IFT 3245
Simulation et modèles

Fabian Bastin
DIRO
Université de Montréal

Automne 2016

Estimation séquentielle

Pour un intervalle de confiance de niveau $1 - \alpha$, si on fixe n, la largeur $I_2 - I_1$ est aléatoire.

Si on veut $I_2 - I_1 \leq w$ pour w fixé, la valeur minimale de n requise est une variable aléatoire N.

Comment prédire ce N?

Pour $X_i \sim$ binomiale$(1, p)$, avec $n = 1000$ on a obtenu $\overline{X}_n = 0.882$, $S_n^2 \approx 0.1042$, et la demi-largeur du intervalle de confiance à 95% était de 0.020. Combien de répétitions additionnelles faut-il pour réduire la demi-largeur à environ 0.005?

Estimation séquentielle

Nous voulons $1.96 S_n/\sqrt{n} \leq 0.005$. En supposant que S_n ne changera pas significativement, cela donne $n \geq (1.96 \times S_n/0.005)^2 \approx 16011.8$. En conséquence, nous pouvons recommander de faire 15012 répétitions additionnelles.

Procédure à deux étapes

Cette approche est valable pour la loi de Student (ou normale).

Faire n_0 répétitions et calculer $S^2_{n_0}$; la prédiction du n requis est

$$\hat{N}^* = \min\left\{n \mid (t_{n-1,1-\alpha/2})S_{n_0}/\sqrt{n} \leq r\right\}.$$

On fera $\max(0, \ \hat{N}^* - n_0)$ répétitions additionnelles.

Bien sûr, il se peut que ce soit insuffisant, ou trop.

Après n_0, recalculer S^2_n et la demi-largeur pour chaque n. On s'arrête dès que $I_2 - I_1 \leq w$.

Procédure à deux étapes

Cette procédure est biaisée, car on tend à s'arrêter à un N où S_N^2 sous-estime la variance.

Mais lorsque $w \to 0$, le bias disparaît, $N/n^* \to 1$ a.p.1 où n^* est la valeur optimale de N si on connaissait σ^2, et $P[|\overline{X}_N - \mu| \leq w/2] \to 1 - \alpha$.

IC pour des estimateurs de quantile

Si X est de répartition F, le q-quantile de F est

$$\xi_q = F^{-1}(q) = \inf\{x : F(x) \geq q\}.$$

Soit $X_{(1)}, \ldots, X_{(n)}$ un échantillon i.i.d. de X, trié, et $\hat{F}_n$ la fonction de répartition empirique. Un estimateur simple de ξ_q est le quantile empirique

$$\hat{\xi}_{q,n} = \hat{F}_n^{-1}(q) = \inf\{x : \hat{F}_n(x) \geq q\} = X_{(\lceil nq \rceil)}.$$

Il est biaisé mais fortement consistent et obéit au théorème de la limite centrale, comme le montre le théorème ci-après.

IC pour des estimateurs de quantile: théorème

(i) *Pour chaque q, $\hat{\xi}_{q,n}$ to as ξ_q quand $n \to \infty$.*

(ii) *Si X a une densité f strictement positive et continue dans un voisinage de ξ_q, alors*

$$\frac{\sqrt{n}(\hat{\xi}_{q,n} - \xi_q) f(\xi_q)}{\sqrt{q(1-q)}} \xrightarrow{D} N(0,1) \quad \textit{quand } n \to \infty.$$

Ce TLC indique qu'il y a beaucoup de bruit (variance) si $f(\xi_q)$ est petit.

De plus, pour l'utiliser afin de construire un intervalle de confiance, il faut estimer $f(\xi_q)$, ce qui est difficile.

IC pour des estimateurs de quantile

Nous pouvons néanmoins construire une méthode non-asymptotique de calcul d'un intervalle de confiance pour ξ_q: supposons que F est continue en ξ_q.

Soit B le nombre d'observations $X_{(i)}$ inférieures à ξ_q.

Puisque $P[X < \xi_q] = q$, B est binomiale(n, q).

Si $1 \leq j < k \leq n$, $X_{(j)} < \xi_q \leq X_{(k)}$ ssi $j \leq B < k$. Alors

$$P[X_{(j)} < \xi_q \leq X_{(k)}] = P[j \leq B < k] = \sum_{i=j}^{k-1} \binom{n}{i} q^i (1-q)^{n-i}.$$

On choisit j et k pour que cette somme soit supérieure ou égale à $1 - \alpha$ (intervalle unilatéral ou bilatéral).

IC pour des estimateurs de quantile

Si n est grand et q n'est pas trop proche de 0 ou 1, on peut approximer la loi binomiale par la loi normale:

$$\frac{B - nq}{\sqrt{nq(1-q)}} \approx N(0,1).$$

On obtient alors $j = \lfloor nq + 1 - \delta \rfloor$ et $k = \lfloor nq + 1 + \delta \rfloor$, où $\delta = \sqrt{nq(1-q)}\Phi^{-1}(1 - \alpha/2)$.

Exemple: valeur à risque

Soit L la perte nette de valeur d'un porte-feuille d'actifs pour une période de temps donnée $[0, T]$. La valeur à risque (VAR) (au temps 0) est la valeur de x_p telle que $P[L > x_p] = p$. C'est le $(1 - p)$-quantile de L.

Valeurs courantes: $p = 0.01$, $T = 2$ semaines (banques), $T =$ mois ou années (assurance, fonds de pension).

On peut critiquer l'utilisation de la VAR, vu qu'elle donne une information très limitée.

Par exemple si $x_{0.01} = 10^7$ dollars, que sait-on sur l'importance réelle de la perte?

Une mesure complémentaire pourrait être $E[L \mid L > x_p]$, par exemple.

Exemple: valeur à risque

Modèles pour estimer la VAR: on doit modéliser l'évolution du prix des actifs (souvent plusieurs milliers, dépendants). Souvent: modèles à facteurs.

On peut remplacer les actifs par des prêts, comptes à payer, etc. Sauf dans les cas simples, on estime la VAR par simulation. Quand p est petit: "importance sampling".