

MATH0062-1 - Éléments du Calcul des Probabilités

Projet 2

Généralités

Le deuxième travail du cours d'éléments du calcul des probabilités portera sur un problème plus réaliste que lors du premier projet. Il s'agira de la modélisation du trajet journalier d'un étudiant se rendant à l'ULiège en utilisant le train et le bus. Ce travail a pour but de familiariser les étudiants avec des systèmes de plusieurs variables aléatoires.

Ce travail devra être réalisé par groupe de maximum deux étudiants. Chaque groupe devra rendre d'une part un rapport dont le nombre de pages conseillé (hormis annexes) est de 6 et d'autre part ses codes sources MATLAB. Ce rapport devra respecter les indications du guideline <http://www.montefiore.ulg.ac.be/~lduchesne/proba/guidelines.pdf>. Toutes vos questions pour ce projet doivent être envoyées à l.duchesne@uliege.be.

Le travail devra être rendu sous forme d'archive .zip au plus tard pour le dimanche 28/04/2019 à 23h59 via la plateforme <https://submit.montefiore.ulg.ac.be/>. Pour ce faire, vous devez tout d'abord inscrire votre groupe au projet. Tous les membres du groupe doivent s'inscrire sur la plateforme et apparaître dans le groupe ainsi créé. La démarche à effectuer pour créer et rejoindre un groupe est décrite sur la page <https://submit.montefiore.ulg.ac.be/student/help>.

Notez qu'au delà de la deadline, il ne sera plus possible de rendre le projet.

Présentation du problème

Vous arrivez régulièrement en retard à vos cours du matin et vous vous intéressez dès lors aux différentes causes de ces retards. Vous relevez 3 causes majeures :

1. le train, qui est régulièrement en retard ou annulé ;
2. le bus 48, également souvent en retard ;
3. les conditions exogènes : ces conditions sont indépendantes des transports en commun mais impactent leur fonctionnement, provoquant des retards.

Pour analyser ce phénomène, vous modélisez votre trajet journalier via trois variables aléatoires discrètes : \mathcal{T} pour le train, \mathcal{B} pour le bus et \mathcal{C} pour les conditions exogènes. Les variables \mathcal{T} et \mathcal{B} modélisent les retards possibles des transports en commun le lendemain et la variable \mathcal{C} modélise les différents problèmes exogènes aux transports en commun pouvant survenir le lendemain également. Ces informations sont reprises à la table 1. Par exemple, si la variable \mathcal{B} prend la valeur 3, cela signifie que le bus 48 aura du retard le lendemain. Évidemment, les différents soucis liés au trajet peuvent s'accumuler (le train et le bus peuvent tous les deux avoir du retard par exemple) mais nous supposons qu'il ne peut y avoir qu'un seul retard par jour venant d'une même source, c'est-à-dire que chaque variable aléatoire ne peut prendre qu'une seule valeur à la fois. Par exemple, la variable

\mathcal{T}	Description
1	A l'heure
2	En retard
3	Ne circule pas
4	Retard dû à un vol de câble

\mathcal{B}	Description
1	A l'heure
2	Un peu en avance
3	En retard
4	Complet/ne circule pas
5	Retard dû à un contrôleur

\mathcal{C}	Description
1	Pas de problème
2	Neige
3	Grève nationale

TABLE 1 – Variables aléatoires décrivant les arrivées possibles des transports et les conditions exogènes

\mathcal{C} ne peut pas prendre les valeurs 2 et 3 en même temps ou, autrement dit, il ne peut pas y avoir de la neige et une grève nationale le même jour.

Heureusement, un de vos amis s'est déjà posé la question et a relevé pendant ses années d'études tous les retards des bus et trains chaque jour. De plus, une étude nationale a relevé durant ces mêmes années le nombre de jours de neige et de grève nationale. Toutes ces statistiques ont permis de générer un tableau tri-dimensionnel décrivant la loi de probabilité conjointe du triplet de variables. La première composante de ce tableau correspond à la variable \mathcal{T} , la deuxième correspond à la variable \mathcal{B} et la troisième correspond à la variable \mathcal{C} . Les chiffres associés à chacune des variables aléatoires de la table 1 sont respectés dans le tableau tri-dimensionnel. Ce tableau est donné sous forme d'un fichier MATLAB.

Questions

1. Manipulation des lois de probabilités

- Calculez les lois de probabilités marginales des variables \mathcal{T} , \mathcal{B} et \mathcal{C} .
- Calculez les lois de probabilités conjointes des paires de variables $(\mathcal{T}, \mathcal{B})$, $(\mathcal{T}, \mathcal{C})$ et $(\mathcal{B}, \mathcal{C})$.
- Calculez les lois de probabilités conditionnelles suivantes $P_{\mathcal{T}|\mathcal{B},\mathcal{C}}(\mathcal{T}|\mathcal{B},\mathcal{C})$, $P_{\mathcal{B}|\mathcal{T},\mathcal{C}}(\mathcal{B}|\mathcal{T},\mathcal{C})$ et $P_{\mathcal{C}|\mathcal{T},\mathcal{B}}(\mathcal{C}|\mathcal{T},\mathcal{B})$.
- Que pouvez-vous déduire des lois de probabilités calculées précédemment en termes de relations entre les variables aléatoires ?

2. Probabilités des événements

- Calculez la probabilité qu'au moins un évènement défavorable ait lieu (c'est-à-dire tous les évènements sauf "A l'heure", "Un peu en avance" et "Pas de problème").
- Calculez la probabilité qu'une grève ait lieu, sachant que le train et le bus sont à l'heure ainsi que la probabilité qu'une grève ait lieu, sachant que le train et le bus ne circulent pas. Lequel de ces deux scénarios vous paraît le plus plausible ?
- Sachant que le train est en retard, calculez la probabilité qu'un contrôle dans le bus ait lieu **et** qu'il n'y ait ni grève nationale ni neige ($\mathcal{C} = 1$).

3. Retards moyens

Considérons maintenant le retard de façon plus précise. Les retards engendrés par chaque situation sont présentés à la table 2.

La situation est la suivante : vous partez de chez vous et marchez jusqu'à la gare. Vous êtes sur le quai à 6h59. Les trains partent toutes les heures (7h00, 8h00, etc). Le trajet en train jusqu'à la gare de Liège-Guillemins prend 26min. Ensuite, vous marchez jusqu'à l'arrêt (il faut moins d'une minute) afin de prendre le bus 48 qui passe toutes les 10 minutes (7h30, 7h40, etc). Le trajet en bus jusqu'au Sart-Tilman dure 20 minutes et votre cours commence à 8h00.







Pour plus de simplicité, on fera les hypothèses suivantes :

- vous prenez le premier train/bus qui vient (\mathcal{T} et \mathcal{B} représentent ceux-ci) ;
- les temps de trajet des trains/bus sont constants (26 et 20min respectivement) ;
- lorsque $\mathcal{T} = 3$ ou $\mathcal{B} = 4$, on considère que le train/bus suivant arrivera toujours à l'heure (donc le retard correspond bien à celui indiqué dans la table) ;
- lorsque $\mathcal{B} = 5$, on suppose que tous les bus sont contrôlés et donc il y a bien 15min de retard (le bus suivant qui devrait être là 10min après est aussi en retard et vous ne l'aurez donc pas avant le bus d'origine).

Retard (min.)	
\mathcal{T}	$\mathcal{R}_{\mathcal{T}}$
1	0
2	8
3	60
4	30

Retard (min.)	
\mathcal{B}	$\mathcal{R}_{\mathcal{B}}$
1	0
2	-2
3	5
4	10
5	15

TABLE 2 – Description des différents retards possibles.

- (a)  Calculez, pour chaque moyen de transport, le retard moyen occasionné. Calculez également les variances de ces retards.
- (b) Soit la fonction retard total $\phi(\mathcal{T}, \mathcal{B}, \mathcal{C})$ par rapport au début du cours (8h00) ¹.
 - i.  Calculez l'espérance de $\phi(\mathcal{T}, \mathcal{B}, \mathcal{C})$ ainsi que sa variance.
 - ii. À quelle notion concrète correspond l'espérance de ϕ ?
 - iii. La relation entre $E\{\phi(\mathcal{T}, \mathcal{B}, \mathcal{C})\}$ et les espérances calculées au point 3a est-elle linéaire ? Si tel était le cas, cette relation linéaire s'appliquerait-elle aussi pour la variance et pourquoi ? Comment expliquez-vous cela ?
- (c) Supposons maintenant que vous possédez un réveil intelligent qui connaît la météo et reçoit des informations concernant les grèves (il connaît donc précisément la variable \mathcal{C}). Vous aimeriez le programmer afin qu'il sonne uniquement si la probabilité d'arriver à l'heure est supérieure à 50 % (vous n'allez pas vous lever pour rien !).
 - i.  Calculez l'espérance conditionnelle et la variance conditionnelle de $\phi(\mathcal{T}, \mathcal{B}, \mathcal{C})$ connaissant \mathcal{C} .
 - ii.  Vérifiez le théorème de l'espérance totale.
 - iii.  Vérifiez le théorème de la variance totale.
 - iv.  Pour quelle(s) valeur(s) de \mathcal{C} votre réveil sonnera-t-il ?

1. Remarquez bien que l'on n'a pas simplement $\phi = \mathcal{R}_{\mathcal{T}} + \mathcal{R}_{\mathcal{B}}$.



4. Borne supérieure des retards

En pratique, les retards ne sont bien sûr pas fixes. Les fluctuations sont modélisées par une variance associée à chaque cas. La table 3 détaille l'espérance et la variance des retards pour chaque moyen de transport.

$\mathcal{R}_{\mathcal{T}}$ (min.)			$\mathcal{R}_{\mathcal{B}}$ (min.)		
\mathcal{T}	Esp.	Var.	\mathcal{B}	Esp.	Var.
1	0	1^2	1	0	1^2
2	8	2^2	2	-2	0.5^2
3	60	1^2	3	5	3^2
4	30	20^2	4	10	1^2
			5	15	5^2

TABLE 3 – Espérances et variances des retards

Pour chacun des moyens de transport (indépendamment), calculez une borne supérieure du retard pour chaque situation telle que la probabilité que le retard dans cette situation soit supérieur à la borne soit inférieure ou égale à 5% :

-  dans le cas où vous n'avez aucune information sur la loi de répartition des retards ;
-  dans le cas où vous supposez la répartition normale.
- Comparez les résultats obtenus aux points 4a et 4b. Comment expliquez-vous les différences ? Ce phénomène arriverait-il si on supposait un autre type de distribution ?

Rapport

Le rapport doit contenir les réponses aux questions posées ci-dessus, ainsi que les raisonnements et les calculs symboliques qui vous ont permis de les obtenir. N'oubliez pas d'inclure dans le rapport les valeurs numériques répondant aux questions posées. Les lois de probabilités que vous avez calculées seront formatées sous forme de tableau dans le rapport.

Suggestions :

- Si vous utilisez \LaTeX , vous pouvez utiliser des macros afin d'alléger votre code. Par exemple : `\newcommand{\T}{\mathcal{T}}` qui remplacera chaque instance de \mathcal{T} par \mathcal{T} .
- Il existe également des [scripts](#) permettant d'exporter une matrice de MATLAB vers un tableau \LaTeX .
- Si vous le désirez, vous pouvez utiliser la fonction matlab [bsxfun](#) qui permet d'appliquer une opération (addition, division,...) à 2 matrices dont les dimensions ne correspondent pas au lieu d'appeler plusieurs fois l'opération dans une boucle.

Données

La loi conjointe des trois variables étudiées est donnée sous forme d'un fichier MATLAB nommé `TBC.mat`. Pour charger la matrice en mémoire, il suffit d'entrer la commande suivante : `load('TBC.mat')`. Vous disposerez alors d'une matrice nommée `TBC` qui sera disponible en mémoire et qui contient la loi conjointe.