



PMS apprentissage 2A : examen partiel

Tous documents autorisés

Vous avez accès libre à internet sous réserve que vous ne cherchiez pas à communiquer avec autrui (il est donc notamment interdit de déposer un message sur internet).

L'utilisation de toute solution visant à dissimuler des communications vers l'extérieur de l'ordinateur utilisé est bien évidemment interdite. Des vérifications des données échangées avec l'extérieur pourront être menées.

*Vous pouvez déposer sous Chamilo **au format pdf** tout document que vous jugerez utile dans la rubrique « Travaux », puis « Documents pour le DS ».*

En dehors de ce qui se trouve dans le polycopié, vous devez mentionner la source de tout résultat utilisé pour mener à bien votre travail, et discuter de sa fiabilité.

GRILLE DE NOTATION

***10-13** : maîtrise correcte de la démarche du statisticien : différents outils pertinents sont mis en œuvre dans un ordre raisonnable et pour ce à quoi ils sont destinés. Même si les démarches entreprises n'aboutissent pas forcément aux meilleurs résultats possibles, la rédaction est claire et suffisante pour que le correcteur comprenne bien ce qui a été entrepris, et les réponses données au problème posé sont raisonnables (aucun résultat ou conclusion manifestement erroné ne figure dans le travail effectué).*

***En dessous de 10** : un ou plusieurs des critères mentionnés dans le créneau ci-dessus fait défaut, ou il n'y a pas suffisamment d'éléments dans la copie pour pouvoir les juger. La qualité de la carte de connaissance sera alors un élément particulièrement pris en compte dans la notation.*

***14-16** : en plus des critères du créneau 10-13, les outils statistiques manipulés sont dans l'ensemble bien mis en œuvre même si tout leur potentiel n'est pas forcément exploité.*

***17-20** : tous les outils statistiques pertinents ont été utilisés et une réponse très pertinente à la question posée a été formulée. Ne subsistent que de petites erreurs ou imprécisions sans gravité.*

PREAMBULE

Recopier le texte ci-après suivi de votre signature au début de votre copie :

"Je certifie que ce travail (ce qui figure dans la copie et les documents déposés ce jour sous Chamilo) est le fruit de mon travail personnel exclusivement. Signature: "

EXERCICE 0

Pensez à mettre dans votre carte de connaissance dans votre copie, ou à la rendre sous Chamilo.

La suite au verso...

Dans ce travail, quand vous mettez en œuvre une méthode statistique, exploitez-la au maximum : par exemple, quand vous mettez en place une estimation de paramètre, donnez si c'est possible la qualité de l'estimateur associé (biais, convergence) ; si vous pouvez obtenir la valeur théorique ces éléments, c'est parfait ; à défaut, vous pourrez avantageusement effectuer des simulations sous R pour avoir une idée de ces qualités. Dans ce cas, il sera important de bien expliciter votre protocole de simulation.

Par ailleurs, il n'est pas utile de redémontrer des résultats du cours.

Si par contre vous utilisez un résultat que nous n'avons pas traité en cours, soit vous le démontrez dans votre copie, soit vous déposez sous Chamilo une démonstration du résultat qui vous semble tenir la route en citant sa source.

Vous êtes responsable d'une étude statistique dont le but est de déterminer le support d'une variable aléatoire parente du jeu de données suivant (disponible également sous Chamilo, rubrique documents, fichier Data_partiel.txt), ainsi que la probabilité que cette dernière soit négative.

-0,97	-0,96	-0,93	-0,91	-0,90	-0,87	-0,85	-0,80	-0,75	-0,73
-0,66	-0,65	-0,63	-0,60	-0,58	-0,51	-0,48	-0,45	-0,44	-0,40
-0,37	-0,33	-0,30	-0,25	-0,12	-0,08	0,25	0,41	0,51	0,68

Des experts avec lesquels vous avez discuté rapidement pensent que ces données peuvent être issues d'une loi « triangle » qui a les caractéristiques suivantes :

- Support $[-t ; t]$ où t est un réel positif
- Densité sur le support : $f_x(x) = (t - x)/(2t^2)$.
- Fonction de répartition sur le support : $F_X(x) = 1 - (t - x)^2/(4t^2)$.
- L'espérance vaut $-t/3$.
- La variance vaut $2t^2/9$.

1. Vous commencez par construire un histogramme pour voir si le modèle proposé tient la route (ou pas). Comme La valeur 0 a une grande importance dans cette étude, vous construirez votre histogramme de en imposant que 0 soit une séparation entre 2 classes.
2. Vous disposez de 4 statisticiens qui ont planché sur le sujet. L'un d'eux a proposé de tracer le nuage de points $(x_i ; \sqrt{1 - F_e(x_i)})$, où F_e est la fonction de répartition empirique du jeu de données. Pensez-vous que c'est une bonne idée ?
3. Pour faire des simulations, un autre propose de simuler une variable T qui suit une loi triangle à partir d'une variable U qui suit une loi uniforme sur $[0,1]$ par $T = t - 2t \cdot \sqrt{1 - U}$. Qu'en pensez-vous ?
4. Les 4 statisticiens ont proposé des méthodes différentes pour estimer le paramètre t :
 - Albert propose -3 fois la moyenne empirique
 - Boris propose d'utiliser le maximum du jeu de données
 - Charles propose l'opposé du minimum du jeu de données
 - David propose de diviser par 2 la différence entre le maximum et le minimum du jeu de données
 Quelle est selon vous la meilleure méthode ? Y a-t-il moyen de faire mieux ?
5. Quelle est selon vous la probabilité que la variable parente des données soit négative ?