Excursions autour de la loi normale

La loi *normale* [dite aussi loi de Gauss, loi gaussienne, voire loi de Laplace-Gauß], notée $\mathcal{N}(0,1)$, est une loi pivotale en théorie des probabilités (théorème de la limite centrale) comme en statistique (validation asymptotique des procédures et approximation de lois complexes). Elle fournit aussi un example simple de loi de famille exponentielle, de loi à paramètre de position-échelle et encore de loi à symétrie sphérique. Nous nous proposons dans ce sujet d'étudier un peu plus en détail cette loi et les conditions de son apparition (ou de sa non-apparition) dans des contextes divers.

1. [T] Trouver la loi de la transformation suivante de deux variables uniformes $\mathcal{U}([0,1])$ indépendantes U_1, U_2,

$$X_1 = \sqrt{-2\log(U_1)}\ \cos(2\pi U_2)\ , \qquad X_2 = \sqrt{-2\log(U_1)}\ \sin(2\pi U_2)\ .$$

2. [T] Déduire de la question précédente l'algorithme de Box–Müller

```
function [x]=myrandn(n)
     u=rand(n,1);
     v=rand(n,1);
     x=sqrt(-2*log(u)).*sin(2*%pi*v)
```

qui permet de générer une loi normale.

Un QQ-plot pour valider le simulateur

3. On cherche à évaluer les performances de cet algorithme, en le comparant à l'algorithme de simulation d'une loi normale implanté sous Scilab, soit la commande de la forme

x = rand(100,1,"normal")

Plus précisemment, on cherche à savoir si l'algorithme rand déjà implanté sous Scilab a un comportement acceptable à l'aune de cet algorithme théoriquement validé dans la question précédente (et inversement)!

Un outil graphique souvent utilisé dans la comparaison de distributions est appelé le *QQ-plot*, où Q représente *quantile.* Étant donné un échantillon iid (indépendant et identiquement distribué), $X_1, \ldots, X_n$, de fonction de répartition F, ce graphique consiste à représenter la suite des points $(i/n, F(X_{(i)}))_{1\leq i\leq n}$, où

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$$

désigne la suite des *statistiques d'ordre associées à l'échantillon* $X_1, \ldots, X_n$, soit la suite réarrangée en ordre croissant

$$X_{(1)} = \min\{X_1, \ldots, X_n\}\,, \quad X_{(2)} = \min\{X_1, \ldots, X_n\}/\{X_{(1)}\}\,, \ldots$$

et

$$X_{(n)} = \max\{X_1, \ldots, X_n\}$$

(3.1) [T] Si X_i suit une loi de fonction de répartition F, continue, montrer que $F(X_i)$ suit une loi uniforme sur $[0,1]$.

(3.2) [T] En déduire que, si F est bien la fonction de répartition des X_i, la suite des $F(X_{(i)}))$ est la statistique d'ordre d'un échantillon uniforme $\mathcal{U}([0,1])$.

(3.3) [T] Donner la loi de $U_{(i)}$ i-ème statistique d'ordre d'un échantillon uniforme $\mathcal{U}([0,1])$ et en déduire sa moyenne.

(3.4) [T] Conclure sur la forme prévisible du graphe des $(i/n, F(X_{(i)}))$, si F est bien la fonction de répartition des X_i et valider ainsi l'idée du QQ-plot.

(3.5) [S] Vérifier que le programme suivant trace le QQ-plot associé à la fonction de répartition de la loi normale:

```
function []=QQplot(x)
   vec1=sortup(x);
   mymean=mean(x)*(x./x);
   mysd=st_deviation(x)*(x./x);
   vec2=cdfnor("PQ",vec1,mymean,mysd);
   dim=size(x);
   vec3=(1:dim(1))/dim(1);
   vec4=[vec2';vec3]';
   vec5=([vec3;vec3])';
   plot2d(vec4,vec5,[-10,1])
```

(3.6) [S] Générer successivement un échantillon de 100 points en utilisant le générateur de Box-Muller donné ci-dessus et celui de Scilab et tracer pour chacun le QQ-plot correspondant. Reprendre l'expérience avec 1000 points et 10000 points.

4. Soit $X_1, \ldots, X_n$ un échantillon iid $\mathcal{E}xp(1)$.

Toutes les lois ne convergent pas vers la loi normale!

(5.1) [T] Montrer que la première statistique d'ordre $x_{(1)}$, valeur minimale de l'échantillon, suit une loi exponentielle $\mathcal{E}xp(n)$, de fonction de répartition $1-\exp(-nx)$. En déduire que $x_{(1)}$ tend presque sûrement vers 0.

(5.2) [T] Déterminer la loi de la n-ème statistique d'ordre $x_{(n)}$ et trouver la normalisation ϱ_n telle que la loi de $\varrho_n x_{(n)}$ se stabilise quand n tend vers l'infini.

(5.3) [S] Evaluer cette stabilisation par une expérience de simulation, en examinant l'évolution des histogrammes de $\varrho_n x_{(n)}$ avec n et l'évolution des QQ-plots associés.

5. Soit $X_1, \ldots, X_n$ un échantillon iid de la loi de Cauchy, de densité sur $\mathbb{R}$

Toutes les moyennes ne convergent pas!

$$\frac{1}{\pi}\frac{1}{1+x^2}$$

et de fonction caractéristique $\varphi(t) = \exp\{-|t|\}$ [résultat admis].

(6.1) [T] Montrer que la moyenne empirique

$$\frac{1}{T}\sum_{t=1}^{T} X_t$$

suit une loi de Cauchy quelque soit T. (*Indication :* Utiliser la fonction caractéristique.)

(6.2) [T] Montrer que la commande Scilab

```
Y=grand(m,1,'nor',0,1)./grand(m,1,'nor',0,1)
```

permet de simuler un échantillon de taille m d'une loi de Cauchy.

(6.3) [S] En utilisant ce simulateur, représenter l'évolution de la moyenne empirique quand T varie de 1 à 1000 pour visualiser le manque de convergence de cette moyenne vers une constante.

(6.4) [S] Sur la même expérience de simulation (c'est à dire en utilisant le même échantillon de Cauchy), représenter [en fonction de T] l'évolution de la moyenne dite *élaguée*

$$\tau_T = \frac{X_{(2)} + \cdots + X_{(T-1)}}{T-2},$$

moyenne empirique de l'échantillon auquel on a oté ses deux points extrêmes. Cette statistique τ_T peut-elle satisfaire un théorème de la limite centrale ?