MA 332 : Modélisation et analyse des processus stochastiques

DM : processus de Poisson, chaînes de Markov à temps discret et files d'attente

Le but de ce DM est d'illustrer, à l'aide de simulations numériques, les notions relatives à la modélisation et à l'analyse des processus stochastiques abordées en cours et durant les TD. Plus précisément, il vous sera demandé de simuler un processus de Poisson (PP), des chaînes de Markov en temps discret (CMTD) et des exemples de files d'attentes (donc de chaines de Markov en temps continu).
Les objectifs de ce devoir sont :

- de vous permettre de réviser les notions abordées en cours, de les rendre plus concrètes intuitives et, enfin, de vérifier de manière *pseudo-expérimentale* les résultats théoriques établis. La répétition d'un grand nombre d'expériences numériques *pseudo-aléatoires* permet en effet de procéder à des moyennes d'ensemble et d'estimer ainsi (à l'aide du théorème central-limite, notamment) la plupart des grandeurs moyennes caractéristiques des processus stochastiques analysés, avec une précision et un taux de confiance choisis à priori.

- de découvrir et de vous familiariser avec le langage de programmation R (cran.r-project.org) et avec l'environnement de développement RStudio (rstudio.com). R est un logiciel libre destiné à l'origine aux statistiques et à la science des données. RStudio est disponible sous la licence libre (AGPLv3) également. R sera utilisé dans votre formation, notamment dans le cadre du cours MA431 - *Mathématiques appliquées à la sécurité*, mais également dans des cours consacrés aux *réseaux complexes* ou au *machine learning.*

- le cas échéant, de permettre votre évaluation dans le cadre du *plan de continuité d'activité et d'enseignement à distance* mis en place par Grenoble-INP.

Ce DM est à réaliser en binôme. Il doit faire l'objet d'un **compte-rendu écrit qui doit être envoyé par mail, au plus tard le lundi 17**

mai 2021, à l'adresse *laurent.lefevre@lcis.grenoble-inp.fr*. L'objet du mail comportera la mention **CR DM MA332** ainsi que les **noms de famille des deux étudiants du binôme**. Le mail comprendra, en pièces jointes, le compte-rendu (rapport) et les codes sources utilisés pour les simulations.

Le DM est organisé comme suit. Dans la section 1, il vous est demandé de simuler des processus de Poisson à partir d'un générateur de variables aléatoires exponentielles (pour générer les intervalles de temps entre les arrivées succesives). Dans la section 2, vous simulerez les trajectoires de chaînes de Markov en temps discret, à l'aide d'un générateur aléatoire de nombres à valeurs dans $[0,1]$. Dans la section 3 vous simulerez le fonctionnement de deux files d'attente simples, à l'aide du simulateur de processus de Poisson développé précédemment et de l'implémentation de la stratégie de service de la file (en l'occurrence FIFO).

1 Processus de Poisson

Exercice 1 On étudie l'arrivée de véhicules à une barrière d'autoroute. Le temps est mesuré en secondes $[s]$. Les arrivées à la barrière de péage son enregistrées, plusieurs jours d'affilées, pendant des intervalles de temps $[t_0, t_0 + T]$ avec $T = 3600\,[s]$ (une heure), pour une heure de référence t_0 choisie. On note $N(t)$ le nombre de véhicules arrivés entre les instants t_0 et $t_0 + t$, pour $0 \leqslant t \leqslant T$. Après un examen attentifs de ces séries temporelles, il est décidé de modéliser le processus d'arrivées $N(t)$ comme un processus de Poisson de paramètre λ. Le paramètre optimal obtenu avec la méthode du maximum de vraisemblance est $\lambda = \frac{1}{3}\left[s^{-1}\right]$.

(i) Faire une simulation des arrivées sur $[t_0, t_0 + T]$, vérifier la distribution exponentielle des intervalles inter-arrivées et représenter le graphe de l'application $t \mapsto N(t)$ obtenu pour cette réalisation particulière.

(ii) Calculer le nombre moyen d'arrivées dans l'intervalle $[0, T]$. Vous réaliserez une *moyenne d'ensemble* sur plusieurs simulations. Vous vérifierez que le résultat obtenu est en accord avec la théorie (à la précision et au taux de confiance près). Pouvez-vous obtenir un résultat similaire avec moins de simulations, mais sur des intervalles $[t_0, t_0 + T]$ plus long? Pourquoi?

(iii) Vous allez maintenant vérifier la distribution uniforme des instants d'arrivée lorsque le nombre total d'arrivées $N(T)$ est connu. Pour ce faire, vous considérerez le cas $T = 60$ et $N(T) = 20$. Vous procéderez en deux temps :

(a) vérifier que dans le cas général, les instants d'arrivées A_n sont bien distribués selon une loi $\Gamma(\lambda, n)$

(b) vérifier que, si vous vous limitez aux simulations pour lesquelles $N(60) = 20$, les instants d'arrivées A_n sont distribués de manière uniforme sur $[0, 60]$

2 Chaînes de Markov en temps discret - CMTD

Exercice 2 On étudie la propagation d'une maladie dans une population. Une personne peut être dans 4 états : saine et non vaccinée (état 0), vaccinée ou immunisée (état 1), inféctée (état 2) et morte (état 3). On suppose que :

- si une personne est malade, elle a 10 % de "chance" d'être morte la semaine d'après. Le tiers des survivants sera vacciné ou guéri et immunisé la semaine suivante. Les deux tiers restant seront toujours malades (i.e. infectés) et non vaccinés la semaine suivante.
- Si une personne est vaccinée, elle le reste à vie, et elle ne peut tomber malade ou mourir.
- Si une personne est saine une semaine, la semaine d'après, elle tombe malade dans 15 % des cas, va se faire vacciner dans 20% des cas et reste saine dans les autres cas.

Il vous est demandé de répondre aux questions suivantes.Vous aurez à prédire les résultats à l'aide des notions théoriques vues au cours, puis à vérifier ces résultats à l'aide d'expériences numériques. Veillez à choisir un nombre suffisant d'expériences aléatoires pour que les résultats numériques obtenus soient significatifs.

1. Modéliser cette situation par une chaine de Markov dont vous donnerez la matrice de transition et le graphe associé. Vous serez précis sur la définition des grandeurs considérées en précisant notamment, le cas échéant, à quels instants elles sont calculées.
2. Supposons qu'à la semaine 0, la population comporte 100% de personnes saines, non vaccinées. Un mois plus tard (i.e. cinq semaines), on se demande quelle sont les proportions de personnes vaccinées, de malades, saines et de mortes dans la population.
   (a) Faites le calcul théorique de la distribution de probabilité correspondante
   (b) Vérifiez ce résultat par simulation, à l'aide de la répétition d'expériences aléatoires que vous définirez
3. Montrez qu'au bout d'un certain temps, toute personne est nécessairement morte ou vaccinée. Donnez la valeur théorique du temps moyen nécessaire pour qu'une personne soit morte ou vaccinée. Estimez ensuite cette valeur à l'aide d'expériences numériques.

4. Robert est sain. Quelle est la probabilité qu'il meure de la maladie ? Faites le calcul théorique, puis vérifiez le résultat obtenu par simulations

3 Files d'attente

Exercice 3 (Files M/M/k)
Cet exercice vous permettra à la fois de réviser le chapitre sur les chaînes de Markov à temps continu et de vérifier des résultats vus en cours qui concernent le comportement asymptotique (régime permanent) de deux files simples markoviennes, respectivement avec un et deux serveurs. Dans ces files, le processus d'arrivées est un processus de Poisson et les temps de service des clients sont des variables aléatoires de distribution exponentielle. Vous pourrez donc ré-utiliser les codes développés dans la première partie de ce DM. Le travail supplémentaire consistera à représenter la file d'attente (buffer) et à tenir compte de la discipline de service (first come, first served). Une fois cette gestion de la file implémentée vous pourrez simuler le fonctionnement du processus et analyser le comportement de la file au cours du temps, notamment en régime stationnaire, afin de vérifier par des expériences numériques les résultats théoriques vus en cours.

1. On considère l'intervalle de temps $[0, T]$ (le temps est en minutes). Simuler[1] une file M/M/1, en prenant $\lambda = 2$, $\mu = 3$, au cours du temps t, pour $t \in [0, T]$. On pourra prendre par exemple $T = 20$, pour commencer (pour cette question uniquement). Représenter graphiquement les résultats obtenus, par exemple à l'aide du graphe de l'application $t \mapsto n(t)$ où $n(t)$ désigne le nombre total de clients dans la file (buffer et serveur) à l'instant t

2. Calculer empiriquement (i.e. à l'aide d'expériences numériques) $q(T)$, le nombre moyen de clients dans le système (i.e. dans le buffer et le serveur réunis) sur $[0, T]$, et comparer avec la formule du cours sur le nombre moyen de clients dans une file M/M/1 (faire la moyenne d'ensemble sur plusieurs simulations). Prendre plusieurs valeurs croissantes de T, que constate-t-on ?

3. On fixe un T suffisant pour pouvoir considérer que le régime stationnaire est établi (d'après la question précédente). Pour différentes valeurs de $\rho := \lambda/\mu$ dans $[0, 1[$, et notamment pour des valeurs de ρ de plus en plus proches de 1 :

   (a) représenter l'évolution du paramètre stochastique
   $$Q = \lim_{T \to \infty} Q(T)$$

[1] Par *simuler*, on entend produire en machine – à l'aide d'un générateur aléatoire – une réalisation particulière du processus stochastiques $N(t)$ sur l'intervalle de temps $[0, T]$

en fonction des valeurs de ρ

(b) représenter l'évolution du paramètre opérationnel

$$\bar{q}_T := \frac{1}{T} \int_0^T q(t)\, dt$$

en fonction des valeurs de ρ. Le résultat “expérimental” obtenu est-il en accord avec la théorie?

4. Faire le même travail (i.e. répondre aux trois questions précédentes) pour une file M/M/2. On prendra pour commencer les cas $\lambda = 2$ et $\mu = 2$.