TP Statistique n°2
Simulation de lois

Rappel : taper `krdc vnc://nom_du_serveur` dans un terminal.
Les sujets et les corrigés des TPs sont mis le lendemain des séances sur ma page :
http://www.math.u-bordeaux1.fr/~chabanol/stat.html
La bibliothèque stixbox (`help stixbox`), gratuite (et disponible à l'agrégation, ce qui n'est pas le cas du "toolbox Statistics" de matlab), contient un certain nombre de fonctions statistiques : elle permet de générer un certain nombre de lois, fournit leur densité, leur fonction de répartition, leur fonction quantile, etc. Attention la loi géométrique est sur $\mathbb{N}$ au lieu de $\mathbb{N}^*$ et le deuxième paramètre de la loi normale est son écart-type, pas sa variance!
Lors de ce TP vous allez créer un certain nombre de fonctions supplémentaires qui serviront dans les TP suivants.
Les questions subsidiaires ne sont à faire que si vous avez fini les autres questions...

1. Représentations graphiques d'échantillons statistiques

1.1. **Rappel : Histogramme.** On utilise la commande `histo`. Si X est un vecteur, `histo` trace un histogramme non normalisé. Pour le normaliser, il faut donner 4 paramètres, par exemple `histo(X,10,0,1)`. Le deuxième paramètre est le nombre de classes (ou de colonnes, en anglais "bins" de l'histogramme), le troisième paramètre vaut 0 ou 1 (en général 0 pour une loi continue, 1 pour une discrète mais ce n'est pas crucial), et le quatrième vaut 1 pour indiquer que l'histogramme est normalisé.
1. Si vous ne l'avez pas fait la dernière fois : Générer un échantillon de 1000 réalisations d'une loi normale centrée réduite (utiliser `randn`*). Obtenir son histogramme, et tracer sa densité sur le même graphe (soit avec* `dnorm`*, soit en utilisant la formule de la densité d'une loi normale centrée réduite). Comparer sa moyenne empirique et son espérance, ainsi que son écart-type empirique et théorique.*

1.2. **Fonction de répartition.** Il est souvent plus fiable de comparer des fonctions de répartition empirique et théorique, qu'un histogramme avec une densité (on a vu que le théorème de Glivenko Cantelli garantit une convergence uniforme de l'empirique vers la théorique). Plusieurs moyens sont possibles pour tracer cette fonction de répartition empirique. Si X est un vecteur aléatoire de taille n, on peut par exemple effectuer les opérations suivantes :

`Y=sort(X);`	Trie le vecteur X .
`stairs(Y,[1:n]/n)`	Dessine la fonction de répartition empirique

2. Comprendre pourquoi les instructions précédentes font bien ce qu'il faut. 3. Générer un échantillon de 1000 réalisations d'une loi normale et tracer sa fonction de répartition empirique. Obtenir sur le même graphe la fonction de répartition théorique (Utiliser `pnorm`*)*
4. Ecrire une fonction `plotrep` *qui prend en entrée un vecteur et trace la fonction de répartition empirique correspondante.*

2. Génération aléatoire

2.1. **Transformations plus ou moins élémentaires de la loi uniforme.** *5. Générer un échantillon X de 1000 réalisations d'une loi uniforme sur* $[0,1]$. *Obtenir à partir de X un échantillon de 1000 réalisations d'une loi uniforme sur* $[1,2]$. *Obtenir un échantillon de 1000 réalisations d'une loi uniforme sur* $[0,2]$, *un échantillon de 1000 réalisations d'une loi uniforme sur* $[-1,1]$ *(faire des histogrammes ou tracer les fonctions de répartition pour vérifier)*

*6. Ecrire une fonction **runif** qui prend en argument n, a et b et fournit un échantillon de n réalisations d'une loi uniforme sur $[a, b]$.*
7. Obtenir à partir d'un échantillon d'une loi uniforme sur [0,1] un échantillon de loi de Bernoulli de paramètre $\frac{1}{2}$ (penser aux opérations booléennes !). Calculer sa moyenne et comparer avec son espérance.
*8. Ecrire une fonction **rbernou** qui prend en argument n, m, p et fournit une matrice $n \times m$ de réalisations d'une loi de Bernoulli de paramètre p.*
9. Obtenir à partir d'un échantillon d'une loi uniforme sur [0,1] un échantillon de loi uniforme discrète sur $\{1, 2, \ldots 10\}$ (Indication : `floor(x)` et `ceil(x)` fournissent respectivement l'entier inférieur et supérieur les plus proches d'un réel x) Obtenir son histogramme, calculer sa moyenne et comparer avec son espérance. Remarque : Ici on est dans le cas d'une loi discrète : donc le mieux est de prendre comme nombre de classes dans l'histogramme le nombre de valeurs observées. Ici ce sera également mieux de prendre 1 en troisième paramètre de l'histogramme : essayer avec 0 pour comparer. Enfin, si on veut comparer avec la loi théorique, on dessine souvent la loi d'une variable discrète à l'aide de `stem` plutôt que `plot`.
*10. Ecrire une fonction **runifd** qui prend en argument n, N et fournit un échantillon de n réalisations d'une loi uniforme discrète sur $\{1, 2, ...N\}$.*
*11. Ecrire une fonction **discrete1** qui génère un n-échantillon d'une loi sur $\{1, 2, 3\}$ avec probabilités respectives $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{6}$.*

*12. Question subsidiaire (à faire à la fin si vous avez fini le reste) : écrire une fonction **discrete2** qui prend en paramètre un entier n et un vecteur p et génère un n-échantillon d'une loi sur $\{1, 2, \ldots, \texttt{length}(p)\}$ avec probabilités respectives $p(1)$, $p(2), \ldots p(\texttt{length}(p))$.*

2.2. **Inversion de la fonction de répartition.**
*13. Ecrire une fonction **rexp** qui prend en paramètre un entier n et un réel a et génère un n-échantillon d'une loi exponentielle de paramètre a (attention la fonction* ln *s'appelle `log`). L'utiliser pour générer un échantillon de loi exponentielle de paramètre 2, vérifier en traçant l'histogramme et la densité sur un même graphe, puis la fonction de répartition empirique et la fonction de répartition théorique. Comparer la moyenne empirique et l'espérance.*
*14. Ecrire une fonction **rcauchy** qui prend en paramètre un entier n et génère un n-échantillon d'une loi de Cauchy. Vérifier en traçant la fonction de répartition.*

2.3. **Utilisation des propriétés des lois.** La somme de N variables aléatoires i.i.d. de loi de Bernoulli de paramètre p suit une loi binomiale de paramètres N, p.
*15. Utiliser ce résultat pour écrire une fonction **rbino2** qui prend en paramètres n, N et p et génère un échantillon de n variables de loi binomiale de paramètres N, p (utiliser votre fonction **rbernou**, et penser à la fonction `sum`; faire `help sum` pour savoir ce que fait `sum` lorsqu'on l'applique à un vecteur ou à une matrice).*
Utiliser `type rbinom` pour voir le code effectivement utilisé par stixbox pour générer une loi binomiale. Est-ce la même méthode ? Sinon, quels sont leurs intérêts respectifs ?
16. Question subsidiaire : Utiliser la définition de la loi géométrique comme instant de premier succès pour générer un échantillon de 1000 réalisations de loi géométrique de paramètre $\frac{1}{6}$ (penser à utiliser une boucle `while`). Comparer la moyenne et l'espérance, et tracer sur un même graphe l'histogramme et la loi de probabilité. `rgeom` *utilise-t-il la même méthode ?*
17. Question subsidiaire. Si X suit une loi normale centrée réduite et si a et b sont deux réels, quelle

est la loi de $aX+b$? Utiliser ce résultat pour générer (en utilisant `randn`*) un échantillon de 1000 réalisations d'une loi normale d'espérance* 10 *et de variance* 25. *(Si vous avez le temps, comparer l'histogramme et la densité).* `rnorm` *utilise-t-il la même méthode ?*

2.4. **Méthode par rejet.** Une méthode souvent utilisée pour générer une variable aléatoire de loi uniforme sur un ensemble E un peu compliqué est de générer une variable X de loi uniforme sur un ensemble $F \supset E$ puis de tester si X appartient à E : si oui, on la garde, et sinon on génère une nouvelle variable X jusqu'à tomber dans E. La commande `find` va être très utile pour générer d'un coup beaucoup de telles variables : si X est un vecteur, `find(X)` retourne les indices pour lesquels X est non nul; et donc (par exemple) `J=find(X<1)` retourne l'ensemble des indices pour lesquels X est inférieur à 1, et `X(J)` désigne le vecteur où on a gardé uniquement les composantes inférieures à 1.

18. Générer deux vecteurs X et Y contenant les abscisses et les ordonnées de 10000 points distribués de manière uniforme sur $[-1,1]\times[-1,1]$. Utiliser `find` *pour déterminer l'ensemble des indices i pour lesquels le point de coordonnées $(X(i),Y(i))$ appartient au disque de centre 0 de rayon 1. Obtenir deux sous-vecteurs Xd et Yd ne contenant que les valeurs de X et de Y correspondant à ces indices. Tracer les points de coordonnées Xd et Yd.*

Pour que les points ne soient pas reliés par des segments de droite il faut dire à matlab d'utiliser des symboles (par exemple mettre `'d'` en option de `plot`, qui va utiliser des losanges).

On peut sur la même figure tracer le cercle unité à l'aide de

`t=[0:pi/30:2*pi]; plot(exp(i*t))`

(à condition que i désigne toujours le nombre imaginaire); on peut également utiliser `polar` qui dessine une courbe en coordonnées pôlaires (faire `help polar`)

Quelle est la loi de la taille de Xd et Yd ?

Xd et Yd suivent-elles une loi uniforme ?

Question subsidiaire : déterminer la loi de Xd, et comparer l'histogramme et sa densité, ainsi que sa fonction de répartition empirique et théorique.

Question subsidiaire numéro 2 : $\sqrt{Xd^2+Yd^2}$ suit-elle une loi uniforme ? Comparer l'histogramme et la densité théorique, ainsi que la fonction de répartition empirique et théorique

19. Générer maintenant deux vecteurs indépendants r et t de loi respectivement uniforme sur $[0,1]$ et uniforme sur $[0,2\pi]$ et tracer les points de coordonnées $(r\cos t, r\sin t)$. On n'obtient ainsi que des points dans le disque. La distribution vous semble-t-elle uniforme ?

20. Question subsidiaire. On considère une cardioïde de paramètre 1, d'équation en pôlaire $r = 1+\cos\theta$. On admettra qu'elle est incluse dans $[-\frac{1}{4},2]\times[-3\frac{\sqrt{3}}{4},3\frac{\sqrt{3}}{4}]$. Générer un échantillon de points de loi uniforme sur l'intérieur de la cardioïde et les représenter ainsi que la cardioïde.

Il pourra être utile d'utiliser `cart2pol` qui permet de passer des coordonnées cartésiennes aux coordonnées pôlaires.

Utiliser votre échantillon pour obtenir une approximation de

$$\frac{\int_{\text{intérieur de cardioide}} x\,dx\,dy}{\text{aire(cardioide)}}$$

Comparer avec la vraie valeur .