MPRO FAT - PROJET VÉLIB

DÉCEMBRE 2020

1. Présentation

On considère un système de vélos partagés, type Vélib, où les vélos sont disponibles dans des stations dédiées et peuvent être empruntés pour faire des trajets d'une station à une autre. Le but de ce projet est de calculer les probabilités stationnaires que chaque station soit vide. Pour cela, il faut :

- Ecrire un simulateur du processus de Markov correspondant pour 5 stations
- Calculer la probabilité stationnaire théorique dans le cas d'un seul vélo
- Comparer au résultat de votre simulation
- Simuler le même réseau avec 100 vélos, initialement répartis de façon uniforme entre les stations et les routes. Calculer la probabilité stationnaire que chaque station soit vide par la méthode décrite en section 5.3. On n'oubliera pas de préciser l'intervalle de confiance.

2. Modalités

Le travail peut se faire en monôme ou binôme. Les groupes ne peuvent pas changer d'un projet à l'autre. Vous pouvez faire la simulation dans le langage de votre choix, nous vous conseillons Python ou Matlab.

- Vous devez déposer pour le **5 janvier à 18h00** un fichier contenant votre modèle du Vélib.
- Le soir du **5 janvier à 18h05** sera disponible sur le Moodle la modélisation que nous souhaitons que vous étudiiez.
- Vous devez déposer sur `https://moodle.r2.enst.fr` un fichier pdf contenant les réponses aux questions théoriques (voir ci-dessous) et les résultats obtenus pour la simulation avec 100 vélos. Le code peut-être en annexe. Il vous est recommandé de travailler en Python Notebook puis d'exporter la feuille de calcul en pdf.

Date limite : 17 janvier 2021 à 23h59

3. Données pour la simulation

On utilisera les données suivantes pour la simulation.

Station	3	4	5	6	7
3		3	5	7	7
4	2		2	5	5
5	4	2		3	3
6	8	6	4		2
7	7	7	5	2	

Temps moyen de trajet tau_ij en minutes

Station	3	4	5	6	7
	2,8	3,7	5,5	3,5	4,6

Taux de départ par heure lambda_i

Station	3	4	5	6	7
3		0,22	0,32	0,2	0,26
4	0,17		0,34	0,21	0,28
5	0,19	0,26		0,24	0,31
6	0,17	0,22	0,33		0,28
7	0,18	0,24	0,35	0,23	

Matrice de routage p_ij

4. Calcul théorique

Une fois écrit votre simulateur, qui sortira forcément des résultats, il faut valider votre code par des résultats théoriques.

1. Utilisez les équations de trafic pour obtenir les relations entre les α_i (notations du document sur les colonies).
2. On considère qu'il n'y a qu'un seul vélo, quelle est alors la taille de l'espace d'état ?
3. Dans ces conditions (un seul vélo), calculer la probabilité que chaque station soit vide.
4. Comparez aux résultats obtenus par simulation.

Pour résoudre numériquement le système d'équations donnant les α_i, on suggère de mettre le système sous forme de calcul matriciel $M\alpha = X$ où α est le vecteur colonne des α_i. On a alors $\alpha = M^{-1}X$. Pour éviter $X = 0$, on pourra remplacer une ligne de M par des 1, on a en effet $n+1$ équations pour n inconnues si on ajoute la condition de normalisation.

5. Simulation

5.1. **Accélération de votre simulation.** Si vous voulez accélérer notablement votre simulation, vous pouvez utiliser la commande *@jit* de la librairie *numba* avant la définition de chaque fonction.

```
from numba import jit

@jit
def pickState(p):
    r=np.cumsum(p)
    r =r/np.sum(p)
    u=np.random.rand()
    w=0
    while(u>r[w]):

        w+=1
    return w
```

Listing 1. Simulation d'une variable aléatoire discrète à valeurs dans $0, 1, \ldots, N$ où N est la longueur du vecteur de probabilité passé en paramètre

5.2. **Simulation d'une variable aléatoire discrète.** Le code ci-dessus met en œuvre l'algorithme de simulation qui suit. Soit X une variable aléatoire discrète et $E = \{x_1, x_2, x_3\}$ son espace d'états. On note $p_i = \mathbf{P}[X = x_i]$ pour $i = 1, 2, 3$ avec $\sum_{i=1}^{3} p_i = 1$. On tire une variable aléatoire u uniforme sur $[0, 1]$, alors :

— Si $u \leq p_1$, $\hat{X} = x_1$ et $\mathbf{P}[\hat{X} = x_1] = p_1$,

— Si $p_1 < u \leq p_1 + p_2$, $\hat{X} = x_2$ et $\mathbf{P}[\hat{X} = x_2] = p_1 + p_2 - p_1 = p_2$,

— Si $p_1 + p_2 < u$, $\hat{X} = x_3$ et $\mathbf{P}[\hat{X} = x_3] = 1 - (p_1 + p_2) = p_3$.

Ainsi, $\hat{X}$ a la même loi que X.

5.3. **Simulation régénérative.** On sait que la probabilité stationnaire d'une chaîne de Markov irréductible, récurrente peut s'écrire

$$\pi(y) = \frac{1}{\tau_x^1} \mathbf{E}_x \left[\sum_{j=0}^{\tau_x - 1} \mathbf{1}_{\{X_j = y\}} \right]$$

où τ_x^1 est l'instant de premier retour en x partant de x. De manière plus générale

$$\bar{z} = \sum_{y \in E} f(y)\pi(y) = \frac{1}{\tau_x^1} \mathbf{E}_x \left[\sum_{j=0}^{\tau_x^1 - 1} f(X_j) \right].$$

Simulons R cycles de x à lui-même, soit $(d_1, \cdots, d_R)$ leur longueur, i.e. $d_j = \tau_x^j - \tau_x^{j-1}$. Soit

$$Y_l = \sum_{j=\tau_x^{l-1}}^{\tau_x^l} f(X_j).$$

Les $\Big((d_k, Y_k), k \in \{1, \cdots, R\}\Big)$ sont indépendants et identiquement distribués donc

$$\hat{\tau}_R = \frac{1}{R} \sum_{k=1}^{R} d_k \xrightarrow{\text{p.s.}} \mathbf{E}_x\left[\tau_x\right] = \bar{\tau}$$

$$\hat{Y}_R = \frac{1}{R} \sum_{k=1}^{R} Y_l \xrightarrow{\text{p.s.}} \mathbf{E}_x\left[\sum_{j=0}^{\tau_x - 1} f(X_j)\right] = \bar{y}$$

$$\sqrt{R}\Big((\hat{\tau}_R, \hat{Y}_R) - (\bar{\tau}, \bar{y})\Big) \xrightarrow{\text{loi}} \mathcal{N}(0, \Gamma)$$

avec

$$\Gamma = \begin{pmatrix} (\tau_x^1) & \operatorname{cov}(\tau_x^1, Y_1) \\ \operatorname{cov}(\tau_x^1, Y_1) & var(Y_1^2) \end{pmatrix}.$$

Un estimateur consistant mais biaisé de $\bar{z}$ est donné par

$$\hat{z}_R = \frac{\hat{Y}_R}{\hat{\tau}_R}.$$

La précision de cet estimateur est donné par

$$\sqrt{R}\,(\hat{z}_R - \bar{z}) \xrightarrow{\text{Loi}} \mathcal{N}(0, \eta^2)$$

où

$$\eta^2 = \gamma_{11}\,\frac{\bar{y}^2}{\bar{\tau}^4} - 2\gamma_{12}\,\frac{\bar{y}}{\bar{\tau}^3} + \gamma_{22}\,\frac{1}{\bar{\tau}^2}.$$

L'intervalle de confiance à 95% est donc donné par

$$\bar{z} \in]\hat{z}_R - \frac{1{,}95\,\eta_R}{\sqrt{R}},\ \hat{z}_R + \frac{1{,}95\,\eta_R}{\sqrt{R}}[.$$

Comme on ne connaît pas a priori ni les γ_{ij}, ni $\bar{y}$, ni $\bar{\tau}$ on les remplace par leur version « empirique » : $\bar{\tau}$ est remplacé par $\hat{\tau}_R$, $\bar{y}$ est remplacé par $\hat{Y}_R$ et γ_{ij} par

$$\hat{\gamma}_{12}^R = \frac{1}{R-1} \sum_{k=1}^{R} d_k Y_k - \hat{\tau}_R \hat{Y}_R$$

$$\hat{\gamma}_{11} = \frac{1}{R-1} \sum_{k=1}^{R} d_k^2 - \hat{\tau}_R^2$$

$$\hat{\gamma}_{22} = \frac{1}{R-1} \sum_{k=1}^{R} Y_k^2 - \hat{Y}_R^2$$