(t) \Leftrightarrow F_T(t, \theta) > 1 - \alpha_2$$

- ▶ Considérer l'intervalle de confiance  $[\theta_L, \theta_U]$  :

$$\{t: \theta_L(t) \leq \theta \leq \theta_U(t)\} = \{t: \alpha_1 \leq F_T(t|\theta) \leq 1 - \alpha_2\},$$

$$P_\theta[\theta_L(T) \leq \theta \leq \theta_U(T)] = P_\theta[\alpha_1 \leq F_T(T|\theta) \leq 1 - \alpha_2] = 1 - \alpha_1 - \alpha_2.$$

# Graphique, pivot de la fonction de répartition

