Introduction au bootstrap

Résumé

Présentation succincte du principe du bootstrap.

Retour au plan du cours

1 Introduction

La motivation du *bootstrap* [1] (Efron, 1982 ; Efron et Tibshirani, 1993) est d'approcher par simulation (*Monte Carlo*) la distribution d'un estimateur lorsque l'on ne connaît pas la loi de l'échantillon ou, plus souvent lorsque l'on ne peut pas supposer qu'elle est gaussienne. L'objectif est de remplacer des hypothèses probabilistes pas toujours vérifiées ou même invérifiables par des simulations et donc beaucoup de calcul.

Le principe fondamental de cette technique de ré-échantillonnage est de substituer à la distribution de probabilité inconnue F, dont est issu l'échantillon d'apprentissage, la distribution empirique $\widehat{F}$ qui donne un poids $1/n$ à chaque réalisation. Ainsi on obtient un échantillon de taille n dit *échantillon bootstrap* selon la distribution empirique $\widehat{F}$ par n tirages aléatoires avec remise parmi les n observations initiales.

Il est facile de construire un grand nombre d'échantillons bootstrap sur lesquels calculer l'estimateur concerné. La loi simulée de cet estimateur est une approximation asymptotiquement convergente sous des hypothèses raisonnables [2] de la loi de l'estimateur. Cette approximation fournit ainsi des estimations du biais, de la variance, donc d'un risque quadratique, et même des intervalles de confiance de l'estimateur sans hypothèse (normalité) sur la vraie loi.

1. Cette appellation est inspirée du baron de Münchhausen (Rudolph Erich Raspe) qui se sortit de sables mouvants par traction sur ses *tirants de bottes*. En France "bootstrap" est parfois traduit par *à la Cyrano* (acte III, scène 13) en référence à ce héros qui prévoyait d'atteindre la lune en se plaçant sur une plaque de fer et en itérant le jet d'un aimant.
2. Échantillon indépendant de même loi et estimateur indépendant de l'ordre des observations.

1.1 Principe du *plug-in*

Soit $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ un échantillon de taille n issue d'une loi inconnue F sur $(\Omega, \mathcal{A})$. On appelle *loi empirique* $\widehat{F}$ la loi discrète des singletons $(x_1, \ldots, x_n)$ affectés des poids $1/n$:

$$\widehat{F} = \sum_{i=1}^{n} \delta_{x_i}.$$

Soit $A \in \mathcal{A}$, $P_F(A)$ est estimée par :

$$\widehat{(P)}_F(A) = P_{\widehat{F}}(A) = \sum_{i=1}^{n} \delta_{x_i}(A) = \frac{1}{n}\text{Card} x_i \in A.$$

De manière plus générale, soit θ un paramètre dont on suppose que c'est une fonction de la loi F. on écrit donc $\theta = t(F)$. Par exemple, $\mu = E(F)$ est un paramètre de F suivant ce modèle. Une *statistique* est une fonction (mesurable) de l'échantillon. Avec le même exemple :

$$\widehat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

et $\overline{x}$ est la statistique qui estime μ. On dit que c'est un estimateur "plug-in" et, plus généralement,

DÉFINITION 1. — *On appelle estimateur* plug-in *d'un paramètre θ de F, l'estimateur obtenu en remplaçant la loi F par la loi empirique :*

$$\widehat{\theta} = t(\widehat{F}).$$

comme dans le cas de l'estimation de μ : $\widehat{\mu} = E(\widehat{F}) = \overline{x}$.

1.2 Estimation de l'écart-type de la moyenne

Soit X une variable aléatoire réelle de loi F. On pose :

$$\mu_F = E_F(X), \quad \text{et} \quad \sigma_F^2 = \text{Var}_F(X) = E_F[(X - \mu_F)^2];$$

Ce qui s'écrit :

$$X \sim (\mu_F, \sigma_F^2).$$

Soit $(X_1, \ldots, X_n)$ n variables aléatoires i.i.d. suivant aussi la loi F. Posons $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Cette variable aléatoire a pour espérance μ_F et pour variance σ_F^2/n. On dit aussi que la statistique

$$\overline{X} \sim (\mu_F, \sigma_F^2/n).$$

Remarquons qu'en moyennant plusieurs valeurs ou observations, on réduit la variance inhérente à une observation. De plus, sous certaines conditions sur la loi F et comme résultat du théorème de la limite centrale, $\overline{X}$ converge en loi vers la loi normale.

L'estimateur plug-in de σ_F est défini par :

$$\begin{aligned} \widehat{\sigma}^2 &= \widehat{\sigma_F}^2 = \sigma_{\widehat{F}}^2 = \mathrm{Var}_{\widehat{F}}(X) \\ &= E_{\widehat{F}}[(X - E_{\widehat{F}}(X))^2] = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2. \end{aligned}$$

L'estimateur plug-in de σ_F est (légèrement) différent de celui du maximum de vraisemblance. L'estimateur plug-in est en général biaisé mais il a l'avantage d'être simple et de pouvoir s'appliquer à tout paramètre θ même lorsque l'on ne peut pas calculer la vraisemblance du modèle.

2 Estimation bootstrap d'un écart-type

Soit $\widehat{\theta} = s(x)$ un estimateur quelconque (M.V. ou autre) de θ pour un échantillon x donné. On cherche à apprécier la précision de $\widehat{\theta}$ et donc à estimer son écart-type.

2.1 Échantillon bootstrap

Avec les mêmes notations, $\widehat{F}$ est la distribution empirique d'un échantillon $\boldsymbol{x} = \{x_1, \ldots, x_n\}$.

DÉFINITION 2. — *On appelle* échantillon bootstrap *de $\boldsymbol{x}$ un échantillon de taille n noté*

$$\boldsymbol{x}^* = \{x_1^*, \ldots, x_n^*\}$$

suivant la loi $\widehat{F}$; $\boldsymbol{x}^*$ est un ré-échantillon de $\boldsymbol{x}$ avec remise.

2.2 Estimation d'un écart-type

DÉFINITION 3. — *On appelle* estimation bootstrap *de l'écart-type $\widehat{\sigma_F}(\widehat{\theta})$ de $\widehat{\theta}$, son estimation plug-in : $\sigma_{\widehat{F}}(\widehat{\theta})$.*

Mais, à part dans le cas très élémentaire où, comme dans l'exemple ci-dessus, θ est une moyenne, il n'y a pas de formule explicite de cet estimateur. Une approximation de l'estimateur bootstrap (ou plug-in) de l'écart-type de $\widehat{\theta}$ est obtenue par une simulation (Monte-Carlo) décrite dans l'algorithme ci-dessous.

Pour un paramètre θ et un échantillon $\boldsymbol{x}$ donnés, on note $\widehat{\theta} = s(\boldsymbol{x})$ l'estimation obtenue sur cet échantillon. Une *réplication bootstrap* de $\widehat{\theta}$ est donnée par : $\widehat{\theta}^* = s(\boldsymbol{x}^*)$.

ALGORITHME 1 : *Estimation de l'écart-type*

Soit $\boldsymbol{x}$ un échantillon et θ un paramètre.
for $b = 1$ à B ***do***
Sélectionner 1 échantillon bootstrap $\boldsymbol{x}^{*b} = \{x_1^{*b}, \ldots, x_n^{*b}\}$ par tirage avec remise dans x.
Estimer sur cet échantillon : $\widehat{\theta}^*(b) = s(\boldsymbol{x}^{*b})$.
end for
Calculer l'écart-type de l'échantillon ainsi construit :

$$\begin{aligned} \widehat{\sigma}_B^2 &= \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{\theta}^*(b) - \widehat{\theta}^*(.))^2 \\ \text{avec} \quad \widehat{\theta}^*(.) &= \frac{1}{B}\sum_{b=1}^{B}(\widehat{\theta}^*(b). \end{aligned}$$

$\widehat{\sigma}_B$ est l'approximation bootstrap de l'estimation plug-in recherchée de l'écart-type de $\widehat{\theta}$.

2.3 Estimation du biais

Avec les mêmes notations :

$$\theta = t(F) \quad \text{et} \quad \widehat{\theta} = s(\boldsymbol{x}),$$

le biais d'un estimateur s'exprime comme

$$\mathcal{B}_F(\widehat{\theta}) = E_F[s(\boldsymbol{x})] - t(F).$$

Un estimateur est sans biais si $E[\widehat{\theta}] = \theta$. Le biais est aussi une mesure de la précision d'un estimateur et on a vu que, généralement, les estimateurs plug-in étaient biaisés.

DÉFINITION 4. — *On appelle estimateur bootstrap du biais, l'estimateur plug-in :*

$$\widehat{\mathcal{B}_F}(\widehat{\theta}) = \mathcal{B}_{\widehat{F}}(\widehat{\theta}) = E_{\widehat{F}}[s(\boldsymbol{x}^*)] - t(\widehat{F}).$$

Comme pour l'écart-type, il n'existe généralement pas d'expression analytique et il faut avoir recours à une approximation par simulation.

ALGORITHME 2 *: Estimation bootstrap du biais*

Soit $\boldsymbol{x}$ un échantillon et θ un paramètre.
for $b = 1$ à B ***do***
Sélectionner 1 échantillon bootstrap $\boldsymbol{x}^{*b} = \{x_1^{*b}, \ldots, x_n^{*b}\}$. par tirage avec remise dans $\boldsymbol{x}$.
Estimer sur cet échantillon la réplication bootstrap de $\widehat{\theta}$: $\widehat{\theta}^*(b) = s(\boldsymbol{x}^{*b})$.
end for
Approcher $E_{\widehat{F}}[s(\boldsymbol{x}^*)]$ par $\widehat{\theta}^*(.) = \frac{1}{B}\sum_{b=1}^{B}(\widehat{\theta}^*(b)$
L'approximation bootstrap du biais est : $\widehat{\mathcal{B}_B}(\widehat{\theta}) = \widehat{\theta}^*(.) - \widehat{\theta}$.

---

3 Compléments

En résumé, on peut dire que le bootstrap repose sur une hypothèse très élémentaire : $\widehat{\theta}^*$ se comporte par rapport à $\widehat{\theta}$ comme $\widehat{\theta}$ par rapport à θ. La connaissance de $\widehat{\theta}^*$ (distribution, variance, biais…) renseigne alors sur celle de $\widehat{\theta}$.

Beaucoup d'autres compléments sont à rechercher dans la littérature et en particulier dans Efron et Tibshirani (1993). Il est ainsi possible de définir des intervalles de confiance bootstrap en considérant la distribution et les quantiles de $\widehat{\theta}^*$ ou même encore des tests à partir des versions bootstrap de leur statistique.

Le bootstrap rapidement décrit ici est dit "non-paramétrique" car la loi empirique $\widehat{F}$ est une estimation non-paramétrique de F. Dans le cas où F serait connue à un paramètre près, il existe également une version dite *paramétrique* du bootstrap.

Pour des estimateurs plus compliqués (fonctionnels) comme dans le cas de la régression non-paramétrique par noyau ou spline, il est facile de construire graphiquement une enveloppe bootstrap de l'estimateur à partir de réplications de l'échantillon. Celle-ci fournit généralement une bonne appréciation de la qualité de l'estimateur obtenu. Attention, dans le cas de la régression il est en principe plus justifié de répliquer le tirage sur les *résidus* plutôt que sur les observations. Ce sont les résidus qui sont en effet supposés i.i.d. et qui vérifient donc les hypothèses nécessaires mais cette approche devient très sensible à l'hypothèse sur la validité du modèle. Il est finalement d'usage de considérer un échantillon bootstrap issu des données initiales (Efron et Tibshirani) :

$$\mathbf{z}^{*b} = \{(\mathbf{x}_1^{*b}, y_1^{*b}), \ldots, (\mathbf{x}_n^{*b}, y_n^{*b})\};$$

c'est ce qui a été choisi dans ce document.

Enfin, l'estimation bootstrap est justifiée par des propriétés asymptotiques (convergence en loi) lorsque le nombre de réplications (B) croit conjointement avec la taille de l'échantillon (n). Comme la loi empirique $\widehat{F}$ converge (en loi) vers celle théorique, la distribution du paramètre $\widehat{\theta} = t(\widehat{F})$ converge (en loi) vers celle théorique de $\theta = t(\widehat{F})$.