**ACT-2005-TP1 (Dû le 23 octobre)**

Renseignements généraux :

1) Rédiger avec un document de WORD ou LATEX puis c’est préférable de convertir le document en entier y inclus les annexes à la fin en fichier PDF avant de mettre dans la boite. La boite de dépôt n’accepte que les fichiers PDF ou Word. Les fichiers type mixtes et photos ne seront pas acceptés et ne seront pas corrigés. Les résultats importants dans le corps de votre document, mettre dans l’annexe les programmes R utilisés. Ajouter vos commentaires personnels, réflexions. Voir votre TP comme un document préliminaire que vous pourriez rédiger un rapport plus tard et remettre au département R et D de votre compagnie.

2) Prendre le temps de le bien faire, vous pourriez former une équipe de 2 mais vous devriez travailler ensemble comme une unité. L’utilisation de R est recommandée pour le travail.

3) Pour la fonction de répartition empirique qui est , utiliser la commande **ecdf** de R et pour les percentiles empiriques la commande **quantile**, voir les documentations de R sur ces fonctions. On pourrait ainsi simuler un échantillon de taille m provenant de qui estime .Voici les étapes et c’est la méthode inverse habituelle avec l’utilisation de .

On commence par simuler un échantillon de taille m de la loi uniforme standard

L’échantillon simulé de est donnée par les percentiles empiriques

,…, .

Il y a plusieurs versions de mais ça ne diffère que peu si la taille de n est assez grande, comme R a déjà une version dans le programme quantile, on va utiliser la version de R.

Remarque : Pour ce premier TP l’attention et l’accent sont mis sur les méthodes non paramétriques, la fonction de distribution empirique est la clé pour construire divers estimateurs.

4) Vous annexez vos programmes à la fin du document, pour les simulations par exemple m=50 fois, inclure le premier échantillon obtenu dans l’annexe que vous allez faire référence dans le document. Les correcteurs vérifieront vos procédures d’une manière plus efficace si vous suivez les instructions. Remarque : R est capable de faire des intégrations, la commande integrate, ou quad dans le package pracma à utiliser au besoin pour vérifier vos calculs.

5) A noter lorsqu’on parle de méthode non paramétrique, la distribution empirique est utilisée d’une manière implicite même explicitement elle ne parait pas; voir les diapos de notre premier cours. Par exemple pour estimer E(X) on utilise la moyenne empirique qui est un estimateur non paramétrique, elle est exprimable comme = mais tandis que la moyenne théorique , voir les diapos du premier cours et au besoin l’exercice 3.16 à la page 29 du livre Loss Models (5ème édition).

6) Le travail est un travail en solo ou en équipe de 2, si vous êtes à 2, veuillez inscrire les noms sur la page couverture de votre TP, on va les gérer et vous auriez la note commune pour votre équipe. On ne configure pas comme équipe sur le site du cours mais on va rentrer les notes individuellement pour avoir plus de flexibilité car si on configure en équipe sur le site et plus tard il n’y a plus de changement possible pour toute la session.

Barèmes :

Le TP est noté sur 100 points :

10 points pour la présentation

No 1 (40 points)

1. Graphique de et histogramme (5 points), commentaires (5 points) 10 points
2. Faire le réchantillonage(10 points) , construire l’intervalle de confiance(10 points), commentaire ( 5 points) 25 points
3. Commentaires 5 points

No 2 (30 points)

a) Trouver les estimés de ces espérances (10 points) et leurs valeurs théoriques (5points) 15 points

b) Construire les 2 intervalles de confiance 10 points

c)Commentaires 5 points

No 3 (20 points)

1. Kaplan -Meier (4 points), graphique (4 points) 8 points
2. Greenwood (4 points) et intervalle de confiance (4 points) 8 points
3. Intervalle via transformation log (-log ( . ) ) 4 points

------------------------------------------------Fin des renseignements---------------------------------------------

(THÈME : INFÉRENCES NONPARAMÉTRIQUES)

**No 1.(Estimation non paramétrique et la fonction de distribution empirique)**

1. **Le coefficient d’asymétrie , est une mesure de symétrie ou d’asymétrie pour la distribution. Si , la distribution est symétrique. Simuler n=100 données d’une loi exponentielle avec une loi exponentielle et voir ces 100 données sont disponibles òu on ne connait pas la loi sous-jacente, estimer d’une manière non paramétrique (voir les dia pos de notre premier cours pour les explications au besoin), examiner la valeur obtenue et valider avec l’allure d’un histogramme construit avec les données. Est-ce que c’est concordant la valeur de estimé et le graphique de l’histogramme (étude sur la symétrie de la densité sous-jacente)? Vos commentaires, svp.**

**Faire le graphique de avec R et comme estime , que remarquez- vous sur le graphique de comparé à celui de ? Vos commentaires.**

**b) On veut non seulement un estimé de , mais aussi un estimé pour la variance de votre estimateur. Pour cela, on va utiliser la méthode de rééchantillonnage, appelée ‘bootstrap’ en anglais .Au lieu de simuler une échantillon, vous allez simuler m=50 échantillons**

**de même taille n=100 mais tiré de car on ne connait pas ;pour chaque échantillon calculer le estimé et la variance des estimés à travers ces 50 échantillons serait un estimé pour la variance de , calculer cet estimé de la variance de , et construire un intervalle de confiance au niveau 0.95 pour et commentez.L’estimateur ponctuel de vient du premier échantillon. Voici la formule qui résume**

**.**

**Le ré-échantillonnage vous permet d’obtenir qui est la variance estimée pour votre estimateur qui est calculé avec votre échantillon original. L’estimé de avec échantillon j est dénoté par .La moyenne de ces 50 est dénotée par .**

**c) Maintenant, la loi est théorique est révélée, calculer le coefficient d’asymétrie théorique avec la loi exponentielle de moyenne=1, est ce que c’est compatible avec votre estimé ponctuel et par intervalle de confiance obtenus dans b? Vos commentaires, svp.**

**Remarque : Avec les procédures paramétriques on pourrait faire mieux si l’estimateur non-paramétrique ne performe pas bien mais il faut chercher le modèle qui ajuste bien les données. A venir pour le TP2!**

**No 2. (Vous conservez les données du no 1 et l’énoncé)**

**a)** **Estimer d’une manière non paramétrique. Le tel que c’est défini par le livre aussi appelé `` limited- loss’’ ou perte limitée pour les valeurs des limites u suivantes :**

**u=,,**

**,**

**est définie comme la fonction de répartition d’une loi exponentielle avec**

**Comment varie cette expression estimée comme fonction de u? Parait-elle logique cette variation? Comparer avec les valeurs théoriques .Vos commentaires, svp.**

**b) Par la méthode bootstrap avec m=50, construire un intervalle de confiance pour les primes des limted loss avec et u=.Primes signifient espérances et 2 intervalles de confiance à construire.**

**c) Comparer vos valeurs estimées dans a), les intervalles de confiance obtenus dans b avec les valeurs théoriques des primes du limited -loss provenant de la loi exponentielle standard (. ? Vos commentaires, svp sur les performances de ces estimateurs non-paramétriques.**

**Remarque : En réalité on ne connait que mais pas pour fixer les valeurs de u mais varie légèrement d’un échantillon à un autre, ce qui rend la correction plus difficile lorsqu’on corrige pour tout le monde, on a fixé u en utilisant plutôt.**

**No 3**

**a) Utiliser les données dans l’annexe obtenir l’estimateur de Kaplan-Meier et faire un graphique de cet estimateur.**

**b) Utiliser la formule de Greenwood( p.313 du livre), calculer un intervalle de confiance au niveau 0.95 pour S (91), méthode linéaire qui est la méthode la plus simple.**

**c) Refaire l’intervalle de confiance avec la transformation log (-log) tel que discuté dans livre et dans le cours avec l’énoncé de la question b). Vos commentaires en comparant avec l’intervalle obtenu dans b).**

**Annexe :**

**Voici le jeu de données :**

**30, 40,57,65,65,84+,90,92,98,101**

**Légende :**

**Donnée avec + signifie censurée à droite**

**----------------------------------------------FIN------------------------------------------------------------------**

**BON TRAVAIL et le TP 2 sera centré sur les procédures paramétriques!**