Probabilités TP 3
Illustration numérique des théorèmes de convergence

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

1. Mise en évidence d'une convergence presque sûre

De manière générale, pour illustrer une convergence presque sûre, on fait une expérience (qui correspond à un ω fixé, et on observe que la suite $X_n(\omega)$ converge.

1.1. **Illustration de la loi des grands nombres.** Rappel : D'après la loi forte des grands nombres, si $(X_n)_{n\in N}$ est une suite de variables aléatoires indépendantes intégrables, de même loi, d'espérance m, alors pour presque tout ω, $\frac{\sum_{i=1}^n X_i(\omega)}{n}$ converge vers m : la moyenne arithmétique converge vers l'espérance.
D'autre part, d'après le théorème de Glivenko-Cantelli (qui se déduit de la loi forte des grands nombres) les fonctions de répartition empiriques d'un échantillon de variables aléatoires indépendantes de même loi convergent vers la fonction de répartition théorique.
On va par exemple travailler avec une variable aléatoire X de loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$ (on va simuler un lancer de dés). On va donc se servir de notre fonction `runifd`.

(1) *Quelle est l'espérance m de X ?*

(2) *Générer un échantillon de X de taille* 5000*; utiliser* `cumsum` *pour obtenir un vecteur avec les valeurs successives des $\sum_{i=1}^n X_i$ pour n allant de 1 à 5000, puis un vecteur avec les valeurs successives des $\frac{\sum_{i=1}^n X_i}{n}$ pour n allant de 1 à 5000.*

(3) *Représenter sur un même graphe les $\frac{\sum_{i=1}^n X_i}{n}$ en fonction de n, ainsi qu'une droite horizontale correspondant à l'espérance : on peut observer la convergence.*

(4) *Constater en traçant les courbes que les $\sqrt{n}(\frac{\sum_{i=1}^n X_i}{n} - m)$ restent d'ordre 1 (mais ne convergent pas p.s.), alors que les $n(\frac{\sum_{i=1}^n X_i}{n} - m)$ divergent et que les $n^{0.1}(\frac{\sum_{i=1}^n X_i}{n} - m)$ tendent vers 0 p.s. : on voit le rôle du $\sqrt{n}$ présent dans le théorème central limite.*

(5) *Question subsidiaire. Illustrer le théorème de Glivenko Cantelli en représentant les fonctions de répartition empiriques de l'échantillon des X_i dont on a gardé respectivement que les* 100, 500,1000 *et* 5000 *premiers termes, ainsi que la fonction de répartition théorique (vous pouvez utiliser votre fonction* `plotrep`*).*

(6) *Question subsidiaire. Refaire les questions 2 et 5 avec une variable aléatoire qui suit une loi uniforme sur* $[0, 1]$.

1.2. **Convergence presque sûre mais pas vers une constante.** La loi forte des grands nombres garantit la convergence presque sûre vers une constante (l'espérance). Mais on peut aussi dans d'autres situations converger vers une variable aléatoire non constante. Dans ce cas-là, pour observer la loi de la variable aléatoire limite, on doit répéter l'expérience plusieurs fois.

(7) *Générer un 100-échantillon E de loi de Bernoulli de paramètre* 0.5.

(8) *Obtenir un vecteur contenant les $\sum_{i=1}^n \frac{E_i}{2^i}$, pour n allant de 1 à 30.*

(9) *La suite* $u_n = \sum_{i=1}^{n} \frac{E_i}{2^i}$ *vous semble-t-elle converger ? Vers quelle valeur ? (faites une figure avec matlab)*

(10) *Refaites l'expérience avec un autre échantillon* : on observe encore une convergence (il y a encore convergence p.s.), mais la limite est différente : elle est aléatoire.

(11) Pour observer la loi de cette valeur limite, on va en fait illustrer la convergence en loi. Pour cela, on a besoin de beaucoup de tels échantillons de E (par exemple 2000), pour lesquels on va retenir uniquement une approximation de la valeur "limite" (par exemple $\sum_{i=1}^{30} \frac{E_i}{2^i}$), et on va observer comment sont réparties ces valeurs limites.
Mettre dans un vecteur les 2000 valeurs successives obtenues pour $\sum_{i=1}^{30} \frac{E_i}{2^i}$ *lorsqu'on effectue 2000 fois les questions 7 et 8. Tracer la fonction de répartition empirique de ces valeurs. A votre avis, quelle est la loi limite ? (NB : on l'a prouvé en TD)*

(12) *Question subsidiaire. Mêmes questions si* E *ne suit pas une loi de Bernoulli mais une loi uniforme sur* $\{-1, 1\}$.
Pour les curieux : la série $\sum_n \frac{A_n}{3^n}$ où les A_n sont i.i.d de loi uniforme sur $\{0, 2\}$ converge en loi vers une loi uniforme sur l'ensemble de Cantor.

2. Mise en évidence d'une convergence en loi

Pour observer une convergence en loi, on répète plusieurs fois l'expérience et on cherche à observer la convergence des fonctions de répartition empirique (ou même théorique!) vers la fonction de répartition limite.

2.1. **Illustration du théorème central limite.** D'après le théorème central limite, si (X_n) est une suite de variables aléatoires indépendantes de même loi et d'espérance μ et de variance σ^2 finies, $\sqrt{n}(\bar{x} - \mu)/\sigma$ converge en loi vers une loi normale centrée réduite (on a noté $\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}$) . Cette quantité reste donc aléatoire, mais on sait comment elle est distribuée asymptotiquement.

(13) On reprend le lancer de dé de tout à l'heure, et on va effectuer plusieurs fois (N_{obs} fois) l'expérience (virtuelle) de le lancer n fois.
X suit donc comme à la question 1 une loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$, d'espérance m et d'écart-type σ.
Calculer σ.
Construire une fonction `TCLde` *qui prend en entrée deux entiers* n *et* N_{obs}, *génère successivement* N_{obs} n*-échantillons de* X, *calcule pour chacun de ces échantillons la valeur "finale"* $\frac{\sqrt{n}}{\sigma}(\bar{x} - m)$, *stocke ces valeurs dans un vecteur, et enfin trace sur un même graphique la fonction de répartition empirique de ces grandeurs ainsi que le graphe de la fonction de répartition empirique de la loi normale centrée réduite.*
Utiliser cette fonction avec $Nobs = 1000$ *et* $n = 10, 50, 300, 1000$.

(14) *Question subsidiaire.* Un autre théorème de convergence en loi dit que si λ est un réel positif, une suite de variables aléatoires de loi binomiale de paramètre $(n, \frac{\lambda}{n})$ converge en loi vers un Poisson de paramètre λ.

(a) *Soit* Z *une variable suivant une loi de Poisson de paramètre 2. Créer un vecteur contenant les* $P(Z = k)$ *pour* k *variant de* 0 *à* 8 *(l'instruction pour la factorielle est* `factorial`*). Tracer la fonction de répartition d'une loi de Poisson de paramètre 2 (utiliser ce vecteur, et l'instruction* `cumsum`*).*

(b) *Générer un 5000 échantillon de loi binomiale de paramètre* $(10, \frac{2}{10})$ *(on prend donc* $n = 10$*). Tracer la fonction de répartition empirique et comparer avec la fonction de répartition d'une loi de Poisson de paramètre 2.*

(c) *Faire de même avec* $n = 100$ *et* $n = 1000$ *et observer la convergence.*

(d) Tracer sur un même graphique les fonctions de répartition théorique de la loi binomiale de paramètre $(n, \frac{2}{n})$ pour $n = 10, 100$ et 1000 et observer là encore la convergence.

(15) On peut aussi observer la convergence en loi de certains processus qu'on appelle des chaînes de Markov. La fonction ci-dessous génère un pas d'une chaîne de Markov sur $\{1, 2, 3\}$. Imaginons par exemple que Toto peut être dans trois états : 1 (bonne santé), 2 (enrhumé), 3 (malade). Le 1er janvier il est en bonne santé. Son état le jour $n+1$ dépend de son état au jour n et pas des jours précédents : s'il est en bonne santé, il le reste le lendemain avec probabilité $\frac{5}{6}$, il devient malade avec probabilités $\frac{1}{12}$. S'il est malade, il le reste avec probabilité $\frac{3}{4}$ et devient en bonne santé avec probabilité $\frac{1}{4}$. S'il est enrhumé, il guérit avec probabilité $\frac{1}{4}$ et devient malade avec probabilité $\frac{1}{4}$. La fonction ci-dessous prend l'état le jour n et donne l'état le jour $n+1$.

```
function Y = Markov( X )
U=rand(1);
if (X==1) Y=1*(U<5/6) + 3*(U>11/12) + 2*(U>5/6&& U<11/12); end
if (X==2) Y=1*(U<1/4) + 3*(U>3/4) + 2*(U>1/4&& U<3/4); end
if (X==3) Y=1*(U<1/4) + 3*(U>1/4); end
end
```

Vérifier que la fonction ci-dessus génère bien ce qu'on veut.
Générer un 5000 échantillon de la chaîne à l'instant 900 et à l'instant 1000 et mettre en évidence une convergence en loi.
La théorie des chaînes de Markov prédit une convergence en loi vers la loi $(\mu(1), \mu(2), \mu(3))$ telle que ${}^t(\mu(1), \mu(2), \mu(3))$ est vecteur propre pour la valeur propre 1 de la transposée de la matrice de transition de la chaîne, qui est ici $\begin{matrix} \frac{5}{6} & \frac{1}{12} & \frac{1}{12} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} \end{matrix}$
Vos observations sont-elles en accord avec ce résultat ?