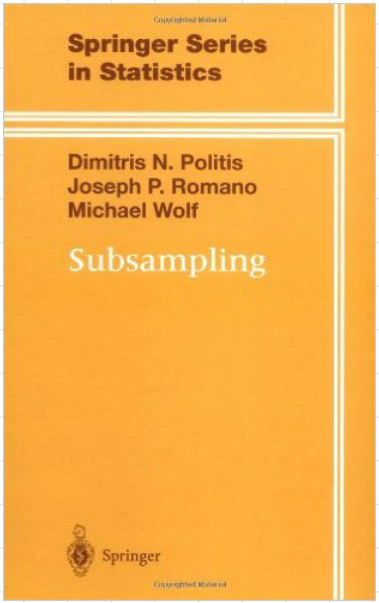
Application du bootstrap et autres techniques de rééchantillonnage

email : etienne.roquain@upmc.fr

ROQUAIN E.



Chapitre 1 : Introduction

Chapitre 2 : Bootstrap

Chapitre 3 : Subsampling

Chapitre 4 : Test par rééchantillonnage

Chapitre 5 : Multitest par rééchantillonnage

TP1 : prise en main Rstudio + 'boot'

TP2 : bootstrap - subsampling données law school

TP3 : tests par rééchantillonage

TP4 : tests en grande dim données cancer prostate

① Test classique

Rappels : $\mathcal{X}_n = (X_1 \dots X_n)$ iid $\sim P \in \mathcal{P}$ modèle (ensemble de lois sur $\mathbb{R}$)

H_0 : "$P \in \mathcal{P}_0$" contre H_1 : "$P \in \mathcal{P}_1$" pour deux sous ensembles $\mathcal{P}_0, \mathcal{P}_1 \subset \mathcal{P}$, $\mathcal{P}_0 \cap \mathcal{P}_1 = \emptyset$, $\mathcal{P}_0 \cup \mathcal{P}_1 = \mathcal{P}$

- Statistique de test $T_n = T_n(\mathcal{X}_n)$ ("grande" sous H_1)
- Test de niveau α : rejet de H_0 si $T_n > c_\alpha$ tel que $\forall P \in \mathcal{P}_0$, $\mathbb{P}(T_n > c_\alpha) \leq \alpha$ (proba de rejet à tort)
- pvalue $p(x) = \inf \{\alpha \in [0,1] : T_n > c_\alpha\}$ (indice de confiance en H_0)

  aussi $p(x) = \sup_{P \in \mathcal{P}_0} \mathbb{P}(T_n(X) > T_n(x))$ (probabilité d'avoir une stat observée aussi extrême que sous H_0)
- fonction puissance du test : $\pi : P \in \mathcal{P} \mapsto \mathbb{P}(T_n > c_\alpha)$ (proba de rejet) (donc $\pi(P) \leq \alpha$ pour $P \in \mathcal{P}_0$)
- puissance $\pi(P)$, $P \in \mathcal{P}_1$ (proba de rejet à raison)

Par exemple : $X_1 \dots X_n$ iid $\sim \mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$ $\quad H_0$: "$\theta = 0$" contre H_1 : "$\theta \neq 0$" $\quad (\mathcal{P}_0 = \{\mathcal{N}(0,1)\}$ et $\mathcal{P}_1 = \{\mathcal{N}(\theta,1), \theta \neq 0\})$

$T_n = \sqrt{n}\,|\bar{X}_n|$ où $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

- sous H_0, $T_n \sim |Z|$ où $Z \sim \mathcal{N}(0,1)$

  donc $c_\alpha = \Phi^{-1}(1 - \frac{\alpha}{2})$ quantile d'ordre $1 - \frac{\alpha}{2}$ d'une $\mathcal{N}(0,1)$

- pvalue $p(X) = \inf\{\alpha \in [0,1] : \underbrace{T_n > c_\alpha}\}$

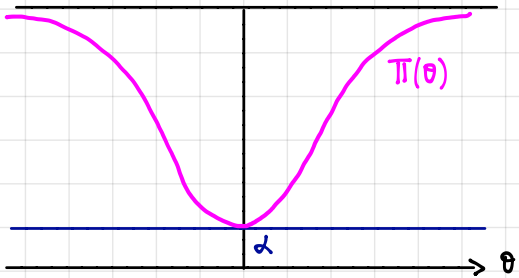
  $\Leftrightarrow \quad \Phi(\sqrt{n}|\bar{X}_n|) > 1 - \frac{\alpha}{2} \quad \Leftrightarrow \quad \alpha > 2(1 - \Phi(\sqrt{n}|\bar{X}_n|))$

  $p(X) = 2(1 - \Phi(\sqrt{n}|\bar{X}_n|))$

- fonction puissance $\theta \in \mathbb{R}$

  $\pi(\theta) = \mathbb{P}_\theta(\sqrt{n}|\bar{X}_n| > c_\alpha)$

  $= 1 - \Phi(c_\alpha - \sqrt{n}\theta) + \Phi(-c_\alpha - \sqrt{n}\theta)$



cadre non gaussien que faire ?

② Test par bootstrap

$\mathcal{X}_n = (X_1 \dots X_n)$ iid $\sim P \in \mathcal{P}$ modèle (ensemble de lois sur $\mathbb{R}$)

H_0 : "$P \in \mathcal{P}_0$" contre H_1 : "$P \in \mathcal{P}_1$" ; Statistique de test $T_n = T_n(\mathcal{X}_n)$

Test de niveau asymptotique α : c_α tel que $\forall P \in \mathcal{P}_0$, $\overline{\lim_n} \{ \mathbb{P}(T_n > c_\alpha) \} \leq \alpha$

$L_n(P) = \mathcal{L}_{X_1 \sim P}(T_n(\mathcal{X}_n)) \simeq L_n(\hat{Q}_n)$ par bootstrap

choix $c_\alpha = c_\alpha(\mathcal{X}_n)$ avec $\mathbb{P}_{X_1^* \dots X_n^* \text{ iid} \sim \hat{Q}_n} \left(T_n(\mathcal{X}_n^*) > c_\alpha \mid \mathcal{X}_n \right) \leq \alpha$

→ bootstrap classique $\hat{Q}_n = \hat{P}_n$ on aura bien $L_n(P) \simeq L_n(\hat{P}_n)$ lorsque $P \in \mathcal{P}_0$ ou $\mathcal{P}_1$

donc $\mathbb{P}(T_n > c_\alpha(\mathcal{X}_n)) \simeq \alpha$ lorsque $P \in \mathcal{P}_0$ ou $\mathcal{P}_1$

Ainsi, la fonction puissance $\pi(P) \simeq \alpha$ pour tout $P \in \mathcal{P}_0$ ou $P \in \mathcal{P}_1$ même puissance que si on rejetait au hasard avec probabilité α

→ bootstrap 'sous contrainte' $\hat{Q}_n$ = "$\hat{P}_n$ contraint"

$\hat{Q}_n$ est une approximation de P sous la contrainte $\hat{Q}_n \in \mathcal{P}_0$ bootstrap sous H_0

Ainsi, sous H_0 contrôle de $\hat{Q}_n \simeq \hat{P}_n$; sous H_1 $L_n(P)$ 'loin' de $L_n(\hat{Q}_n)$ ($P \in \mathcal{P}_1$, $\hat{Q}_n \in \mathcal{P}_0$) donc $\mathbb{P}(T_n > c_\alpha(\mathcal{X}_n)) \simeq 1$

pvaleur par bootstrap

$$p(x_n) = \underset{X_1^* \dots X_n^* \text{ iid} \sim \hat{Q}_n}{\mathbb{P}} \left(T_n(x_n^*) \geq T_n(x_n) \mid x_n \right)$$

(bien sous H_0) $\hat{Q}_n \in \mathcal{P}_0$

Approximation Monte Carlo : B fixé (par ex 10 000)

$x_n^{*(1)} \dots x_n^{*(B)}$ B tirages d'échantillons bootstrap de x_n

$$\tilde{p}(x_n) = \frac{1}{B} \sum_{j=1}^{B} \mathbb{1}\{ T_n(x_n^{*(j)}) \geq T_n(x_n) \}$$

(proportion de stat sous H_0 qui dépasse la stat observée)

Théorie $C_\alpha(x_n)$ obtenu par bootstrap

Résultat type si $P \in \mathcal{P}_0$, $d(\hat{Q}_n, P) \xrightarrow{P} 0$ (+ autres hypothèses type tableau triangulaire)
sur loi de la stat $T_n(x_n)$ $X_1 \dots X_n \sim P_n$

$\Rightarrow \forall P \in \mathcal{P}_0$, $\mathbb{P}\left(T_n(x_n) > C_\alpha(x_n) \right) \xrightarrow[n \to +\infty]{} \alpha$ (le niveau est le bon)

Exemple : test non paramétrique de la moyenne

$x_n = (X_1 \dots X_n)$ iid $\sim F_\theta$ avec $F_\theta(x) = G(x-\theta)$ où G est centrée et continue — deux paramètres | θ intérêt, G nuisance

H_0 : "$\theta = 0$" contre H_1 : "$\theta \neq 0$" (ie loi centrée contre non centrée)

Approximation de P sous H_0 par $\hat{Q}_n$ la loi uniforme sur $\{ X_i - \overline{X}_n \,,\, 1 \leq i \leq n \}$

- $T_n = \sqrt{n}\,|\overline{X}_n|$ a une loi sous H_0 approchée par $L_n(\hat{Q}_n) = \mathcal{L}\left(\sqrt{n}\,|\overline{X}_n^*| \mid x_n \right)$, $X_1^* \dots X_n^*$ iid $\sim \hat{Q}_n$
- $c_\alpha(x_n) = 1-\alpha/2$ quantile de $L_n(\hat{Q}_n)$ (ou MC approx)
- Test final par bootstrap rejette H_0 si $T_n > c_\alpha(x_n)$
- $p(x) = \mathbb{P}_{X_1^* \dots X_n^* \text{ iid} \sim \hat{Q}_n}\left(\sqrt{n}\,|\overline{X}_n^*| \geq \sqrt{n}\,|\overline{x}_n| \mid x_n \right)$ ($\sqrt{n}\,|\overline{x}_n|$ observé)

Bonus : approximation de la fonction puissance $\pi : \theta \in \mathbb{R} \mapsto \mathbb{P}_\theta(T_n > c_\alpha(x_n))$ (restreinte au paramètre d'intérêt)

$$\pi(\theta) = \mathbb{P}_\theta\left(\sqrt{n}(\overline{X}_n - \theta) > c_\alpha(x_n) - \sqrt{n}\,\theta \right) + \mathbb{P}_\theta\left(\sqrt{n}(\overline{X}_n - \theta) < -c_\alpha(x_n) - \sqrt{n}\,\theta \right)$$

$$\simeq \mathbb{P}_{X_1^* \dots X_n^* \text{ iid} \sim \hat{Q}_n}\left(\sqrt{n}\,\overline{X}_n^* > c_\alpha(x_n^*) - \sqrt{n}\,\theta \mid x_n \right) + \mathbb{P}_{X_1^* \dots X_n^* \text{ iid} \sim \hat{Q}_n}\left(\sqrt{n}\,\overline{X}_n^* < -c_\alpha(x_n^*) - \sqrt{n}\,\theta \right)$$

($c_\alpha(x_n^*)$ 'double' bootstrap)

③ Test par subsampling

$\mathcal{X}_n = (X_1 \dots X_n)$ iid $\sim P \in \mathcal{P}$ modèle (ensemble de lois sur $\mathbb{R}$)

H_0 : "$P \in \mathcal{P}_0$" contre H_1 : "$P \in \mathcal{P}_1$" ; Statistique de test $T_n(\mathcal{X}_n) = \tau_n\, S_n(\mathcal{X}_n)$ (taux)

$$\mathcal{L}_P\big(\tau_n S_n(\mathcal{X}_n)\big) \simeq \text{loi uniforme sur } \Big\{\tau_b S_b(X_i, i\in S),\ S\subset\{1,\dots,n\},\ |S|=b\Big\}$$

$$c_\alpha = c_\alpha(\mathcal{X}_n) \text{ avec } \frac{1}{\binom{n}{b}} \sum_{\substack{|S|=b \\ S\subset\{1\dots n\}}} \mathbb{1}\{\tau_b S_b(X_i, i\in S) > c_\alpha(\mathcal{X}_n)\} \lesssim \alpha$$

ça marche sans s'occuper de H_0, juste $\frac{b}{n}\to 0$, $b\to+\infty$, $\frac{\tau_b}{\tau_n}\to 0$.

on fait comme d'habitude !

si $\begin{cases} \text{sous } H_0 : \tau_n S_n(\mathcal{X}_n) \xrightarrow{\mathcal{L}} L(P) \text{ continue} \\ \text{sous } H_1 : S_n(\mathcal{X}_n) \xrightarrow{P} \text{cste} > 0 \end{cases}$ (souvent vrai)

alors $\begin{cases} \text{sous } H_0 : c_\alpha(\mathcal{X}_n) \simeq \text{quantile de } L(P) \longrightarrow \mathbb{P}\big(\tau_n S_n(\mathcal{X}_n) > c_\alpha(\mathcal{X}_n)\big) \xrightarrow[n\to+\infty]{} \alpha \text{ ok} \\ \text{sous } H_1 : c_\alpha(\mathcal{X}_n) \simeq \tau_b \times \text{cste} \end{cases}$

$\hookrightarrow \mathbb{P}\big(\tau_n S_n(\mathcal{X}_n) > c_\alpha(\mathcal{X}_n)\big)$

$\simeq \mathbb{P}(\tau_n \text{ cste} > \tau_b \text{ cste}) \xrightarrow[n\to+\infty]{} 1$

Théorie c'est un théorème !

pvaleur par subsampling

$$p(x_n) = \frac{1}{\binom{n}{b}} \sum_{\substack{|S|=b \\ S \subset \{1,\dots,n\}}} \mathbb{1}\{ \tau_b\, S_b(X_i, i \in S) \geq \tau_n\, S_n(x_n) \}$$

même idée : sous H_0 $\tau_b\, S_b(X_i, i \in S)$ $\simeq$ même loi que $\tau_n\, S_n(x_n)$ pvaleur unif ok

sous H_1 $\tau_b\, S_b(X_i, i \in S) \simeq \tau_b$ cste $<$ τ_n cste $\simeq \tau_n\, S_n(x_n)$ pvaleur $\simeq 0$

Approximation Monte Carlo : B fixé (par ex 10 000)

$S_1 \dots S_B$ B sous ensembles de $\{1,\dots,n\}$ de tailles b tirés uniformément et de façon iid

$$\tilde{p}(x_n) = \frac{1}{B} \sum_{j=1}^{B} \mathbb{1}\{ \tau_b\, S_b(X_i, i \in S_j) \geq \tau_n\, S_n(x_n) \}$$

Exemple test de la moyenne $x_n = (X_1 \dots X_n)$ iid $\sim F_\theta(x) = G(x-\theta)$ où G est centrée continue

H_0 : "$\theta = 0$" contre H_1 : "$\theta \neq 0$" on peut appliquer le subsampling avec

$$T_n(x_n) = \sqrt{n}\, |\overline{X}_n|$$

(τ_n ↑ , ↖ $S_n(x_n)$)

④ Test exact par randomisation

deux cas : symétrisation ; permutation

Dans certains cas, on peut avoir un contrôle du niveau qui est non asymptotique

Test par symétrisation $x_n = (X_1 \dots X_n)$ iid $\sim F_\theta(x) = G(x-\theta)$ où G est centrée continue et symétrique / 0

H_0 : " $\theta = 0$ " contre H_1 : " $\theta \neq 0$ " ; $T_n(x_n) = \sqrt{n}\,|\overline{X_n}| = \sqrt{n}\left|\frac{1}{n}\sum_{i=1}^n X_i\right|$

Alors nous avons la propriété suivante

Lemme : si $\varepsilon_n = (\varepsilon_{1,n}, \dots, \varepsilon_{n,n})$ sont iid avec $\varepsilon_{1,n} = \begin{cases} -1 & \text{proba } 1/2 \\ +1 & \text{proba } 1/2 \end{cases}$ (signes aléatoires)

Alors sous H_0, $(X_1, \dots, X_n) \sim (\varepsilon_{1,n} X_1, \dots, \varepsilon_{n,n} X_n)$

Preuve.

invariance de la loi de l'observation sous H_0

Théorème : si $c_\alpha(x_n) = \inf\{x \in \mathbb{R} : \mathbb{P}(T_n(\underbrace{\varepsilon_n \times x_n}_{= (\varepsilon_{1,n} X_1, \dots, \varepsilon_{n,n} X_n)}) > x \mid x_n) \leq \alpha$ avec ε_n comme plus haut

Alors sous H_0, $\mathbb{P}(T_n(x_n) > c_\alpha(x_n)) \leq \alpha \quad (\forall n)$

Preuve.

Remarque : sous H_1, $T_n(\varepsilon_n x_n) = \sqrt{n}\left|\frac{1}{n}\sum_{i=1}^n \overbrace{\varepsilon_{i,n} X_i}^{\text{centrée}}\right|$ donc moyenne tuée dans $c_\alpha(x_n)$ et $\mathbb{P}(T_n(x_n) > c_\alpha(x_n)) \simeq 1$ si la moyenne est grande

Test par permutation

$\mathcal{X}_n = (X_1, \dots, X_n) = \big((Y_1, \dots, Y_{n_1}), (Z_1, \dots, Z_{n_2})\big)$ deux échantillons indépendants

$(Y_1 \dots Y_{n_1})$ iid $\sim Q_1$ 'sain'

$(Z_1 \dots Z_{n_2})$ iid $\sim Q_2$ 'malade'

On cherche à tester H_0 : "$Q_1 = Q_2$" contre H_1 : "$Q_1 \neq Q_2$"

Statistique de test $T_n(\mathcal{X}_n)$ à choisir ($T_n(\mathcal{X}_n) = |\bar{Z} - \bar{Y}|$ par ex si on veut détecter des différences dans les moyennes ou alors test de rang etc

Alors nous avons la propriété suivante

Lemme : si σ permutation aléatoire uniforme dans $\mathfrak{S}_n = \text{permut}(\{1, \dots, n\})$ (n tirages sans remise)

Alors sous H_0, $(X_1, \dots, X_n) \sim (X_{\sigma(1)}, \dots, X_{\sigma(n)})$

Preuve.

invariance de la loi de l'observation sous H_0

Théorème : si $c_\alpha(\mathcal{X}_n) = \inf\{x \in \mathbb{R} : \mathbb{P}\big(T_n(\underbrace{\mathcal{X}_n^\sigma}_{= (X_{\sigma(1)}, \dots, X_{\sigma(n)})}) > x \mid \mathcal{X}_n\big) \leq \alpha$ avec σ comme plus haut

Alors sous H_0, $\mathbb{P}\big(T_n(\mathcal{X}_n) > c_\alpha(\mathcal{X}_n)\big) \leq \alpha \quad (\forall n)$

Preuve.

Remarque : sous H_1, $T_n(\mathcal{X}_n^\sigma) = \sqrt{n} \left| \frac{1}{n_2} \sum_{i=n_1+1}^{n} X_{\sigma(i)} - \frac{1}{n_1} \sum_{i=1}^{n_1} X_{\sigma(i)} \right|$ la différence entre les deux groupes est 'tuée' dans $c_\alpha(\mathcal{X}_n)$

$\rightarrow$ $\mathbb{P}\big(T_n(\mathcal{X}_n) > c_\alpha(\mathcal{X}_n)\big)$ devrait être proche de 1

p-valeurs par randomisation comme d'habitude

⑤ Test de type Kolmogorov - Smirnov

Rappels : $\mathcal{X}_n = (X_1, \dots, X_n)$ iid de loi P avec fonction de répartition F continue

$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}$, $x \in \mathbb{R}$, fonction de répartition empirique de $\mathcal{X}_n$

P_0 loi avec F_0 fonction de répartition continue ($= F$ ou pas)

Distance de Kolmogorov-Smirnov (renormalisée) entre $\hat{F}_n$ et F_0 :

$$\sqrt{n} \, \| \hat{F}_n - F_0 \|_\infty = \sqrt{n} \sup_{x \in \mathbb{R}} | \hat{F}_n(x) - F_0(x) | \; ; \quad \text{loi associée } L_n(P \mid P_0) = \mathcal{L}_P \left(\sqrt{n} \, \| \hat{F}_n - F_0 \|_\infty \right)$$

Propriétés :

- $$\| \hat{F}_n - F_0 \|_\infty = \max_{1 \leq i \leq n} \left\{ \max \left(\frac{i}{n} - F_0(X_{(i)}) \, , \; F_0(X_{(i)}) - \frac{i-1}{n} \right) \right\}$$

  calculable avec 'ks.test' en R

- en $P = P_0$, $L_n(P_0 \mid P_0) = \mathcal{L}\left(\sqrt{n} \sup_{t \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{U_i \leq t\}} - t \right| \right)$

  $U_1 \dots U_n$ iid $\sim \mathcal{U}(0,1)$

  ne dépend pas de P_0
  loi accessible par Monte Carlo et des simulations d'uniformes

Test d'adéquation à une loi

Pour F_0 continue (donnée), on cherche à tester H_0 : "$P = P_0$" contre H_1 : "$P \neq P_0$"

Statistique de test $T_n(x_n) = \sqrt{n}\, \| \hat{F}_n - F_0 \|_\infty$

Test de niveau α : rejette H_0 si $T_n(x_n) > c_\alpha$ avec c_α quantile d'ordre $1-\alpha$ de $L_n(P_0 | P_0)$

↓ approchable par Monte Carlo

$$= \mathcal{L}_{P_0}\left(\sqrt{n}\, \| \hat{F}_n - F_0 \|_\infty \right)$$

simulable avec des uniformes

p-valeur $p(x_n) = \mathbb{P}\left(Z \geq T_n(x_n) \mid x_n \right)$

où $Z \sim L_n(P_0 | P_0)$

$$\simeq \frac{1}{B} \sum_{j=1}^{B} \mathbb{1}\{ Z_j \geq T_n(x_0) \}$$ où $Z_1 \dots Z_B$ iid $\sim L_n(P_0 | P_0)$

(chaque Z_j venant de $U_1^{(j)} \dots U_n^{(j)}$ iid $U(0,1)$)

sous H_1 il suffit que $\hat{F}_n(x)$ s'éloigne de $F_0(x)$ en un seul x pour que T_n soit 'grand'

marche bien dans le 'ventre mou' de la distribution

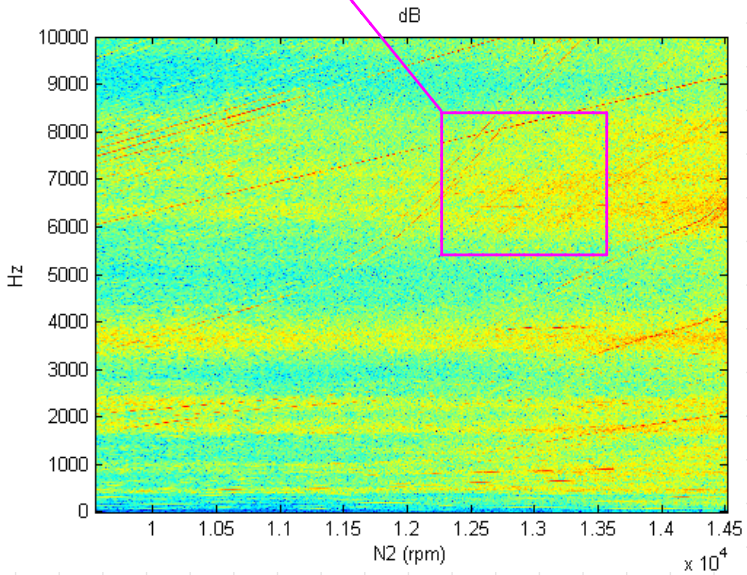
R s.test tout cela est fait en R avec la fonction 'R s.test'

Remarque : ici pas besoin de bootstrap !

Données moteur d'avion

Besoin de diagnostiquer l'état de santé d'un moteur

loi ?



fréquences des vibrations

régime rotation moteur

→ il est supposé que les scores des vingtiles (après décomposition dans une base) suivent une loi Gamma

→ cette hypothèse est-elle raisonnable ?

→ adéquation à la famille des lois gamma

Test d'adéquation à une famille de lois

$\mathcal{X}_n = (X_1, \dots, X_n)$ iid loi P avec F fonction de répartition associée

$\mathcal{P}_0$ famille paramétrique $\mathcal{P}_0 = \{Q_\theta, \theta \in \Theta\}$ disons $\Theta \subset \mathbb{R}$, Q_θ de fonction de répartition G_θ

On cherche à tester H_0 : "$P \in \mathcal{P}_0$" contre H_1 : "$P \notin \mathcal{P}_0$" (hypothèse nulle composite)

Statistique de test naturelle $T_n(\mathcal{X}_n) = \sqrt{n} \| \hat{F}_n - G_{\hat{\theta}_n} \|_\infty$ pour $\hat{\theta}_n = \hat{\theta}_n(\mathcal{X}_n)$ estimateur de θ dans le modèle $\mathcal{P}_0$, défini par H_0

Loi de $T_n(\mathcal{X}_n)$ notée $L_n(P \mid \mathcal{P}_0)$

Exemple : adéquation à une famille de lois gamma

$\mathcal{P}_0 = \{\gamma(\theta, 1), \theta > 0\}$, $\hat{\theta}_n = \overline{X}_n$ par la méthode des moments

$T_n(\mathcal{X}_n)$ distance de Kolmogorov Smirnov au 'meilleur candidat' parmi les lois gamma

Problème : sous H_0 comment obtenir $L_n(P \mid \mathcal{P}_0)$? (dépend du $P \in \mathcal{P}_0$)

- **par bootstrap :** $L_n(P|\mathcal{P}_0) \simeq L_n(\hat{Q}_n|\mathcal{P}_0)$

où $\hat{Q}_n$ est une approximation de P sous la contrainte $\hat{Q}_n \in \mathcal{P}_0$ — bootstrap sous H_0

On choisit $\hat{Q}_n = Q_{\hat{\theta}_n}$

ce qui donne $L_n(\hat{Q}_n|\mathcal{P}_0) = \mathcal{L}\left(\sqrt{n}\,\|\hat{F}_n^* - G_{\hat{\theta}_n^*}\|_\infty \;\middle|\; \mathcal{X}_n\right)$ où $X_1^* \dots X_n^* \overset{iid}{\sim} Q_{\hat{\theta}_n}$

bootstrap paramétrique

- **par subsampling** $b/n \to 0$, $b \to +\infty$ à choisir

calibration sous H_0 gratuite grâce au changement d'échelle

$L_n(P|\mathcal{P}_0) \simeq L_{n,b}(\hat{P}_n|\mathcal{P}_0)$

=

loi uniforme sur $\left\{ \sqrt{b} \sup_{x\in\mathbb{R}} \left| \frac{1}{b}\sum_{i\in S} \mathbb{1}\{X_i \le x\} - G_{\hat{\theta}_b}(x) \right|,\ S \subset \{1,\dots,n\},\ |S| = b \right\}$

$= \hat{\theta}_b(X_i, i\in S)$

subsampling classique

→ $c_\alpha(\mathcal{X}_n)$ en prenant le $(1-\alpha)$-quantile de $L_n(\hat{Q}_n|\mathcal{P}_0)$ ou $L_{n,b}(\hat{P}_n|\mathcal{P}_0)$ (+ Monte Carlo)

→ pvalue $p(\mathcal{X}_n) = P(Z \ge T_n(\mathcal{X}_n))$ où $Z \sim L_n(\hat{Q}_n|\mathcal{P}_0)$ ou $L_{n,b}(\hat{P}_n|\mathcal{P}_0)$

théorie cf [Beran (1986) et Romano (1988)] bien de niveau asymptotiquement α si $\{Q_\theta, \theta\in\Theta\}$ régulier