MDI230 - PROJET VÉLIB

JUIN 2018

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 la probabilité de trouver une station vide. Pour cela, il faut :

- Proposer une modélisation du Vélib,
- Calibrer le modèle avec des données réelles (Rouen 2016),
- Simuler le processus de Markov correspondant pour 5 stations,
- Calculer la probabilité stationnaire théorique pour valider la simulation.

2. Modalités

Le travail peut se faire en monôme ou binôme. Les groupes ne peuvent pas changer entre les différentes étapes. Vous pouvez faire la simulation dans le langage de votre choix, nous vous conseillons Python ou Matlab, tous deux présents sur les machines des salles de TP de l'école.

Ce projet compte pour 75% de la note de l'UE. La partie modélisation et calibration compte pour 25%, et le rapport avec la simulation et l'étude théorique compte pour les 50% restant.

3. Échéances

- Avant le 19/06 : Lecture du document sur le modèle des colonies disponible sur le site pédagogique.
- 19/06 8h30-10h : Recherche de la modélisation du problème.
- **19/06 à 10h : Rendu de la feuille présentant le modèle en séance.**
- 19/06 à 10h : Solution du problème de modélisation.
- 19/06 10h15 - 11h45 : Recherche de la calibration du modèle.
- **19/06 à 11h45 : Rendu de la feuille présentant la calibration du modèle en séance.**
- 19/06 à 11h45 : Solution de la calibration.
- Du 19/06 au 08/07 : Travail sur la simulation, les calculs théoriques, et l'écriture du rapport.
- 26/06 8h30 - 11h45 : Séance de travail programmée obligatoire.
- **08/07 à 00h00 : Date limite du rendu du rapport final contenant le travail effectué sur le site pédagogique.**

4. Modélisation

On suppose que la capacité des stations est illimitée. On suppose aussi que tous les trajets se font d'une station à une autre exclusivement, il n'est pas possible de revenir à la même station.

1. Lisez le document sur le modèle des colonies disponible sur le site pédagogique (extrait du livre *Lectures on stochastic networks* par F. Kelly et E. Yudovin).
2. Inspirez-vous de ce modèle pour proposer une modélisation du Vélib.
   — Donnez l'espace d'états, et sa taille.
   — Tracez le diagramme des transitions possibles (limiter le dessin à quelques stations)
   — Indiquez les taux de transitions.
   — Donnez les paramètres de votre modèle, précisez les unités.

5. Calibration

Suite à la partie modélisation, nous avons donc des paramètres pour le modèle choisi. Nous disposons des données statistiques calculées sur les heures entre 12h et 14h tous les jours de la semaine pendant 2 mois en 2016 sur les 21 stations du Vélib de Rouen (vous pouvez consulter les données temps réels sur vlsstats.ifsttar.fr).

3. Donner les paramètres du modèle en fonction des données fournies.

6. Simulation

On propose d'étudier le Vélib de Rouen dans sa configuration de 2016 qui possédait 21 stations. Pour la simulation, on ne conserve que 5 stations (stations 3 à 7), les paramètres calculés ainsi que les conditions initiales sont donnés dans le fichier Excel Donnees_simulations.xlsx mis à disposition sur le site pédagogique après la séance du 19/06.

4. Simulez les trajectoires du processus de Markov.
5. En déduire la probabilité que chaque station soit vide après 1 heure.
6. Calculer l'intervalle de confiance de ce résultat.
7. Les conditions initiales ont-elles une influence sur le résultat ?
8. Calculez le pourcentage de temps pendant lequel une station est vide.
9. Quel est selon vous le résultat le plus proche de la probabilité stationnaire (entre l'observation de la configuration finale de la question 5 et le pourcentage de temps de la question 8) ?
10. Selon vous, vaut-il mieux augmenter la durée de la simulation (1 heure), ou le nombre de simulations pour améliorer la précision du résultat ?

On rappelle la méthode pour tirer une variable aléatoire discrète. 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.

On rappelle que l'intervalle de confiance d'une quantité $\theta = \mathbf{E}[X]$ estimée sur n tirages $X_1, \ldots, X_n$ de moyenne $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ est l'intervalle de la forme $[\hat{\theta}_n - \varepsilon, \hat{\theta}_n + \varepsilon]$ dans lequel on est sûr à $\alpha\%$ que se trouve θ. On a $\varepsilon = \beta\sigma_n/\sqrt{n}$ avec :

— σ_n l'écart-type empririque sur les tirages :

$$\sigma_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \hat{\theta}_n)^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j)^2 - \frac{n}{n-1}\hat{\theta}_n^2.$$

Si les X_j valent 0 ou 1, alors $X_j^2 = X_j$ et,

$$\sigma_n^2 = \frac{n}{n-1}(\hat{\theta}_n - \hat{\theta}_n^2).$$

— $\beta = 1,96$ pour $\alpha = 0,95$.

Plus de détails sur les intervalles de confiance et comment ils sont calculés sont disponibles sur le site pédagogique.

7. Calcul théorique

On chercher maintenant à valider la simulation par les résultats théoriques

11. Utilisez les équations de trafic pour obtenir les relations entre les α_i (notations du document sur les colonies).
12. On considère qu'il n'y a qu'un seul vélo, quelle est alors la taille de l'espace d'état ?
13. Dans ces conditions (un seul vélo), calculer la probabilité que chaque station soit vide.
14. 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.