

TP Statistique n°2

Simulation de lois

Rappel : taper `krdc vnc://nom_du_serveur` dans un terminal.

Les sujets et les corrigés des TP sont mis le lendemain des séances sur ma page :

<http://www.math.u-bordeaux1.fr/~chabanol/stat.html>

La bibliothèque `stixbox` 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!

Tapez `help stixbox` dans matlab pour voir un certain nombre de commandes.

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 (il y a un théorème (de Glivenko Cantelli) qui 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.

5. Question subsidiaire : Générer un échantillon d'une loi binomiale et comparer l'espérance et la moyenne empirique, ainsi que l'écart-type et l'écart-type empirique. Tracer sa fonction de répartition empirique, et comparer avec la fonction de répartition théorique (utiliser `pbinom`).

2. GÉNÉRATION ALÉATOIRE

2.1. Transformations plus ou moins élémentaires de la loi uniforme.

6. 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 $[1, 2]$. Puis 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)

7. Écrire une fonction `runif` qui prend en argument n, a et b et fournit un échantillon de n réalisations d'une loi uniforme sur $[a, b]$.

8. Obtenir à partir d'un échantillon d'une loi uniforme sur $[0, 1]$ un échantillon de loi de Bernoulli de paramètre $\frac{1}{3}$ (penser aux opérations booléennes!). Calculer sa moyenne et comparer avec son espérance.

9. Écrire une fonction `rbernou` qui prend en argument n et p et fournit un échantillon de n réalisations d'une loi de Bernoulli de paramètre p .

10. Obtenir à partir d'un échantillon d'une loi uniforme sur $[0, 1]$ un échantillon de loi uniforme discrète sur $\{1, 2, \dots, 10\}$ (Indication : `floor(x)` et `ceil(x)` fournissent respectivement l'entier inférieur et supérieur les plus proches d'un réel x) Tracer son histogramme ou sa fonction de répartition empirique, 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`.

11. Question subsidiaire. Écrire une fonction `runifd` qui prend en argument n, N et fournit un échantillon de n réalisations d'une loi uniforme discrète sur $\{1, 2, \dots, N\}$.

12. Écrire une fonction `discrete1` qui génère un n -échantillon d'une loi sur $\{1, 2, 3\}$ avec probabilités respectives $\frac{1}{2}, \frac{1}{3}, \frac{1}{6}$.

13. Question très 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 $\{1, 2, \dots, \text{length}(p)\}$ avec probabilités respectives $p(1), p(2), \dots, p(\text{length}(p))$.

2.2. Inversion de la fonction de répartition.

14. On rappelle que la fonction de répartition d'une loi exponentielle de paramètre λ est sur \mathbb{R}^+ $x \mapsto (1 - \exp(-\lambda x))$. Soit X une variable aléatoire de loi uniforme sur $[0, 1]$. Déterminer une fonction g telle que $g(X)$ suive une loi exponentielle de paramètre λ .

Écrire 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.

15. Question subsidiaire. Écrire 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.

16. La fonction ci-dessous qui prend en paramètre un réel $0 < p < 1$ génère une réalisation d'une variable aléatoire. Quelle est sa loi ?

```
function X = mystere(p )
k=1;
while (rand(1)>p)
k=k+1;
end
X=k
end
```

Générez un échantillon de 1000 réalisations, et comparez la moyenne et l'espérance de la loi simulée.

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<0.5)` retourne l'ensemble des indices pour lesquels X est inférieur à 0.5, et `X(J)` désigne le vecteur où on a gardé uniquement les composantes inférieures à 0.5.

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 X_d et Y_d ne contenant que les valeurs de X et de Y correspondant à ces indices. Tracer les points de coordonnées X_d et Y_d .

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 en le considérant comme une courbe paramétrée, à l'aide de

```
t=[0:pi/30:2*pi]; plot(cos(t),sin(t))
```

Quelle est la loi de la taille de X_d et Y_d ?

X_d et Y_d suivent-elles une loi uniforme ?

Question subsidiaire : $\sqrt{X_d^2 + Y_d^2}$ 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 polaire $r = 1 + \cos \theta$. 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.

Utiliser votre échantillon pour obtenir une approximation de

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

Comparer avec la vraie valeur .