IFT 3245
Simulation et modèles

Fabian Bastin
DIRO
Université de Montréal

Automne 2016

Détection et réduction du biais initial

Idéalement, nous souhaiterions générer l'état initial selon la loi d'équilibre, En effet, dans un tel cas, il n'y aurait aucun biais, mais une telle manière de procéder est en général très délicate.

Pratiquement, il n'est dès lors pas possible de procéder ainsi, et il nous faudra considérer le biais résultant de l'état initial de la simulation.

Si $(\mu_i, \sigma_i^2, \rho_{i,i+j}) \to (\mu, \sigma^2, \rho_j)$ quand $i \to \infty$, la question est alors de savoir comment détecter et réduire le biais $|E[\bar{C}_n] - \mu|$?

Détection et réduction du biais initial

Les solutions pratiques sont avant tout heuristiques.

Considérons un processus en temps discret $\{C_i, \ i \geq 1\}$ (le cas continu est similaire), et supposons que l'on ne compte pas les n_0 premières observations:

$$\bar{C}_{n_0,n} = \frac{1}{n - n_0} \sum_{j=n_0+1}^{n} C_j.$$

Souvent, nous pouvons écrire $|E[C_i] - \mu| \leq \kappa_0 \beta^i$ (i.e. $|E[C_i] - \mu| \in O(\beta^i)$) pour $\kappa_0 < \infty$ et $\beta < 1$.

Détection et réduction du biais initial

Dans ce cas, nous avons

$$\begin{aligned}
\left|E[\bar{C}_{n_0,n}] - \mu\right| &\leq \frac{\kappa_0}{n-n_0}\sum_{i=n_0+1}^{n}\beta^i = \frac{\kappa_0\beta^{n_0+1}(1-\beta^{n-n_0})}{(n-n_0)(1-\beta)} \\
&\in O\left(\frac{\beta^{n_0+1}}{(n-n_0)(1-\beta)}\right) \approx O\left(\frac{\beta^{n_0}}{n(1-\beta)}\right)
\end{aligned}$$

si n est très grand par rapport à n_0.

En d'autres termes, $|E[\bar{C}_n] - \mu|$ est en $O(1/(n(1-\beta)))$, si $n \gg n_0$.

En augmentant n_0, nous diminuons le biais, mais ce faisant, nous augmentons aussi la variance, qui est dans $O(1/(n-n_0))$.

Détection et réduction du biais initial

Quand n tend vers l'infini, le carré du biais devient négligeable par rapport à la variance, ainsi que, par conséquent, sa contribution dans le MSE.

Il n'existe cependant aucune méthode fiable et universelle permettant de choisir n_0 (ou t_0) en pratique.

Une manière populaire de procéder est d'utiliser l'heuristique de Welch.

Heuristique de Welch

1. Faire k simulations de longueur n_1. Soit C_{ij} l'observation j de la répétition i.
2. Posons $\bar{C}_j = \sum_{i=1}^{k} C_{ij}/k$.
3. Lissage des hautes fréquences par une moyenne mobile de largeur w:

$$\bar{C}_j(w) = \begin{cases} \frac{1}{2w+1} \sum_{s=-w}^{w} \bar{C}_{j+s}, & j = w+1, \ldots, n_1 - w, \\ \frac{1}{2j-1} \sum_{s=-(j-1)}^{j-1} \bar{C}_{j+s}, & j = 1 \ldots, w. \end{cases}$$

4. Regardons le graphique de $\bar{C}_j(w)$ en fonction de j et soit n_0 la valeur de j où le processus semble stable (choix subjectif).

heuristique de Welch

Schématiquement, la procédure peut être résumée comme suit.

Simulation 1	$C_{11}, C_{12}, \dots, C_{1,j-w}, \dots, C_{1,j+w}, \dots, C_{1,T_1}$
Simulation 2	$C_{21}, C_{22}, \dots, C_{2,j-w}, \dots, C_{2,j+w}, \dots, C_{2,T_1}$
$\vdots$	
Simulation N	$C_{N1}, C_{N2}, \dots, C_{N,j-w}, \dots, C_{N,j+w}, \dots, C_{N,T_1}$
Moyenne	$\overline{C}_1, \overline{C}_2, \dots, \overline{C}_{j-w}, \dots, \overline{C}_{j+w}, \dots, \overline{C}_{T_1}$
Moyenne mobile	$\overline{C}_{j-w}(w)$

Horizon tronqué

Une fois n_0 (ou t_0) choisi, nous faisons k répétitions indépendantes de longueurs $T_1, \ldots, T_k$, où les T_i sont des variables déterministes ou aléatoires.

Pour la répétition i, posons

$$X_i = \frac{1}{T_i - t_0} \sum_{j=N(t_0)+1}^{N(T_i)} C_j.$$

Un estimateur global est alors $\bar{X}_k = (1/k) \sum_{i=1}^{k} X_i$.

Si k est strictement plus grand que 2, nous pouvons exploiter le fait que les X_i sont i.i.d. pour estimer la variance et calculer un intervalle de confiance. Il s'agit l'approche "réplication-suppression."

Horizon tronqué

Dans le cas discret, nous avons

$$X_i = \frac{1}{n - n_0} \sum_{j=n_0+1}^{n} C_j.$$

Pour un budget de calcul fixe kn, comment choisir k?

Typiquement, $k = 1$ minimise le MSE, mais pas toujours, et il est plus difficile d'estimer la variance lorsque $k = 1$.

Le MSE est minimisé pour k strictement plus grand que 1 lorsque

(a) les autocorrélations diminuent très lentement <u>et</u>
(b) le biais initial est très faible ou diminue très vite.

Horizon tronqué

Supposons que

$$|E[C_j] - \mu| = \kappa_0 \beta^j$$

et

$$\text{Cov}[C_i,\ C_{i+j}] = \sigma^2 \rho_j = \sigma^2 \alpha^j$$

pour $j \geq 0$, où $\beta < 1$, $0 < \alpha < 1$, $\kappa_0 > 0$, et $\sigma^2 > 0$.

Budget de calcul $B = nk$. Choisir k et n_0 pour minimiser

$$\begin{aligned} \text{MSE}[\bar{X}_k] &= (E[X_i - \mu])^2 + \text{Var}[X_i]/k \\ = & \left(\frac{\kappa_0(\beta^{n_0+1} - \beta^{n+1})}{(n-n_0)(1-\beta)} \right)^2 + \frac{\sigma^2}{(n-n_0)k} \times \\ & \left(1 + 2 \sum_{j=1}^{n-n_0-1} \frac{\alpha^j (n - n_0 - j)}{n - n_0} \right). \end{aligned}$$

Horizon tronqué

Il faut prendre $k > 1$ si β et κ_0 sont petits et si σ^2 et α sont grands.

Mais si n_0 et n sont fixés et $k \to \infty$, le biais ne disparait pas et le TLC devient

$$\frac{\sqrt{k}(\bar{X}_k - \mu)}{\sigma_{n_0,n}} \xrightarrow{D} N(0,1) + \frac{\sqrt{k}\beta_{n_0,n}}{\sigma_{n_0,n}},$$

où $\beta_{n_0,n} = E[X_i] - \mu$ et $\sigma^2_{n_0,n} = \mathrm{Var}[X_i]$.

Horizon tronqué

Un intervalle de confiance basé sur l'hypothèse que les X_i sont i.i.d. normales est asymptotiquement valide seulement si $\sqrt{k}\beta_{n_0,n}/\sigma_{n_0,n} \to 0$ quand $k \to \infty$.

Autrement dit, le biais doit converger vers zéro plus vite que l'écart-type de $\bar{X}_k$.

Si $|E[C_j] - \mu| = \kappa_0\beta^j$ et $\rho_j = \sigma^2\alpha^j$, alors

$$\beta_{n_0,n} = \frac{\kappa_0(\beta^{n_0+1} - \beta^{n+1})}{(n-n_0)(1-\beta)} \in O\left(\frac{\beta^{n_0}}{n-n_0}\right)$$

et

$$\frac{\sigma^2_{n_0,n}}{k} = \frac{\sigma^2}{(n-n_0)k}\left(1 + 2\sum_{j=1}^{n-n_0-1}\frac{\alpha^j(n-n_0-j)}{n-n_0}\right) \in O\left(\frac{1}{(n-n_0)k}\right).$$

Horizon tronqué

Ainsi,

$$\frac{\sqrt{k}\,\beta_{n_0,n}}{\sigma_{n_0,n}} \in O\left(\sqrt{k/(n-n_0)}\,\beta^{n_0}\right).$$

Cette expression converge vers 0 lorsque $k \to \infty$ si et seulement si

$$\ln\left(\sqrt{\frac{k}{n-n_0}}\beta^{n_0}\right) \to -\infty$$

quand k tend vers l'infini, ou, en d'autres termes, si et seulement si $\ln[k/(n-n_0)]/2 + n_0 \ln\beta \to -\infty$. Comme β est supposé strictement plus petit que 1, cela revient à exiger

$$\frac{\ln[k/(n-n_0)]}{2\ln\beta} + n_0 \to \infty.$$

Autrement dit, si on augmente k, il faut aussi augmenter n et n_0 assez vite.