Methodes de Bootstrap et applications actuarielles

Patrice BERTAIL

Laboratoire de Statistique, CREST-INSEE
et Université Paris X, Nanterre
FRANCE

Formation CARITAT, Paris le 12 mars 2007

Plan de l'exposé

Contexte et notations

Echantillon de n variables aléatoires $\mathbf{X}_n = (X_1, X_2, .., X_n)$ indépendantes identiquement distribuées (i.i.d.) de fonction de répartition commune F.

But : étudier les caractéristiques d'une statistique $T_n(\mathbf{X}_n)$ estimant un paramétre θ_F.

Problème :

- Est-il possible sans faire d'hypothèse sur la loi F ou des hypothèses minimales sur ses moments) de connaitre les performances de $T_n(\mathbf{X}_n)$ en terme de biais, de variance?
- Peut-on donner une approximation de sa distribution?

Finalité (historique) :

- Estimer la variance de statistique complexe
- Construire des intervalles de confiance pour des paramètres complexes.

Pas d'hypothèses fortes sur les lois sous-jacentes (méthode non-paramétrique) mais un domaine de validité bien connu

Principe de l'approximation

- Calcul exact : si T_n simple et la loi F sont spécifiées (Ex: moyenne dans le cas gaussien).
- Approximations asymptotiques (déterministes):
  - Distribution asymptotique limite de la statistique centrée correctement dilatée : pas toujours immédiat, ne rend pas toujours bien compte de la distribution à distance finie pour les petits échantillons. Par exemple, dans le cas de la moyenne, si F dissymétrique, ne rend pas compte de cette dissymétrie (erreur de l'ordre dissymétrie absolue $/\sqrt{n}$.
  - Développement d'Edgeworth (Feller (1971)) approximations au second ordre plus satisfaisantes mais sont souvent difficiles à mettre en oeuvre.
- Approximation stochastique (estimation) de la loi de T_n par méthodes de rééchantillonnage.

LE JACKKNIFE

Le principe du Jackknife: Tukey (1958): ôter de l'échantillon initial les observations à tour de rôle pour recalculer la valeur de la statistique (Gray, Schucany et Watkins (1972) et Miller (1974)).

De nombreux avantages pratiques :

- Détections de points aberrants,
- Calcul de la contribution de chacun des individus à la statistique (calcul empirique de la fonction d'influence)
- Calcul d'un estimateur de θ_F en moyennant ayant de meilleurs propriétés (correction du biais),
- Calcul d'un estimateur de la variance (sous la forme d'une variance empirique des contributions).
- Calcul de l'histogramme des n valeurs obtenues $\longrightarrow$ intervalles de confiance si n est grand et la statistique T_n simple (moments, fonction différentiable de moment).

Le bootstrap : une méthode de type plug-in, Efron (1979,1982a)

- Echantillon "Bootstrap" $(X_1^*, X_2^*, ..X_n^*)$ i.i.d. de distribution $\hat{F}_n$ estimant F. Interprétation: si cet estimateur $\hat{F}_n$ est proche (en un certain sens) de F, "on fait comme si" les observations X_i avaient comme loi $\hat{F}_n$ au lieu de F.
- Distribution Bootstrap et caractéristiques Bootstrap (variance, moments, fractiles bootstrap etc...) $=$ distribution (resp. caractéristiques) de la statistique $T_n(\mathbf{X}_n^*)$ sous la loi $\hat{F}_n$, connue, au lieu de la loi F.
- Distribution de $T_n(\mathbf{X}_n^*)$ sous la loi $\hat{F}_n$ est l'estimateur empirique naturel de la distribution de $T_n(\mathbf{X}_n)$ sous la loi F.

Question : **Dans quelle mesure la distribution de $\mathbf{T}_n(\mathbf{X}_n)$ sous $\hat{F}_n$ permet d'avoir une bonne idé e de la distribution de $T_n(\mathbf{X}_n)$ sous $\mathbf{F}$** i.e. de la vraie distribution.
Réponse : essentiellement un problème d'estimation. Le paramètre que l'on cherche à estimer est ici la distribution d'une statistique : est-ce que la méthode plug-in marche pour des distributions?

Quelques définitions

- Distribution de la statistique pour un échantillon de taille n $k_n(x, F) = \Pr_{F^{\otimes n}}\{T_n(Y_{1,\ldots,}Y_n) \leq x\},\ n \geq 1,$ distribution de la statistique T_n sous $F^{\otimes n}$.
- Distribution de la statistique pour un échantillon de taille m $k_m(x, F) = \Pr_{F^{\otimes m}}\{T_m(Y_{1,\ldots,}Y_m) \leq x\},\ m \geq 1,$ distribution de la statistique T_m sous $F^{\otimes m}$.
  Ce sont des paramètres (des fonctions) dépendants de F : une méthode d'estimation simple est simplement de "plugger" un estimateur de F dans cette expression.
- $k_n(x, \hat{F}_n)$ n'est rien d'autre que la distribution bootstrap.
- Si $m < n$, $k_m(x, \hat{F}_n))$ est la distribution bootstrap de taille m (m-out-n bootstrap). Validité universelle si $m^2/n- > 0$ (cf distribution de sous-échantillonnage).

Autant de méthodes de bootstrap que d'estimateurs de F

Cas particulier : le bootstrap naif

Si $\hat{F}_n = F_n$ fonction de répartition empirique des observations, alors l'échantillon bootstrap s'obtient simplement en tirant avec remise de manière équiprobable des valeurs dans l'échantillon initial.

$$\begin{aligned} k_m(.,F_n) &= \Pr_{F_n^{\otimes m}}\{T_m(Y_1,...,Y_m) \leq x\} \\ &= E_{F_n^{\otimes m}}\left(\mathbb{I}_{\{T_m(Y_1,...,Y_m)\leq x\}}\right) \\ &= n^{-m}\sum_{i_1=1}^{n}....\sum_{i_m=1}^{n}\mathbb{I}_{\left\{T_m(X_{i_1},...,X_{i_m})\leq x\right\}}, \end{aligned}$$

En particulier

$$k_n(.,F_n) = n^{-n}\sum_{i_1=1}^{n}....\sum_{i_n=1}^{n}\mathbb{I}_{\left\{T_n(X_{i_1},...,X_{i_n})\leq x\right\}},$$

Autant de méthodes de bootstrap que d'estimateurs de F

Cas particulier : le bootstrap paramétrique

Si le modèle est paramétrique $F_\theta, \theta \in \Theta$ et si $\hat{\theta}$ est un estimateur de θ par exemple l'estimateur du maximum de vraisemblance alors on peut prendre $\hat{F}_n = F_{\hat{\theta}}$ l'échantillon bootstrap s'obtient en tirant des valeurs dans la distribution estimée $F_{\hat{\theta}}$.

Exercice : $X_1, X_2, ..., X_n$ i.i.d. de loi normale $N(m, \sigma^2)$.

Déterminer une approximation de la loi de $CV = S_n^2/\overline{X}_n$, où S_n^2 est la variance empirique et $\overline{X}_n$ la moyenne empirique. Principe du bootstrap paramétrique : générer des observations de loi $N(\overline{X}_n, S_n^2)$.

Autant de méthodes de bootstrap que d'estimateurs de F

Cas particulier : le bootstrap lisse

Lorsque les estimateurs ou les variances dependent fortement de la densité alors $\hat{F}_n = F_n$ n'est pas un bon choix.
C'est le cas en particulier pour

- Les fractiles empiriques d'une distribution (sa variance depend de la densité)
- Un estimateur de la densité (par noyau ou ondelette)
- Des estimateurs multidimensionnels faisant intervenir une notion de profondeur (Estimateur de profondeur ou de longueur minimal).

Cas particulier : le bootstrap lisse (smooth bootstrap)

Meilleur choix : estimateur lissé de la fonction de répartition empirique (convolué de F_n avec un noyau).

$$\widehat{F}_n(x) = 1/n \sum_{i=1}^{n} K(((x - X_i)/h))$$

où K est une fonction de répartition connue centrée réduite (par exemple la loi normale centrée réduite.

Pratiquement : il suffit de tirer un échantillon bootstrap naif (i.i.d. avec remise dans les observations) et de les perturber par des v.a. gaussiennes indépendantes de loi $N(0, h^2)$ avec h petit : typiquement h de l'ordre de $1.049 * n^{-1/5}$*Ecart-type de l'échantillon.

Autres variations autour du bootstrap : le jackknife

Le bootstrap SANS REMISE de taille m : ou sous-échantillonnage

Si $\hat{F}_n = F_n$ fonction de répartition empirique des observations, alors l'échantillon bootstrap sans remise de taille m s'obtient simplement en tirant sans remise de manière équiprobable des valeurs dans l'échantillon initial.

$$K_m^U(x, F_n) = 1/C_n^m \sum_{i_1 \neq i_2, \ldots \neq i_m} \mathbb{I}_{\{T_m(X_{i_1,\ldots,}X_{i_m}) \leq x\}}$$

Cas particuliers :

- $m = n - 1 \rightarrow$ Histogramme du jackknife!
- $m = n - b_n \rightarrow$ Histogramme du " b_n-jackknife ou b_n-delete jackknife" =distribution de sous-échantillonnage de taille $n - b_n$.

Bootstrap et calcul de Monte-Carlo

Pas toujours possible de calculer explicitement la distribution ou les caractéristiques de $T_n(\mathbf{X}_n)$ sous la loi $\hat{F}_n$.

Bootstrap naif : n^n calculs. Bootstrap lisse ou bootstrap paramétrique en général pas d'expression explicite pour la distribution bootstrap.

Bootstrap et calcul de Monte-Carlo

Principe de la simulation :

- Itérer B fois
  - Génération d'un echantillon indépendants de loi $\hat{F}_n$ (rééchantillonnage)
  - Calcul sur l'échantillon tiré de la valeur de la statistique
- Utilisation des B valeurs pour construire une approximation de la distribution de $T_n(\mathbf{X}_n)$ sous $\hat{F}_n$.

Bien distinguer **deux niveaux**:

- **Un niveau méthodologique** : passage de F à la loi $\hat{F}_n$: méthode "plug-in" ou empirique.
- **Un niveau pratique : utiliser une technique de type Monte-Carlo** pour obtenir une approximation de la distribution Bootstrap. A ce titre la méthode du bootstrap est une méthode de calcul intensif.

Bootstrap naif et calcul de Monte-Carlo

On sélectionne de façon aléatoire équiprobable B échantillons, parmi l'ensemble des réalisations possibles de l'échantillon Bootstrap. Connaissant la réalisation $(x_1, x_2, ..x_n)$, on génère B échantillons:

$$\mathbf{x}_n^{*(b)} = (x_1^{*(b)}, x_2^{*(b)}, x_n^{*(b)})b = 1, , B$$

en tirant avec remise les $x_i^{*(b)}, i = 1, ..n$, dans l'ensemble $\{x_1, x_2, .., x_n\}$, puis on calcule $T_n(\mathbf{x}_n^{*(b)}), b = 1, ..B$ les valeurs de la statistique sur chacune des réalisations de l'échantillon Bootstrap obtenu.

Bootstrap naif et calcul de Monte-Carlo

L'histogramme des valeurs $T_n(\mathbf{x}_n^{*(b)}), b = 1, ..B$ donné par

$$K^*_{n,B}(u) = 1/B \sum_{b=1}^{B} 1_{(T_n(\mathbf{x}_n^{*(b)}) \leq x)},$$

donne une approximation de la distribution Bootstrap,

$$K(x, F_n) = P_{F_n}(T_n(X_n^*) \leq x).$$

On sait que dans tous cas les l'erreur commise par l'étape Monte-Carlo est de l'ordre $1/\sqrt{B}$. Indication sur le choix du nombre de rééchantillonnage nécessaire pour obtenir une précision donnée (B=10000 pour une précision à 1% sur toute la loi). Voir des indications plus précises dans le cas de la construction d'intervalles de confiance (9999 est meilleur...).

Bootstrap généralisé (pondéré) et plan de rééchantillonnage

- Plan de rééchantillonnage associé à $\mathbf{X}_n = (X_1, X_2, ., X_n) =$ vecteur de fréquences $\mathbf{W}_n^{(n)} = (W_1^{(n} \; W_2^{(} \;)W_3^{(n)}, \ldots ., W_n^{(n)})$, avec $\sum_{i=1}^n \; W_i^{(n)} = 1$ et $W_i^{(n)} \geq 0, \; i = 1, ., n,$
- Loi du plan de rééchantillonnage : $\mathbf{W}_n^{(n)}$ de loi $W^*(n)$ conditionnellement aux X_i (eventuellement dépendant des observations).

Bootstrap généralisé (pondéré) : terminologie

- Fonction de répartition empirique pondéré associée : $F_{n,W}(x) = \sum_{i=1}^{n} W_i^* I_{(X_i \leq x)}$.
- Echantillon bootstrap généralisé : si nW_i^* prend des valeurs entières, $(X_1^*, X_2^*, .., X_n^*)$ est l'échantillon dans lequel X_i apparaitrait nW_i^* fois.
- Statistique bootstrap généralisée : La "statistique" associée au plan de rééchantillonnage est définie à partir de la fonctionnelle $\theta(F)$ par $\theta(F_{n,W})$ où simplement par $T_n(\mathbf{X}^*)$.

Bootstrap généralisé (pondéré) : cas particuliers.

- Le Bootstrap naif : $n\mathbf{W}_n^*$ de loi multinomiale $Mult(n, 1/n)$.
- $P(W_1^*, = 1,\ \ W_i = 0, ...W_n^* = 1) = 1/n\ ,\ i = 1, ..n$: jackknife.
- $W_i = \gamma_i/(\sum_{j=1}^n \gamma_j)$, γ_i, i.i.d. $\gamma(1)$: bootstrap bayesien=distribution a posteriori pour un a priori de type Dirichlet pour la fonction de repartition (bayesien non-paramétrique).

Choix du meilleur type de bootstrap : souvent en fonction du probleme considéré... Va à l'encontre de l'idée selon laquelle le bootstrap est un outil universel.

Validité du Bootstrap : quid?

La conjecture du Bootstrap (Efron(1979)) : les statistiques étudiées sous la loi $\mathbf{F}_n$ connue et sous la loi inconnue $\mathbf{F}$ des observations ont un comportement analogue en terme de loi, de biais, de variance etc....

En d'autre termes, que la distribution de la statistique est robuste à une petite déviation par rapport à F et que l'on peut remplacer sans trop de problème la loi F par l' approximation F_n qui est très proche.

Validité du Bootstrap : quid?

On sera amené à comparer les comportements respectifs de $R(T_n, F) = T_n(\mathbf{X}_n) - \theta_F$ et de $R(T_n^*, F_n) = T_n(\mathbf{X}_n^*) - T(\mathbf{X}_n)$ ou de versions dilatées **(**resp. $n^{1/2}R(T_n, F)$ et $n^{1/2}R(T_n^*, F_n)$) évalués respectivement sous la loi F et sous la loi F_n , conditionellement à $X_1, \ldots X_n$.

Si ces quantités ont des comportements identiques en terme de loi asymptotique (à distance finie), on dit que **le Bootstrap fonctionne asymptotiquement (à distance finie).**

Validité du Bootstrap et méthode plug-in

La principale justification du Bootstrap naif repose sur les propriétés de la fonction de répartition empirique (voir Csörgö et Révész (1981), Shorack et Wellner (1986)).

Glivenko et Cantelli

$$||F_n - F||_\infty \to 0,\ p.s.$$

A t fixé, **Théorème central limite et fonctionnel** :

$$n^{1/2}(F_n(t) - F(t)) \underset{n\to\infty}{\overset{L}{\to}} N\left(0, F(t)(1-F(t)\right)$$

Théorème de Kolmogorov :

$$n^{1/2}||F_n - F||_\infty \underset{n\to\infty}{\overset{L}{\to}} K_F$$

K_F étant une variable aléatoire ne dépendant de F qui si cette dernière n'est pas continue.

Validité du Bootstrap et méthode plug-in

Ces propriétés expriment la proximité de la fonction de répartition empirique avec la vraie loi, pour la distance de Kolmogorov. D'autres résultats sont également disponibles pour d'autres métriques (Skorokhod, métriques de Zolotarev).

Question : est ce que les distributions de statistique sont robustes à un un changement local de fonction de répartition? En général, NON. Exemple les valeurs extrêmes (3 domaines d'attraction) $\longrightarrow$ explique l'echec du bootstrap naif dans ce cas. Idem pour les statistiques dont les distributions asymptotiques sont non continues en certain point : estimateur à seuil(Hodge-Lehmann), ou ayant des points d'hyperefficacité.

Validité du Bootstrap asymptotique et vitesse de convergence

Comment établir la validité du Bootstrap?

Proximité de type asymptotique : comparer la loi limite non dégéné rée de $n^{1/2}R(T_n, F)$ avec la loi limite conditionnelle à X_n de $n^{1/2}R(T_n^*, F_n)$ (eventuellement avec une autre standardisation). Si ces lois sont identiques, alors on peut se servir de la distribution Bootstrap comme d'une approximation de la vraie loi.

Avantage : la distribution bootstrap est facile à obtenir pratiquement, ce qui n'est pas toujours le cas de la distribution limite. Dans le cas gaussien pas très interessant (eventuellement pour avoir une approximation de la variance).

Validité du Bootstrap : le cas simple de la moyenne

Bickel et Freedmann(1981) ont montré à l'aide de la distance de Mallows que, dans le cas $\mathcal{R} = \mathbb{R}$, sous $E_F X_i^2 < +\infty$,

$$n^{1/2}(\overline{X}_n^* - \overline{X}_n) \xrightarrow{L.} N(0, \sigma^2) \; p.s.$$

$$S_n^{2*} \xrightarrow{pr.} \sigma^2, \; p.s.$$

d'oú l'on déduit

$$n^{1/2} S_n^{*-1/2}(\overline{X}_n^* - \overline{X}_n) \xrightarrow{L.} N(0, 1) \; p.s.$$

Giné et Zinn(1989) ont montré que la condition $E_F X^2 < \infty$ est une condition nécessaire et suffisante pour la validité du bootstrap p.s. en $\mathbf{X}_n$.

Attention au bootstrap quand il n'existe pas des moments au moins d'ordre 2 : dans ce cas le bootstrap n'est pas valide en particulier pour des lois de Cauchy ou des lois de Pareto avec index trop petit.

Validité au second ordre du Bootstrap et vitesse de convergence

Etude de la proximité: en terme d'ecart entre la vrai distribution de la statistique et la distribution bootstrap.

Pour cela on étudie la vitesse de convergence d'une distance entre les deux distributions. Les métriques les plus utilisées sont la distance de Kolmogorov et les distances de Mallows (ou distance de Wasserstein, cf Bickel et Freedman(1981)).

Avantage : sous la condition d'existence de moments d'ordre 4 et si la distribution des variables aléatoires n'est pas réticulé, la distribution Bootstrap **centrée réduite** est plus proche de la vraie distribution que la distribution asymptotique.

Comparaison des vitesses : asymptotique et bootstrap

- L'écart entre la vraie distribution et la distribution asymptotique : $O(n^{-1/2})$ (Théorème de Berry-Esséen).
- L'écart entre la vraie distribution et la distribution Bootstrap: $O_P(n^{-1})$: on parle de propriétés au second ordre.
- pour des statistiques symétriques (exemple $n^{1/2}|\overline{X}_n - \theta|$), on peut même avoir une vitesse de l'ordre $O_P(n^{-2})$.

Validité au second ordre du Bootstrap et vitesse de convergence

On explique ce comportement par le fait que la distribution Bootstrap prend en compte, de manière empirique, l'assymétrie (voire l'applatissement) de la distribution sous-jacente, ce que ne fait pas la distribution asymptotique.

Un exemple : si $X_1, X_2, ..., X_n$, $n = 49$ i.i.d. de $\gamma(1)$ et on s'interesse à la moyenne. L'erreur commise en utilisant l'approximation asymptotique gaussienne est supérieure à coefficient d'asymétrie absolu $/\sqrt{n} = 2/7 \approx 0.28$. Ce qui signifie que en construisant un intervalle de confiance a 95% en prenant une gaussienne, le niveau reel est plutot 66%... Avec le bootstrap, si on bootstrappe la moyenne standardisé par son ecart-type estimé, l'erreur est de l'ordre $1/n \simeq 1/49 \simeq 0.02$.

Une amélioration de l'approximation asymptotique : les développement d'Edgeworth(1907), Feller (1971), Chibishov(1972)

Le développement d'Edgeworth de la moyenne : si on pose $E_F(X - E_F X)^2 = 1$ et $\mu_3 = E_F(X - E_F X)^3 < +\infty$ pour le coéfficient d'assymétrie et si

- **Condition de Cramer** $\overline{\lim_{t\to\infty}} |E_F \exp(itX)| < \infty$
- **Condition de moments** $E_F(X)^4 < +\infty$

alors on a un développement de la moyenne empirique de la forme

$$P_F(n^{1/2}(\overline{X}_n - \mu) \leq x) = \Phi(x) - \frac{1}{6}\mu_3 n^{-1/2}(x^2 - 1)\phi(x) + O(n^{-1})$$

uniformément en x où ϕ et Φ respectivement la densité et la fonction de répartition de la loi normale.

Une amélioration de l'approximation asymptotique : les développement d'Edgeworth

Plus généralement si R_n est une statistique standardisée asymptotiquement gaussienne, sous conditions d'existence de moments, il existe des polynômes $p_i(x, F), i = 1, .., L$, et une fonction

$$F_{E_L}^{(n)}(x) = \Phi(x) + \sum_{i=1}^{L} n^{-i/2} p_i(x, F)\phi(x)$$

tels que, uniformement en x, on ait

$$P_F(R_n) = F_{E_L}^{(n)}(x) + o(n^{-L/2}).$$

$F_{E_L}^{(n)}(x)$ est **le développement d'Edgeworth d'ordre L** $(\mathbf{DE}_L)$ de R_n.

Développements d'Edgeworth et correction au second ordre

Abramovitch et Singh (1985) montrent que, si $\hat{p}_1(x)$ est un estimateur de $p_1(x, F)$ tel que pour tout $\varepsilon > 0$ $P_F\{|\hat{p}_1(R_n) - p_1(R_n, F)| > \epsilon\} = O(n^{-1})$ alors la statistique $L_n(\mathbf{X_n})$ définie par $L_n(\mathbf{X}_n) = R_n - n^{-1/2}\hat{p}_1(R_n)$ admet le DE1 uniforme en x :

$$P_F(L_n(\mathbf{X}_n) \leq x) = \Phi(x) + O(n^{-1}).$$

$x - n^{-1/2}\hat{p}_1(x)$ est une transformation normalisante (au sens de Fisher(1925)) dans la mesure où l'on corrige de la dissymétrie du problème. On dit alors que la **statistique** $L_n(\mathbf{X}_n)$ est **correcte au second ordre**.

Distribution Bootstrap et développement d'Edgeworth

Freedman(1980) et Singh(1981). Si F est suffisamment régulière (condition de Cramer) et admet des moments d'ordre au moins 3, la distribution Bootstrap standardisée est équivalente à un dé veloppement d'Edgeworth D.E. d'ordre 1 et tient donc compte de la dissymétrie de la distribution à distance finie de la moyenne. La distribution Bootstrap s'interprète comme un développement d'Edgeworth empirique i.e. la moyenne Bootstrap admet comme D.E. celui de la moyenne, dans lequel les moments exacts sont remplacés par les moments empiriques.

$$n^{1/2}||P_{F_n}(n^{1/2}(\overline{X}_n^* - \overline{X}_n)/S_n) - \Phi(x) + \tfrac{1}{6}\hat{\mu}_3 n^{-1/2}(x^2 - 1)\phi(x)||_\infty \xrightarrow{\Pr} 0$$

où $\hat{\mu}_3 = n^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^3 / \left(n^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2\right)^{3/2}$ est le coefficient d'assymétrie empirique. D'où si $E_F|X_i|^3 < \infty$,

$n^{1/2}||P_{F_n}(n^{1/2}(\overline{X}_n^* - \overline{X}_n)/S_n) - P_F(n^{1/2}(\overline{X}_n - \mu)/\sigma)||_\infty \xrightarrow{\Pr} 0.$

Si de plus $E_F X_i^6 < \infty$ alors

$$n \log\log(n)^{-1/2}||P_{F_n}(n^{1/2}(\overline{X}_n^* - \overline{X}_n)/S_n) - P_F(n^{1/2}(\overline{X}_n - \mu)/\sigma)||_\infty = H_F$$

où H_F est une constante dépendant des moments d'ordre 6.

Généralisation à des statistiques plus générales

Ces résultats de Singh(1981) ont été généralisés par Athreya(1983), Babu et Singh(1984a), Bhattacharya(1987), Hall(1992) à des statistiques s'exprimant comme des fonctions (trois fois différentiables) de la moyenne et par suite comme des fonctions de moment de F, puis pour des X_i à valeurs dans un Banach (Götze(1989)).

Weng(1989) obtient le D.E. du Bootstrap bayesien, Barbe et Bertail(1994) des conditions de validité du Bootstrap généralisé.

Généralisation à des statistiques plus générales Bhattacharya et Qumsiyeh(1989) montrent que le D.E. de la distribution Bootstrap est la version empirique du D.E. de la distribution exacte, est valide à tout ordre $o(n^{-L/2})$ sous conditions d'existence des moments d'ordre $(L+2)^2$, presque sûrement, uniformément en x, mais est aussi valide en p-moyenne, à savoir pour la norme $(E(.)^p)^{1/p}$, uniformément en x.
A x fixé, la distribution Bootstrap est pour cette norme toujours meilleure que le D.E.

Généralisation : le processus empirique

Bickel et Freedman (1981) ont établi pour l'échantillon Bootstrap dans le cas $\mathcal{R} = \mathbb{R}$ un résultat analogue à la convergence du processus empirique vers le mouvement brownien $B(F)$
soit $n^{1/2}(F_n^* - F_n) \underset{n\to\infty}{\overset{w}{\to}} B(F),\ a.s.$.
Mason et Newton(1991) ont obtenu un résultat analogue pour le Bootstrap généralisé.

Généralisation : le processus empirique

Applications :

- Construction de bandes de confiance pour F.
- Les tests bootstrap de type Kolmogorov-Smirnov sont valides.
- Si on peut établir une convergence asymptotique avec le processus empirique alors sous des conditions très faibles, le bootstrap est asymptotiquement valide (Csörgö et Mason (1989)).
- en particulier il est possible de bootstrapper tous les M-estimateurs classiques et la plupart des R-L-estimateurs.

Généralisation : le processus fractile

Si F est continuement dérivable et si f désigne la densité associée supposée strictement positive, le processus fractile, $Z_n = n^{1/2}(F_n^{-1} - F^{-1})$, converge vers $B(F^{-1})/f \circ F^{-1}(.)$ sur tout intervalle $[t_0, t_1], (t_0, t_1) \in]0,1[^2$,
On a alors

$$||P_F\{n^{1/2}(F_n^{-1}(t) - F^{-1}(t)) \leq x\} - P_{F_n}\{n^{1/2}(F_n^{*-1}(t) - F_n^{-1}(t)) \leq x \mid \mathbf{X}$$

de l'ordre de $n^{1/4}(loglogn)^{-1/2}$ p.s..
L'ordre de la différence est en proba $O(n^{-1/4})$ alors qu'il ést de $O(n^{-1/2})$ pour la distribution asymptotique!!!

Généralisation : le processus fractile

NE JAMAIS UTILISER LE BOOSTRAP NAIF POUR UN FRACTILE (cf méthode de provisionnement).

Lui préférer le BOOTSTRAP LISSE : ordre de l'approximation $O(n^{-3/4})$.

Généralisation : fonctionnelles différentiables

Le parametre d'intérêt est une fonction de F : $T(F)$ défini sur un espace convexe contenant au moins les fonctions de répartion discrete muni d'une métrique D.

- Differentiabilité: il existe une fonction $T^{(1)}(x, F)$ (la fonction d'influence) vérifiant $E_F T^{(1)}(X, F) = 0$ telle que :

$$T(G) - T(F) = \int T^{(1)}(x, F) d(G - F)(x) + o(D(G, F))$$

- Développement stochastique (Serfling(1981))

$$n^{1/2}(T(F_n) - T(F)) = n^{-1/2} \sum_{i=1}^{n} T^{(1)}(X_i, F) + n^{1/2} o(D(F_n, F)).$$

- Idée : approcher $T(F_n) - T(F)$ par sa partie linéaire = "delta-méthode fonctionnelle" = outil fondamental de la robustesse (voir Huber(1981)).
- Version Bootstrap :

$$n^{1/2}(T(F_n^*) - T(F_n)) = n^{-1/2} \sum_{i=1}^{n} \psi_T(X_i^*, F) - n^{-1/2} \sum_{i=1}^{n} \psi_T(X_i, F)$$

$$+ R_n^*.$$

- Validité du Bootstrap s'en déduit par application du résultat de validité de la moyenne et du processus empirique.

Beran obtient des résultats d'optimalité minimax de la distribution Bootstrap pour des fonctionnelles statistiques deux fois différentiables sur l'espace des fonctions de répartition à support compact et admettant un développement d'Edgeworth possédant des propriétés d'uniformité en F.

Conclusion : pour des fonctionnelles régulières, suffisament différentiables, la distribution bootstrap est au sens minimax le meilleur estimateur régulier possible d'une distribution d'une statistique empirique dans un cadre i.i.d.

Validité du bootstrap : les modèles économétriques

Généralisation du Bootstrap : le modèle de corrélation

Si l'on considère un modèle non conditionnel (dit modèle de corrélation (Freedman(1981)) aux variables explicatives:
$Y = X\beta + \epsilon$ avec $E_F\epsilon = 0$ et $V_F\epsilon = \sigma^2 I_n$
où $Y = (Y_i)_{1\leq i\leq n}$, $X = (X_i^{'})_{1\leq i\leq n}$ et $\epsilon = (\epsilon_i)_{1\leq i\leq n}$
où les $(Y_i, X_i^{'})$ sont des variables aléatoires i.i.d de loi multidimensionnelle F tels que $E(X_i'\epsilon_i) = 0$ alors β admet une forme fonctionnelle $\beta = (E_F X'X)^{-1} E_F X\prime Y$ et les résultats sur le cas i.i.d s'appliquent directement au couple $(Y_i, X_i^{'}$ (voir Freedman(1981)).

Application : Rééchantillonnage de couples d'observations $(Y_i, X_i^{'}), i = 1, ..., n$ pour conserver la structure de la corrélation.

Généralisation du Bootstrap: le modèle de régression

Hypothèses conditionnelles : ϵ_i sont i.i.d. de loi F inconnue et que la matrice X est supposée déterministe ou modèle conditionnel aux variables explicatives avec $E_F(\epsilon/X) = 0$ et $V_F(\epsilon/X) = \sigma^2 I_n$).

Remarques : les variables aléatoires Y_i ne sont pas i.i.d.

Estimateur des m.c.o: $\hat{\beta}_n = (X'X)^{-1}X\prime Y = \beta + (X'X)^{-1}X'\epsilon$

L'idée de base (Freedman(1981)) : rééchantillonner des résidus estimés centrés.

Principe : construire un estimateur de la loi des résidus et se ramener à du rééchantillonnage i.i.d.

Généralisation du Bootstrap: le modèle de régression

- Etape estimation Résidus estimés du modèle $\hat{\epsilon} = Y - X\hat{\beta}_n$, Résidus centrés: $\tilde{\epsilon} = (\tilde{\epsilon}_i)_{i=1,.n}$ $\tilde{\epsilon}_i = \hat{\epsilon}_i - n^{-1}\sum_{i=1}^n \sum_{i=1}^n \hat{\epsilon}_i$, Fonction de répartition empirique des résidus estimés centré

$$\tilde{F}_n(u) = \frac{1}{n}\sum_{i=1}^{n} h_{\tilde{\epsilon}_i}(u)$$

.

- Phase de rééchantillonnage $\epsilon_i^*, i = 1, .., n$, i.i.d. de loi $\tilde{F}_n$ conditionnellement à (Y, X).
- Pseudo modèle Bootstrap

$$y_i^* = x_i'\hat{\beta}_n + \epsilon_i^*,$$

Généralisation du Bootstrap: le modèle de régression Relève de la catégorie Bootstrap semiparamétrique

La version bootstrap est du même type que le modèle initial i.e. de la forme $F_{(\hat{\beta},\tilde{F}_n)}$ avec

$$E_{\tilde{F}_n}\epsilon_i^* = 0 \qquad V_{\tilde{F}_n}\epsilon_i^* = \frac{1}{n}\sum_{i=1}^{n}\tilde{\epsilon}_i^2 \underset{n\to\infty}{\overset{pr}{\longrightarrow}} \sigma^2.$$

Analogues Bootstrap des estimateurs des m.c.o. et de la variance résiduelle

$$\hat{\beta}_n^* = \left(X'X\right)^{-1} X'Y^* = \hat{\beta}_n + \left(X'X\right)^{-1} X'\epsilon^*$$

$$\hat{\sigma}_n^{*2} = \frac{1}{n}\sum_{i=1}^{n}(y_i^* - x_i'\hat{\beta}_n^*)^2$$

Validité asymptotique et au second ordre : le modèle de régression

- Freedman (1981) montre que cette procédure fonctionne asymptotiquement, sous condition d'existence de moments d'ordre au moins 4 pour la loi des résidus.
- Robinson(1987), Hall(1988) établissent la validité au second ordre de la procédure sous conditions de moments d'ordre au moins 6.
- Le principal argument : proximité de $\tilde{F}_n$ avec F_n et de F_n avec F, en terme de distance de Mallows.
- TRES IMPORTANT: le recentrage des résidus crucial pour obtenir la proximité de $\tilde{F}_n$ avec F et la validité du Bootstrap. Centrage automatique si la constante est inclut dans le modèle.

Gérénralisation à divers modèles économétriques

- Shorack(1982) : M-estimateurs dans la régression robuste,
- Staniewski(1985) : l'estimateur des moindre-carrés non linéaires dans la régression non-linéaire,
- Härdle et Bowman(1988), Härdle et Marron(1990), Hall(1992) : régression non paramétrique.
- Freedman(1984): estimateurs des doubles et triples moindres carrés dans des modèles à équations simultanées.

UN SEUL ET MÊME PRINCIPE : ESTIMATION DU MODELE DE BASE, ESTIMATION DE LA LOI DES RESIDUS QUI VERIFIENT LE MODELE DE BASE, PSEUDO-MODELE BOOTSTRAP, REESTIMATION (rééchantillonnage) DANS LE MONDE BOOTSTRAP

Intervalle de confiance exact

$T_n(\mathbf{X}_n)$ un estimateur du paramètre réel θ_F, $T_n(\mathbf{X}_n^*)$ la statistique Bootstrap.

$G_n(u)$ (resp. $G_n^*(u)$) : distribution à distance finie de $T_n(\mathbf{X}_n)$ (resp. $T_n(\mathbf{X}_n^*)$ conditionnellement à $\mathbf{X}_n$).

Quantile d'ordre γ de G_n (resp.de G_n^*) est défini par $g_n(\gamma) = G_n^{-1}(\gamma)$ resp. $g_n^*(\gamma) = G_n^{*-1}(\gamma)$.

Pour un niveau α donné, si la fonction de répartition G_n connue, intervalle de confiance pour le paramètre θ_F, exact, de niveau $1-\alpha$

$$I_n[\alpha] = [g_n(\alpha/2), g_n(1-\alpha/2)].$$

La méthode percentile

Intervalle construit avec les quantiles de la loi Bootstrap (qui est un estimateur de la vraie loie):
Intervalle percentile :

$$PI_n[\alpha] = [g_n^*(\alpha/2), g_n^*(1-\alpha/2)].$$

Si le Bootstrap fonctionne asymptotiquement alors,
quand $(n \to \infty)$, $PI_n[\alpha]$ est un intervalle de confiance asymptotiquement de niveau $1-\alpha$ (voir Beran(1984a), Beran et Millar(1986) pour une généralisation au cas multimensionnel).
Avantage sur la méthode asymptotique : pas besoin d'estimateur de la la variance ou de la loi asymptotique.

La méthode t-percentile

Idée de la méthode t-percentile : utiliser les quantiles de la statistique Bootstrap convenablement standardisée: $\tilde{R}_n(T_n^*, F_n)$, analogue de la quantité $\tilde{R}_n(T_n, F)$, de manière à ce que celle-ci soit asymptotiquement pivotale.
$\tilde{R}_n(T_n, F)$: appelé pivot ou racine de l'intervalle de confiance.
Exemple de la moyenne : la racine est $\tilde{R}_n(T_n, F) = (T_n(\mathbf{X}_n) - \theta_F)/S_n(X_n)$ où $S_n(\mathbf{X}_n)$ est un estimateur de la variance de $T(\mathbf{X_n})$.
Analogue Bootstrap est $\tilde{R}_n(T_n^*, F_n) = (T_n(\mathbf{X}_n^*) - T_n(\mathbf{X}_n))/S_n(X_n^*)$ définie conditionnellement à $(X_1, X_2, .., X_n)$.

Avantage de La méthode t-percentile

- Automatiquement correct au second ordre $O_P(n^{-1})$ pour des intervalles de confiance unilatéraux, voire correct au 3ème ordre pour des intervalles de confiances bilatéraux $O_P(n^{-2})$
- Tiens compte automatiquement des phénomènes de dissymétries
- Pas besoin de correction de biais et d'accélération (coeff. d'assymétrie).

Inconvénient par rapport à la méthode percentile

- Nécessité d'avoir un bon estimateur (de faible biais $O_P(n^{-1})$) de la variance.
- Théoriquement, possibilité d'utiliser un estimateur bootstrap de la variance (et donc du double bootstrap) mais souvent décevant en pratique.

Méthodes percentiles corrigés du biais et accéléré

- But : obtenir a partir d'une méthode percentile simple un intervalle correct au second ordre.
- Outils : inverser le développement d'Edgeworth de la statistique non-standardisée
- Necessité d'estimer le biais (facile avec la distribution bootstrap) et le coefficient d'assymétrie (plus difficile dans la plupart des situations)
- En général, plus simple et plus direct de construire la méthode t-percentile.

Choix du nombre de rééchantillonnages

Construction et résultats sur les intervalles de confiance théorique dans la mesure où ils sont construits avec les distributions Bootstrap exactes et non sur leur approximation .

Problême du choix de rééchantillonnages B nécessaire pour obtenir une précision suffisante sur les bornes de l'intervalle de confiance mais aussi sur la couverture effectivement atteinte.

Choix du nombre de rééchantillonnages

Optique conservative : impact de la phase Monte-Carlo $O_P(1/\sqrt{(}B))$.

Si on utilise une méthode percentile alors on ne peut avoir mieux que $O_P(1/\sqrt{(}n))$ et donc dans ce cas B¿¿n suffit pour conserver les propriétés du Bootstrap.

Par contre si on utilise une méthode t-percentile, avec une erreur $O_P(1/n))$ alors il faut au moins $B = n^2$ simulations pour avoir une précision correcte à la fois sur la borne et la couverture de IC et $B = n^4$ si on considère des IC bilateraux.

Un résultat étonnant de Hall(1986) dans le cas de la méthode t-percentile.

- Erreur commise sur le niveau de l'intervalle construit par la méthode t-percentile après rééchantillonnage est de l'ordre de B^{-1}
- Si B est tel que, pour un niveau $1-\alpha$ désiré, $(B+1)(1-\alpha)$ est entier alors l'erreur commise n'est plus que de l'ordre de $(nB)^{-1}$.
- En pratique si $\alpha = 5\%$ ou 1% alors on doit prendre $B = 99$, 999, 9999 ou un nombre de ce type.
- Donne une bonne borne sur la couverture mais pas sur le quantile

les echecs du Bootstrap

De nombreuses situations dans lesquelles le bootstrap ne fonctionne pas même asymptotiquement.

- cas des variables aléatoires ayant des queues de distributions trop épaisses (pour lesquels le moments d'ordre 2) n'existe pas,
- Les statistiques faisant intervenir des valeurs extrèmes,
- les U ou V statistiques dégénérées (du type statistique de Wilcoxon sous H_0) dont le comportement asymptotique n'est pas gaussien.
- Le bootstrap ne fonctionne pas en tout point de l'espace des paramètres où la distribution limite présente une discontinuité(Estimateur de Hodge-Lehmann, Stein, ondelettes, phénomènes hyperefficacité en certains points).

Les solutions aux echecs

- Une méthode Bootstrap adapté au problême condidéré. Parfois un recentrage adéquat au point problématique suffit. Exemple : $(X_1, ..., X_n)$ i.i.d. F, validité du bootstrap de $(\overline{X}_n)^2$?.
- Une méthode bootstrap par troncature (eliminer ou tenir compte précisément des points ou ca ne fonctionne pas): nécessité d'avoir une bonne connaissance du phénomène aux points d'hyperefficacité.
- Une solution universelle pour des tailles d'échantillons n assez grande : le sous-échantillonage ou bootstrap sans remise.

Validité asymptotique du sous échantillonnage

H_0 : il existe une vitesse τ_n, telle que
$\tau_n \left(T_n(X_{1,\ldots,}X_n) - \theta(P)\right)$ converge vers une distribution K_P quand $n \to \infty$.
H_1 : K_P est non-dégénérée et continue (au moins dans un voisinage d'un point dont on cherche les quantiles).
H_2 : $\frac{m}{n} \to 0$ et $\frac{\tau_m}{\tau_n} \to 0$ alors

$$||K_m^U(.,F_n) - K_n(.,F)||_\infty \underset{n\to\infty}{\to} 0 \text{ Pr}.$$

Si de plus $\frac{m}{n^{1/2}} \to 0$ alors le m out of n bootstrap (avec remise) est aussi universellement consistant.
Choix optimal $m = n^{2/3}$ sans correction de population finie.

Validité asymptotique du sous échantillonnage

- Necessite la connaissance de la vitesse de convergence pour construire la distribution de sous échantillonnage.
- Fonctionne même si cette vitesse est estimée (possible avec la distribution de sous-echantillonage), Bertail et al. (1999)
- Fonctionne en particulier pour les valeurs extrêmes, Bertail et al. (2004).
- Pas de validité au second ordre possible sans extrapolation...
- En pratique le choix optimal de m est un vrai casse tête...

En pratique : le cas des extrêmes

- construire la distribution de sous-echantillonnage sans standardisation i.e par exemple directement la distribution du max pour une succession de valeurs de m (typiquement entre $m^{1/4}$ et $m^{2/3}$
- représenter un ou plusieurs quantiles de la distribution de sous echantillonnage en fonction de m (ex. quantile à 75% ou 90%)
- On constate a partir d'un certain m que le quantile devient très volatile : prendre la valeur juste en dessous.
- si on connait la forme de la vitesse de convergence, l'utiliser directement. Sinon l'estimer et reconstruire la distribution de sous echantillonnage avec cette standardisation estimée.

cf Bertail, Politis, White (2004) pour des applications aux extremes et à la VAR en finance.