

# Chapitre IV

## Simulations et bootstrap

### 1 Les simulations

#### 1.1 Pourquoi faire des simulations ?

Une partie des résultats que l'on a obtenus jusqu'ici sont de type asymptotique mais en pratique on va les utiliser pour un nombre donné  $n$  (éventuellement grand mais jamais infini) d'observations. Chaque fois que l'on démontre un résultat de normalité asymptotique tel que  $\sqrt{n}(Z_n - z) \rightsquigarrow \mathcal{N}(0, \sigma^2)$  et que l'on remplace la f.r. de  $Z_n$ ,

$$\mathbb{P}[Z_n \leq t] = \mathbb{P}[\sigma^{-1}\sqrt{n}(Z_n - z) \leq \sigma^{-1}\sqrt{n}(t - z)],$$

que l'on ne sait pas calculer, par son équivalent asymptotique  $\Phi(\sigma^{-1}\sqrt{n}(t - z))$  où  $\Phi$  désigne la f.r. gaussienne standard, on fait une approximation dont on ignore la qualité pour une valeur donnée de  $n$ . Ceci s'applique en particulier au T.L.C. ordinaire. Pour avoir une idée de l'erreur que l'on commet en remplaçant  $\mathbb{P}[Z_n \leq t]$  par  $\Phi(\sigma^{-1}\sqrt{n}(t - z))$ , on peut recourir à une simulation. Le même problème se pose en fait déjà avec la loi des grands nombres lorsque l'on remplace  $n^{-1} \sum_{i=1}^n X_i$  par  $\mathbb{E}[X]$ .

En Statistique, on fait une simulation parce que l'on a fabriqué une procédure dont on n'est pas capable d'évaluer précisément les performances non-asymptotiques. On connaît la loi limite de telle suite d'estimateurs ou de statistiques de test mais on ignore si, avec le nombre  $n$  d'observations dont on dispose, on est proche de cette loi limite. De manière plus générale, on peut souhaiter évaluer la loi d'une variable ou la probabilité d'un événement dans une situation où l'on est incapable de mener à bien les calculs. Une simulation numérique permet souvent de résoudre le problème.

#### 1.2 Comment faire des simulations ?

Supposons que l'objet d'intérêt est une statistique  $T = T(X_1, \dots, X_n)$  fonction de variables i.i.d.  $X_i$ ,  $1 \leq i \leq n$  où  $n$  est fixé mais pas nécessairement très grand et que l'on souhaite obtenir des informations sur la loi  $P_T$  de  $T$  lorsque les  $X_i$  ont une fonction de répartition  $F$  donnée. Par exemple si les  $X_i$  sont des v.a. de Bernoulli de paramètre  $\theta$  et  $\hat{\theta}_n = n^{-1} \sum_{i=1}^n X_i$ , on sait que  $\sqrt{n}(\hat{\theta}_n - \theta)[\theta(1 - \theta)]^{-1/2} \mapsto \mathcal{N}(0; 1)$ . Donc, d'après le Théorème de Slutsky,  $T_n = \sqrt{n}(\hat{\theta}_n - \theta)[\hat{\theta}_n(1 - \hat{\theta}_n)]^{-1/2}$  converge vers une loi gaussienne standard quand  $n \rightarrow +\infty$ . Il est alors légitime, lorsque  $n$  n'est pas très grand, de se demander si la loi de  $T_n$  est proche d'une loi gaussienne standard. Comme cette loi dépend non seulement de  $n$  mais aussi de  $\theta$ , on s'efforcera de résoudre le problème pour différentes valeurs de  $\theta$ , en particulier 1/2, 1/4 et, à l'opposé, des valeurs proches de zéro.

Pour simuler la loi  $P_T$  d'une statistique  $T = T(X_1, \dots, X_n)$  quand les  $X_i$  sont i.i.d. de fonction de répartition  $F$ , on commence par générer  $nk$  variables indépendantes  $Y_{i,j}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq k$  de fonction de répartition  $F$ , où  $k$ , étant choisi par le statisticien, peut être très grand (la limitation n'est que celle des capacités de calcul). On verra ci-après comment faire cela. En posant  $T_j = T(Y_{1,j}, \dots, Y_{n,j})$ ,  $1 \leq j \leq k$ , on construit ainsi un  $k$ -échantillon  $T_1, \dots, T_k$  de loi  $P_T$  et l'on désigne par  $\nu_k$  sa loi empirique.

**Exemple 1 : Probabilité d'un évènement.** Pour évaluer la probabilité  $P_T(A)$  d'un évènement  $A$ , il suffit de calculer sa fréquence empirique  $\nu_k(A) = k^{-1} \sum_{j=1}^k \mathbb{1}_A(T_j)$ . L'Inégalité de Hoeffding binômiale (I.2.3) nous dit alors que

$$\mathbb{P}[|P_T(A) - \nu_k(A)| \geq x] \leq 2 \exp[-2kx^2],$$

qui peut être rendu arbitrairement petit par un bon choix de  $k$ . En prenant  $k = 10^6$  et  $x = 2 \times 10^{-3}$ , cette probabilité est majorée par  $2e^{-8} < 7 \times 10^{-4}$ . Donc, avec une probabilité très voisine de un, la fréquence empirique de  $A$  fournit une excellente approximation de  $P_T(A)$ .

**Exemple 2 : Évaluation de la loi de  $T$ .** Pour connaître la loi de  $T$ , il suffit de connaître sa fonction de répartition  $F_T$ . Cette fois-ci, on utilisera l'Inégalité de Massart du Théorème 5 pour les déviations de la fonction de répartition empirique :

$$\mathbb{P}\left[\sup_t |F_T(t) - F_k(t)| \geq x/\sqrt{k}\right] \leq 2 \exp[-2x^2],$$

où  $F_k$  est la fonction de répartition empirique des  $T_j$ . En particulier, avec une probabilité très voisine de un ( $> 0,9993$ ),  $\sup_t |F_T(t) - F_k(t)| \leq 2/\sqrt{k}$ . Pour  $k = 40000$ , valeur tout à fait banale pour les machines actuelles, on obtient une approximation de  $F_T$  à  $10^{-2}$  près et l'on peut ainsi vérifier, par exemple, si la loi de  $T$  est proche de la loi gaussienne standard.

### 1.3 Génération de variables aléatoires de loi donnée

Comme on l'a vu, pour faire une simulation, on a besoin de générer (le plus souvent en grand nombre) des variables aléatoires i.i.d. de loi donnée, c'est à dire de f.r.  $F$  donnée. Pour ce faire, on peut s'appuyer sur le résultat élémentaire suivant qui fournit une méthode générale pour construire des variables aléatoires de loi donnée sur  $\mathbb{R}$  par inversion de la f.r.  $F$ .

**Lemme 3** Soit  $U$  une variable aléatoire uniforme sur  $[0; 1]$  et  $F$  une fonction de répartition sur  $\mathbb{R}$ , alors la v.a.  $X = F^{-1}(U)$  (où  $F^{-1}$  désigne l'inverse généralisée de  $F$ ) a la fonction de répartition  $F$ .

*Démonstration :* Soit  $t \in \mathbb{R}$ , alors, d'après (III.2.5),

$$\mathbb{P}[X \leq t] = \mathbb{P}[F^{-1}(U) \leq t] = \mathbb{P}[U \leq F(t)] = F(t). \quad \square$$

D'après ce lemme, si l'on sait générer une suite de variables aléatoires indépendantes  $U_1, \dots, U_n$ , uniformes sur  $[0; 1]$ , les variables  $X_i = F^{-1}(U_i)$  sont i.i.d. de f.r.  $F$ . En pratique on sait générer sur ordinateur des variables pseudo-aléatoires (en fait déterministes

mais ayant apparemment le même comportement que de vraies variables aléatoires) uniformes sur  $[0; 1]$ . Si la fonction  $F$  se calcule et s'inverse raisonnablement bien, on peut mettre en oeuvre cette méthode.

**Exemple 1 : Loi exponentielle.** La loi exponentielle standard  $\mathcal{E}(1)$  est la loi de densité  $e^{-x}$  sur  $\mathbb{R}_+$ . Sa f.r. s'écrit  $1 - \exp(-x)$  sur  $\mathbb{R}_+$ , d'inverse  $F^{-1}(u) = -\log(1 - u)$ , donc  $X = -\log(1 - U)$  est  $\mathcal{E}(1)$  et, puisque  $U$  et  $1 - U$  ont la même loi  $\mathcal{U}([0; 1])$ ,  $Y = -\log U$  est aussi  $\mathcal{E}(1)$ .

**Exemple 2 : Loi de Cauchy.** La f.r. de la loi de Cauchy s'écrit  $F(x) = 1/2 + \pi^{-1} \arctan(x)$  d'où  $F^{-1}(u) = \tan[\pi(u - 1/2)]$  et  $X = \tan[\pi(U - 1/2)]$  a la loi requise.

Malheureusement la méthode précédente ne s'applique pas très souvent soit parce que l'on ne sait pas calculer  $F$  explicitement (lois de Gauss ou lois Gamma), soit parce que  $F$  s'inverse difficilement. On a alors recours à d'autres méthodes plus adaptées et très bien expliquées dans certains ouvrages. On peut en particulier trouver l'excellent livre de Luc Devroye (fourni gratuitement par l'auteur) à l'adresse suivante :

<http://luc.devroye.org/rnbookindex.html>

## 2 Le Bootstrap

### 2.1 Introduction du problème

En Statistique on est souvent amené à se poser le problème suivant : on dispose de  $n$  observations i.i.d. de loi (f.r.) inconnue  $F$  et l'on veut estimer une caractéristique  $\theta = t(F)$  de  $F$ , comme sa moyenne ou sa médiane, par exemple. Si l'on dispose d'un estimateur sans biais  $\hat{\theta}$  de  $\theta$ , dont on veut évaluer les performances, il serait souhaitable de connaître sa variance pour en déduire un intervalle de confiance pour  $\theta$ , soit par application de l'Inégalité de Bienaymé-Tchébychev, soit par approximation gaussienne.

Prenons l'exemple de la moyenne,  $t(F) = \mathbb{E}_F[X]$ , et de son estimateur empirique  $\hat{\theta} = \bar{X}_n$ . Dans ce cas, on sait que la variance de  $\hat{\theta}$  est  $n^{-1}\sigma^2(F)$ , où  $\sigma^2(F)$  désigne la variance des  $X_i$  qu'en général on ne connaît pas vu que l'on ne connaît pas  $F$  et que l'on estime par la variance empirique  $n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sigma^2(F_n)$  où  $F_n$  est la f.r. empirique.

On peut aussi voir ceci autrement : si les  $X_i$  sont fixés, ce qui revient à dire que l'on va travailler conditionnellement aux observations,  $F_n$  l'est aussi et l'on peut considérer un  $n$ -échantillon  $X_1^*, \dots, X_n^*$  de loi  $F_n$ . On notera que si les  $X_i$  sont distincts,  $F_n$  est la f.r. de la loi uniforme sur  $\{X_1, \dots, X_n\}$ , de sorte que tirer les  $X_i^*$  revient à tirer  $n$  fois au hasard dans le premier échantillon, avec remise. On peut alors considérer l'estimateur  $\hat{\theta}^*$  construit à partir du nouvel échantillon :  $\hat{\theta}^* = \bar{X}_n^* = n^{-1} \sum_{i=1}^n X_i^*$ . Sa variance est  $n^{-1} \text{Var}_{F_n}(X_1^*)$ , c'est à dire  $n^{-1}\sigma^2(F_n)$  que l'on peut calculer. Donc la méthode habituelle d'estimation de la variance de  $\hat{\theta}$  consiste simplement à la remplacer par celle de  $\hat{\theta}^* = \hat{\theta}(X_1^*, \dots, X_n^*)$  qui est le même estimateur, mais construit sur un  $n$ -échantillon de la loi empirique  $F_n$ .

### 2.2 La méthode du bootstrap pour la variance

Dans le cas de la moyenne empirique, la situation est particulièrement simple parce que l'on dispose d'une formule explicite pour la variance de cet estimateur, ce qui est un cas exceptionnel. Si l'on prend pour  $t(F)$  la médiane et pour  $\hat{\theta}(X_1, \dots, X_n)$  la médiane empirique, on ne dispose pas d'une formule simple de calcul de la variance de  $\hat{\theta}(X_1, \dots, X_n)$ ,

seulement d'une formule asymptotique qui fait intervenir la densité, a priori inconnue, de la loi des  $X_i$ . On peut néanmoins appliquer le même principe : on considère un  $n$ -échantillon  $X_1^*, \dots, X_n^*$  de la loi empirique  $F_n$ , dit *échantillon bootstrap*, on construit la médiane empirique correspondante  $\hat{\theta}^* = \hat{\theta}(X_1^*, \dots, X_n^*)$  et l'on prend comme estimateur de  $\text{Var}_F(\hat{\theta})$  la variance de  $\hat{\theta}^*$ , la différence étant que l'on ne peut calculer la première, puisque  $F$  est inconnue, alors que l'on peut, en principe, calculer la seconde car  $F_n$  est connue. On dit que  $\text{Var}_{F_n}(\hat{\theta}^*)$  est *l'estimateur bootstrap* de  $\text{Var}_F(\hat{\theta})$ .

En fait, le calcul explicite de la variance d'une médiane empirique est compliqué, même si l'on connaît la loi des observations. On a alors recours à une simulation pour résoudre le problème. On veut évaluer la variance de la variable  $\hat{\theta}^*$ . On sait bien qu'asymptotiquement la variance empirique converge vers la vraie variance, donc si l'on dispose d'un grand nombre  $k$  de variables i.i.d.  $\hat{\theta}_1^*, \dots, \hat{\theta}_k^*$  de même loi que  $\hat{\theta}^*$ , on aura une très bonne approximation de la variance de  $\hat{\theta}^*$  par la variance empirique des  $\hat{\theta}_j^* : k^{-1} \sum_{j=1}^k (\hat{\theta}_j^* - \bar{\theta}^*)^2$ . Il suffit pour cela de choisir  $k$  suffisamment grand. Or il n'est pas difficile de fabriquer des variables aléatoires de même loi que  $\hat{\theta}^*$  dès que l'on dispose d'un bon générateur de variables aléatoires et, dans ce cadre, le choix de  $k$  appartient au statisticien.

**Procédure pratique** Supposons que l'on dispose d'un générateur aléatoire qui permette de tirer au hasard un entier  $I$  dans  $\{1; \dots; n\}$ . Il est facile d'en construire un à partir d'un générateur d'une v.a. uniforme  $U$  sur  $[0; 1]$  : il suffit de poser  $I = \inf\{i \in \mathbb{N} \mid i \geq nU\}$  car, p.s.,  $nU \in ]0, n]$ . Dans ce cas  $X_I$  est une valeur au hasard dans  $\mathcal{X} = \{X_1, \dots, X_n\}$ , les  $n$  points étant pris sans suppression des ex-aequo. On peut donc tirer un nombre aussi grand que l'on veut de variables aléatoires de loi uniforme sur  $\mathcal{X}$ , c'est à dire de f.r.  $F_n$ . On va en tirer  $kn$ , toutes indépendantes et notées  $X_{j,i}^*$  avec  $1 \leq j \leq k$  et  $1 \leq i \leq n$ . Si l'on s'intéresse à la variance d'une certaine variable aléatoire  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$  où les  $X_i$  sont i.i.d. de loi inconnue  $F$ , on va calculer, pour  $1 \leq j \leq k$  les quantités bootstrap de  $\hat{\theta}$  :

$$\hat{\theta}_j^* = \hat{\theta}(X_{j,1}^*, X_{j,2}^*, \dots, X_{j,n}^*).$$

Ces variables sont i.i.d. de même loi que  $\hat{\theta}^* = \hat{\theta}(X_1^*, \dots, X_n^*)$  où  $X_1^*, \dots, X_n^*$  est un  $n$ -échantillon de loi  $F_n$  et forment donc un  $k$ -échantillon de la loi de  $\hat{\theta}^*$ . Donc si  $k$  est suffisamment grand, la variance empirique des  $\hat{\theta}_j^*$  est très voisine de la variance théorique de  $\hat{\theta}^*$  et l'on peut remplacer l'une par l'autre. En pratique, un  $k$  de l'ordre de quelques centaines convient le plus souvent.

**Pourquoi le bootstrap fonctionne-t-il ?** La deuxième partie de la procédure qui consiste à remplacer la variance de  $\hat{\theta}^*$  par la variance empirique des  $\hat{\theta}_j^*$  fonctionne automatiquement d'après la loi des grands nombres du moment que  $k$  est suffisamment grand. Le bon choix de  $k$  est une question d'expérience mais ceci est le plus souvent indiqué dans les livres. Sa taille n'est limitée que par la puissance de calcul dont on dispose et ce n'est plus un problème aujourd'hui en général. Ceci dit, cet aspect explique que le bootstrap soit une procédure récente qui ne s'est répandue que depuis les années 80 avec l'essor du calcul facile et bon marché.

Le vrai problème est au premier niveau où l'on remplace la variance de  $\hat{\theta}(X_1, \dots, X_n)$  où  $X_1, \dots, X_n$  est un  $n$ -échantillon de loi  $F$  par celle de  $\hat{\theta}(X_1^*, \dots, X_n^*)$  où  $X_1^*, \dots, X_n^*$  est un  $n$ -échantillon de loi  $F_n$ . Notons  $v(G)$  la variance de  $\hat{\theta}(Y_1, \dots, Y_n)$  lorsque  $Y_1, \dots, Y_n$  est un  $n$ -échantillon de loi  $G$ . Quand on applique le bootstrap, on remplace tout simplement

$v(F)$  par  $v(F_n)$ . Or on sait que, si  $n$  n'est pas trop petit,  $F_n$  est proche de  $F$  avec une grande probabilité d'après l'Inégalité de Massart. Si la fonction  $v$  est une fonction continue de la fonction de répartition, on peut espérer que  $v(F_n)$  soit proche de  $v(F)$ . Clairement, ceci sera vrai lorsque  $n \rightarrow +\infty$  si  $v$  est continue au point  $F$ . Dans ce cas, on dit que *le bootstrap fonctionne*.

## 2.3 La méthode générale du bootstrap

Le problème que l'on se pose est d'évaluer une quantité  $t(F)$  fonction d'une loi inconnue  $F$  à partir de la connaissance d'un  $n$ -échantillon  $X_1, \dots, X_n$  de loi  $F$ , par exemple la variance d'un estimateur  $T(X_1, \dots, X_n)$  fonction d'un  $n$ -échantillon de loi  $F$ , ou son moment d'ordre 4, ou un quantile de sa loi, ou sa fonction de répartition, etc. Si  $t(F)$  dépend continuellement de la loi  $F$  des observations, il est naturel de l'estimer par  $t(F_n)$  qui est la même quantité fondée sur un  $n$ -échantillon  $X_1^*, \dots, X_n^*$  de loi  $F_n$  où  $F_n$  désigne la loi empirique des  $X_i$ , dans la mesure où l'on sait que  $F_n$  est voisine de  $F$ . Comme on ne peut généralement pas calculer simplement  $t(F_n)$  de manière théorique, on l'évalue par simulation comme indiqué plus haut en tirant un nombre  $k$  suffisamment grand de  $n$ -échantillons de loi  $F_n$ , en recalculant  $k$  fois la variable aléatoire d'intérêt pour chacun des échantillons et en utilisant les  $k$  valeurs obtenues pour estimer ce que l'on ne peut calculer. Dans la suite, on notera systématiquement  $X_1^*, \dots, X_n^*$  un *échantillon bootstrap*, c'est à dire de loi  $F_n$ .

## 2.4 Exemples et applications

### 2.4.1 Biais d'un estimateur

Soit  $\hat{\theta}(X_1, \dots, X_n)$  un estimateur d'une certaine quantité  $\theta(F)$  fonction de la loi  $F$  des  $X_i$ . Dans beaucoup de situations cet estimateur est biaisé et l'on sait qu'un bon estimateur doit avoir un petit biais. Il est souvent intéressant d'évaluer le biais

$$b(F) = \mathbb{E}_F [\hat{\theta}(X_1, \dots, X_n)] - \theta(F)$$

de l'estimateur. L'estimateur bootstrap de  $b(F)$  est tout simplement

$$b(F_n) = \mathbb{E}_{F_n} [\hat{\theta}(X_1^*, \dots, X_n^*)] - \theta(F_n).$$

Si, par exemple,  $\theta(F) = \sigma^2(F)$  est la variance de  $F$  et  $\hat{\theta}$  est la variance empirique biaisée  $n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ , on sait que son espérance est  $n^{-1}(n-1)\sigma^2(F)$  et son biais  $b(F) = -n^{-1}\sigma^2(F)$ . L'estimateur bootstrap du biais est donc

$$b(F_n) = -n^{-1}\sigma^2(F_n) = -n^{-2} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

que l'on peut dans ce cas calculer explicitement. Le plus souvent ce n'est pas le cas parce que l'on ne sait pas calculer explicitement  $\mathbb{E}_{F_n} [\hat{\theta}(X_1^*, \dots, X_n^*)]$ . On est alors amené à recourir à une simulation pour évaluer  $b(F_n)$ . Pour ce faire on tire  $k$  échantillons bootstrap indépendants  $X_{j,1}^*, X_{j,2}^*, \dots, X_{j,n}^*$ ,  $1 \leq j \leq k$  et l'on remplace l'espérance inconnue par la moyenne empirique  $k^{-1} \sum_{j=1}^k \hat{\theta}(X_{j,1}^*, \dots, X_{j,n}^*)$ . On obtient finalement comme estimateur du biais

$$k^{-1} \sum_{j=1}^k \hat{\theta}(X_{j,1}^*, X_{j,2}^*, \dots, X_{j,n}^*) - \theta(F_n).$$

### 2.4.2 Loi d'un estimateur

Il est souvent intéressant de connaître la loi de l'erreur d'un estimateur  $\hat{\theta}(X_1, \dots, X_n)$  de  $\theta(F)$ , c'est à dire la loi sous  $F$  de  $\hat{\theta}(X_1, \dots, X_n) - \theta(F)$ . Le principe est le même. L'estimateur bootstrap est la loi de  $\hat{\theta}(X_1^*, \dots, X_n^*) - \theta(F_n)$  que l'on peut aisément approcher par la loi empirique de cette même quantité, c'est à dire la loi empirique des  $k$  variables aléatoires  $\hat{\theta}(X_{j,1}^*, \dots, X_{j,n}^*) - \theta(F_n)$  avec  $1 \leq j \leq k$ . On sait (cf. l'Inégalité de Massart) que la distance entre la vraie loi et l'empirique est avec une grande probabilité, inférieure à  $2/\sqrt{k}$ . Si l'on veut rendre ceci très petit, il est souhaitable de choisir une grande valeur de  $k$ , de l'ordre de quelques milliers au moins.

### 2.4.3 Intervalles de confiance

Comment construit-on le plus souvent des intervalles de confiance pour un paramètre inconnu  $\theta = \theta(F)$ . Une méthode (entre autres) est la suivante. On considère un (bon) estimateur  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$  de  $\theta$ , la loi  $Q_\theta$  de  $\hat{\theta}(X_1, \dots, X_n) - \theta$  et l'on choisit un intervalle  $[a_\theta, b_\theta]$  (unilatère ou bilatère) tel que  $Q_\theta([a_\theta, b_\theta]) \geq 1 - \alpha$ . On a alors

$$\mathbb{P}_\theta [a_\theta \leq \hat{\theta} - \theta \leq b_\theta] = \mathbb{P}_\theta [\hat{\theta} - b_\theta \leq \theta \leq \hat{\theta} - a_\theta] \geq 1 - \alpha. \quad (\text{IV.2.1})$$

Si  $a_\theta$  et  $b_\theta$  sont indépendants de  $\theta$ , c'est terminé. Sinon, on utilise le fait que  $a_\theta$  est typiquement continue en  $\theta$  et que  $\hat{\theta}$  est proche de  $\theta$  pour remplacer  $a_\theta$  par  $a_{\hat{\theta}}$  et de même pour  $b_\theta$ . On conclut par un argument de continuité de la loi  $Q_\theta$  (supposée de f.r. continue). De toutes façons, on ne connaît le plus souvent pas  $Q_\theta$  que l'on a déjà remplacé par une approximation. Le cas type en variables i.i.d. étant  $\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, \sigma^2(\theta))$  avec  $\sigma$  continue, d'où  $\sqrt{n}[\sigma(\hat{\theta}_n)]^{-1}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, 1)$ . La même procédure s'applique si l'on connaît plutôt la loi (limite) de  $\hat{\theta}_n/\theta$ . Dans ce cas on peut faire une transformation pour se ramener à  $\log \hat{\theta}_n - \log \theta$ .

Une idée naturelle, lorsque la loi  $Q_\theta$  n'est pas connue, est d'utiliser le bootstrap pour l'approximer, c'est à dire de remplacer  $Q_\theta$  par son estimateur bootstrap qui est la loi  $Q^*$  de  $\hat{\theta}(X_1^*, \dots, X_n^*) - \theta(F_n)$ . En pratique on la remplacera par son estimée empirique  $Q_k^*$  fondée sur les  $k$  variables i.i.d.  $V_j = \hat{\theta}(X_{j,1}^*, \dots, X_{j,n}^*) - \theta(F_n)$ ,  $1 \leq j \leq k$ , pour  $k$  suffisamment grand. On choisira alors un intervalle  $[a^*, b^*]$  contenant une proportion  $(1 - \alpha)$  des  $V_j$ , c'est à dire un nombre à peu près égal (il doit être entier) à  $(1 - \alpha)k$  des  $V_j$ . D'un point de vue pratique, on ordonne les  $V_j$ , on ignore les  $k_1$  plus petits et les  $k_2$  plus grands avec  $k_1 + k_2 \sim \alpha k$  et on garde l'intervalle  $[a^*, b^*]$  qui contient les autres. On conclut alors que, approximativement,  $Q^*([a^*, b^*]) \sim 1 - \alpha$  ou encore

$$\mathbb{P} [a^* \leq \hat{\theta}(X_1^*, \dots, X_n^*) - \theta(F_n) \leq b^* \mid X_1, \dots, X_n] \sim 1 - \alpha.$$

On obtient ainsi un intervalle de confiance bootstrap en remplaçant  $a_\theta$  et  $b_\theta$  par  $a^*$  et  $b^*$  respectivement dans (IV.2.1).

### 2.4.4 Régression linéaire

Soit un modèle classique de régression

$$Y_i = \sum_{j=1}^p \beta_j X_i^j + \varepsilon_i \quad \text{avec } X_i^1 = 1, \quad 1 \leq i \leq n,$$

où les erreurs  $\varepsilon_i$  sont i.i.d. avec des moments d'ordre 4 mais pas nécessairement gaussiennes. On peut s'intéresser aux estimateurs des moindres carrés et en particulier à l'estimateur sans biais  $s_n^2$  de la variance ou à une quantité comme  $(\hat{\beta}_j - \beta_j)/(\delta_j s_n)$  où  $\sigma^2 \delta_j^2$  est la variance de  $\hat{\beta}_j$ . Ces quantités ont des lois bien connues dans le modèle gaussien mais pas dans le cas général et il peut être intéressant d'estimer certaines de leurs caractéristiques. Comment fait-on pour calculer des estimateurs bootstrap dans ce cadre où les observations ne sont plus i.i.d. ?

Il faut trouver quelque chose qui joue le rôle des observations usuelles. Ici ce sont les  $\varepsilon_i$  qui sont bien i.i.d. mais ne sont pas observables. On les remplace donc par ce que l'on nomme les *résidus* et que l'on détermine comme suit. On calcule l'estimateur des moindres carrés  $\hat{\beta}$  de  $\beta$  et les pseudo-observations  $\hat{Y}_i = \sum_{j=1}^p \hat{\beta}_j X_i^j$  coordonnées du vecteur  $\hat{Y}$ , projection de  $Y$  sur l'espace vectoriel  $V$  engendré par les colonnes de la matrice  $X$ . Le vecteur  $\hat{\varepsilon} = \hat{Y} - Y$ , projection de  $Y$  (donc de  $\varepsilon$ ) sur  $V^\perp$ , forme le vecteur des résidus (avec lequel on calcule la somme des carrés résiduelle). Ses coordonnées ne sont pas indépendantes mais ont une loi empirique  $F_n$  qui est centrée ( $\sum_{i=1}^n \hat{\varepsilon}_i = 0$ ) comme la vraie loi des  $\varepsilon_i$  parce que  $\hat{\varepsilon}$  est orthogonal à  $V$ , donc à  $X^1$ . Pour créer un échantillon bootstrap des  $\varepsilon_i$ , on échantillonne selon la loi  $F_n$  ce qui permet d'avoir des  $\varepsilon_i^*$  et un échantillon bootstrap des  $Y_i$  donné par

$$Y_i^* = \sum_{j=1}^p \hat{\beta}_j X_i^j + \varepsilon_i^*, \quad 1 \leq i \leq n.$$

A partir de cet échantillon bootstrap, on peut calculer les estimateurs des moindres carrés,  $\hat{\beta}^* = (X^t X)^{-1} X^t Y^*$  et  $s_n^{*2}$ . On n'a pas besoin de bootstrapper la matrice  $X$  qui est connue. On peut, par exemple, utiliser cette méthode pour évaluer la variance de  $s_n^2$ , ou trouver des intervalles de confiance pour  $(\hat{\beta}_j - \beta_j)/(\delta_j s_n)$  en utilisant la technique du paragraphe précédent. En effet, cette quantité a une loi qui est inconnue (pas nécessairement de Student) si les erreurs ne sont pas gaussiennes.

#### 2.4.5 Quand est-ce que le bootstrap fonctionne ?

Avant d'utiliser la méthode du bootstrap il convient de s'assurer qu'elle fonctionne correctement, c'est à dire que les hypothèses de continuité évoquées plus haut sont effectivement satisfaites. Les études théoriques ont été faites dans de très nombreux articles et quelques livres et l'on peut s'y référer. Il n'y a pas de théorème général qui couvre toutes les situations, seulement une suite de théorèmes spécifiques. En gros, on peut dire que le bootstrap fonctionne dans les situations où les statistiques d'intérêt ont des asymptotiques gaussiennes. Un exemple pour lequel le bootstrap ne fonctionne pas est le suivant : si l'on a une loi à support borné supérieurement par  $A$  et  $X_{(n)}$  désigne la plus grande des observations, la loi limite de  $A - X_{(n)}$  n'est pas gaussienne (cf. l'exemple des lois uniformes sur  $[0; \theta]$ ). Dans ce cas on ne peut pas la remplacer par son équivalent bootstrap qui serait la loi de  $X_{(n)} - X_{(n)}^*$  où  $X_{(n)}^*$  désigne la plus grande des observations d'un échantillon bootstrap (de loi  $F_n$ ).

## Une courte bibliographie pour ceux qui voudraient en savoir davantage

BICKEL, P.J. and DOKSUM, K.A. (1977). *Mathematical Statistics*. Holden Day, San Francisco.

BOROVKOV, A. (1987). *Statistique Mathématique*. Editions MIR, Moscou.

BOROVKOV, A. (1998). *Mathematical Statistics*. Gordon & Breach, Amsterdam.

DEVROYE, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New-York.

EFRON, B. and TIBSHIRANI, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

FELLER, W. (1968). *An Introduction to Probability Theory and its Applications, Vol. I (Third Ed.)*. John Wiley, New York.

MASSART, P. (1990). The tight constant in the D.K.W. inequality. *Ann. Probab.* **18**, 1269-1283.

TOMASSONE, R., LESQUOY E. et MILLIER, C. (1983). *La Régression, Nouveaux Regards sur une Ancienne Méthode Statistique*. Masson, Paris.

van der VAART, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.