T.P. n°7

Convergence en loi, approximation, TCL
Estimation, Intervalles de confiance
Méthode de Monte-Carlo

Illustration(s) de convergence(s) en loi

Exercice 1. (Retour sur **EDHEC 2013**) On a pu voir en classe que, si (U_n) est une suite de v.a. indépendantes de même loi uniforme $\mathcal{U}([0;1])$, alors,

$$I_n = \min(U_1, ..., U_n) \xrightarrow{\mathcal{L}} 0.$$

Le programme suivant vise à illustrer cette observation. Recopier et compléter les instructions.

```
//simulation de la variable I_n
function y=edhec2013(n)
    U=zeros(1,n);
    for k=1:n
        U(k)=..................
    end
    y=............................
endfunction

//fréquence de la valeur obtenue par I_n sur un échantillon de taille 100
function y=freq_I(n)
    U=zeros(1,100)
    for k=1:100
        U(k)=..................
    end
    y=.............................
endfunction

I=feval([1:10:1000], freq_I)
plot2d([1:10:1000], I, style=5)
```

Exercice 2. Soit (X_n) une suite de v.a. indépendantes de même loi uniforme $\mathcal{U}([0;1])$.

(1) Rappeler l'énoncé du **Théorème central limite**.
(2) Justifier que

$$\sqrt{\frac{12}{n}} \sum_{k=1}^{n} X_n - \sqrt{3n} \xrightarrow{\mathcal{L}} X, \qquad \text{où} \quad X \hookrightarrow \mathcal{N}(0,1).$$

(3) Utiliser cette observation pour compléter le programme `SciLab` suivant permettant de simuler la variable

$$S_n^* = \sqrt{\frac{12}{n}} \sum_{k=1}^{n} X_n - \sqrt{3n}$$

```
function y=normale(n)
    s=0;
    for k=1:n
        s=s+..........................
    end
    y=................................
endfunction
```

(4) Compléter le programme suivant pour représenter simultanément le diagramme obtenu avec un échantillon de taille $N = 1000$ et $n = 12$ ainsi que la densité de la loi normale $\mathcal{N}(0;1)$.

☞ On en profitera pour rappeler comment fonctionne la commande `dsearch( )`.

```
n=12;
//simulation d'un échantillon de S_n*
S=zeros(1,1000);
for k=1:1000
    ................................
end

//répartition des effectifs par classe, calcul des fréquences et diagramme
classes=-10:.2:10
nbc=....................; //nombre de classes
[i, eff]=dsearch(S,classes);
freq=...............................
bar(................., ..................)

//courbe de la densité de la loi normale centrée réduite
Y=feval(..............)
plot2d(........, ......., style=5) // style=5 correspond à une courbe rouge
```

(5) Reprendre le programme suivant afin de comparer la fonction de répartition et le diagramme des effectifs cumulés.

Exercice 3. (D'après **ECRICOME 2017**) Dans le Devoir Maison n°9, on a notamment montré (Exercice 5), que la suite de variables aléatoires (T_n) définie par

$$T_n(\Omega) = [\![2; n+1]\!], \qquad PT_n = k) = \frac{(k-1)}{n^k}\binom{n+1}{k}, \qquad k \in [\![2; n+1]\!].$$

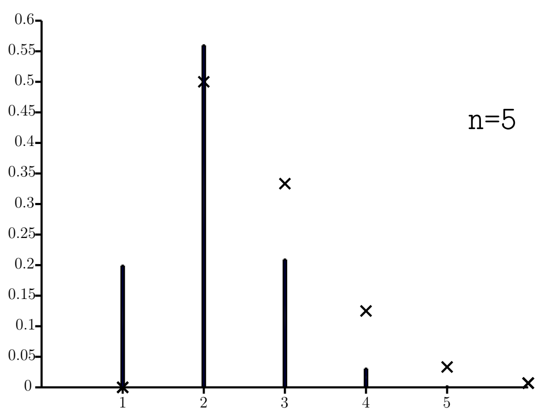
convergeait en loi vers une v.a. Y de loi

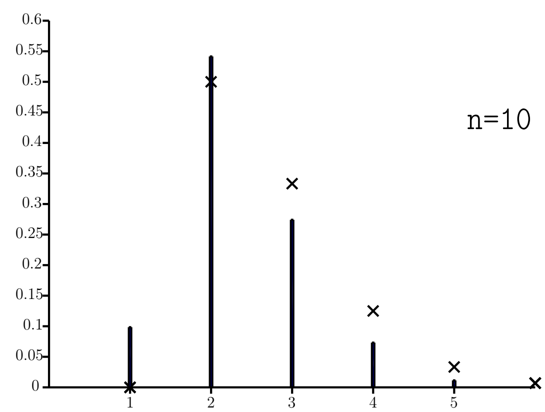
$$Y(\Omega) = [\![2; +\infty[\![, \qquad P(Y = k) = \frac{k-1}{k!}, \qquad k \in [\![2; +\infty[\![.$$

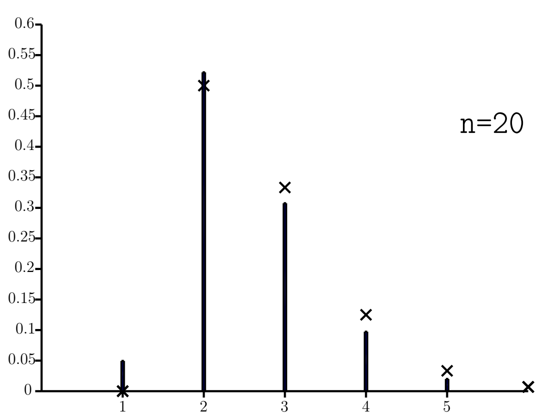
(1) On suppose qu'on dispose, sous `SciLab`, d'une fonction `y=T(n)` permettant de simuler la variable T_n. On écrit alors le script ci-dessous.

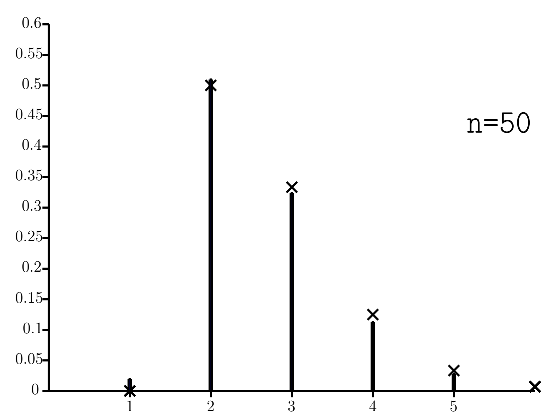
```
function y=freqT(n)
    y = zeros(1,n)
    for i=1:100000
        k = T(n)
        y(k) = y(k)+1
    end
    y = y/100000
endfunction

function y=loitheoY(n)
    y = zeros(1,n)
    for k=1:n
        y(k) = (k-1)/prod(1:k)
    end
endfunction

clf()
n = input('n=?')
plot2d(loitheoY(6),style=-2)
x = freqT(n)
bar(x(1:5))
```

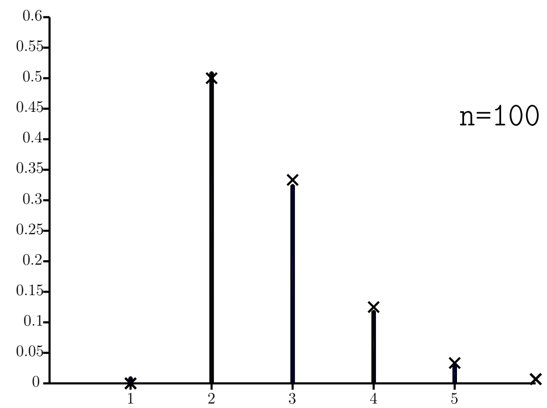
L'exécution de ce script pour les valeurs de n indiquées a permis d'obtenir les graphes ci-dessous:

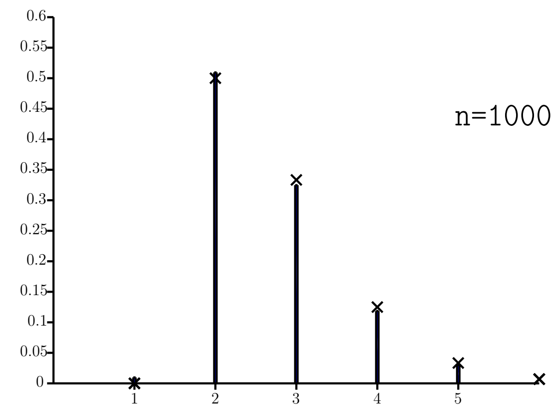












(a) Expliquer ce que représentent les vecteurs renvoyés par les fonctions `freqT` et `loitheoY`. Comment ces vecteurs sont-ils représentés graphiquement dans chaque graphique obtenu ?
(b) Expliquer en quoi cette succession de graphiques permet d'illustrer la convergence en loi de (T_n) vers Y.

Exercice 4. (Loi du khi-deux) Soient $X_1, X_2, ..., X_n$ des v.a. indépendantes suivant une loi normale $\mathcal{N}(0,1)$. On dit que la variable T_n définie par

$$T_n = \sum_{k=1}^{n} X_k^2$$

suit la loi du khi-deux à n degrés de liberté, ce qu'on note $T_n \hookrightarrow \chi_n^2$.

(1) Écrire une fonction `y=khi_deux(n)` simulant la loi du khi-deux à n degrés de liberté. Stocker 10000 réalisations pour $n = 6$ dans un vecteur `echantillon`.
(2) Donner une estimation du réel t tel que $P(T_5 < t) = 0,95$.

☞ On commencera par taper l'instruction `echantillon=gsort(echantillon, 'g', 'i')` et à cette occasion rappeler ce que fait et comment fonctionne la commande `gsort( )`

(3) On considère une urne contenant n jetons numérotés 1 à n. On tire k jetons un par un avec remise. On note, pour tout i entre 1 et n, N_i le nombre de jetons piochés portant le numéro i. On pose ensuite

$$X_k = \frac{n}{k} \sum_{i=1}^{n} \left(N_i - \frac{k}{n} \right)^2.$$

☞ On peut montrer que $X_k \xrightarrow[k \to +\infty]{\mathcal{L}} \chi_{n-1}^2$.

(a) Écrire une fonction `N=jeton(n,k)` renvoyant un vecteur N dont les composantes sont $N_1, N_2, ..., N_n$.
(b) On suppose $k = 6$. Utiliser la fonction précédente pour calculer X_{10000}.

Estimation(s)

Exercice 5. (D'après **ECRICOME 2020**) Soit a un réel strictement positif. On note f la fonction définie sur $\mathbb{R}$ par

$$f(x) = \begin{cases} 0, & \text{si } x < a \\ \dfrac{3a^3}{x^4}, & \text{si } x \geq a \end{cases}$$

(1) Montrer (à nouveau) que f est bien une densité de probabilité. On note X une v.a de densité de f. Vérifier que X admet espérance et variance, que

$$E(X) = \frac{3a}{2}, \qquad V(X) = \frac{3a^2}{4},$$

et que la fonction de répartition de X est donnée par

$$F_X(x) = \begin{cases} 0, & \text{si } x < a \\ 1 - \dfrac{a^3}{x^3}, & \text{si } x \geq a \end{cases}$$

(2) Montrer que $Y = aU^{-1/3}$ et X suivent la même loi, où $U \hookrightarrow \mathcal{U}([0;1])$. En déduire l'écriture d'une fonction `SciLab` d'en-tête `function Y=simulX(a, m, n)` qui renvoie une matrice à m lignes et n colonnes dont chaque composante est une simulation indépendante de X.

On considère alors un $n-$échantillon $(X_1, ..., X_n)$ de X.

(3) On pose $V_n = \dfrac{2}{3n} \displaystyle\sum_{k=1}^{n} X_k$.
Montrer que V_n est un estimateur sans biais de a et que son risque quadratique est égal à

$$r(V_n) = \frac{a^2}{3n}.$$

(4) On pose $W_n = \min(X_1, ..., X_n)$.
  (a) Déterminer la fonction de répartition de W_n puis une densité.
  (b) En déduire que W_n admet une espérance que l'on précisera.
  Déterminer un réel λ_n tel que $Z_n = \lambda_n W_n$ soit un estimateur sans biais de a.

  (c) Montrer que le risque quadratique de Z_n est égal à

$$r(Z_n) = \frac{a^2}{3n(3n-2)}$$

(5) (a) Compléter la fonction ci-dessous afin qu'elle réalise m simulations de la variable aléatoire V_n et renvoie les résultats obtenus sous forme d'une matrice ligne à m éléments.

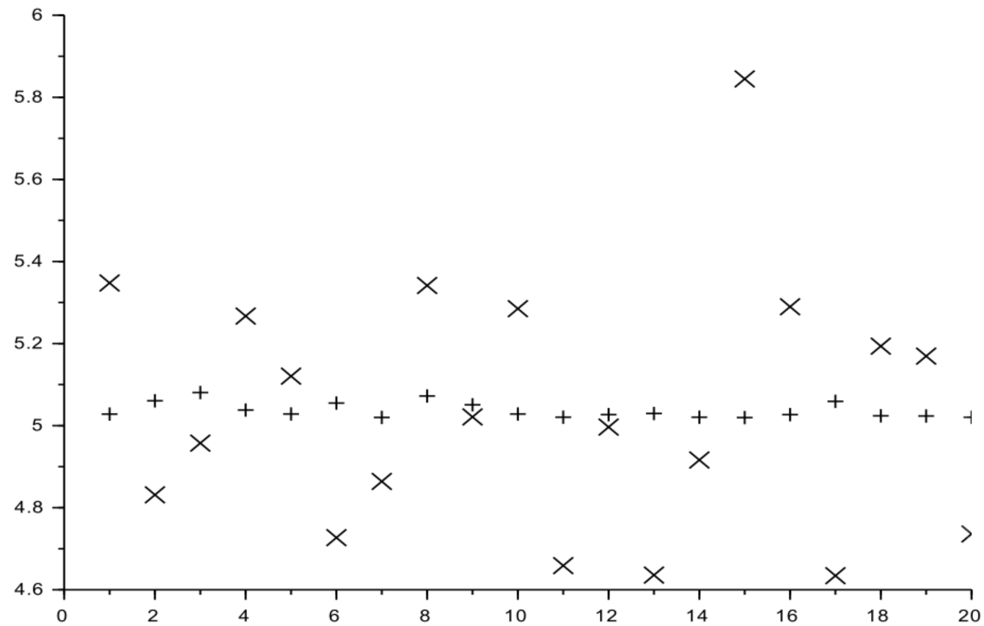
```
function V=simulV(a, m, n)
    X=simulX(a, m, n)
    V=zeros(1, m)
    for k=....................
        V(k)=..................
    end
endfunction
```

Pour la suite, on prend $n = 100$ et on suppose que l'on dispose d'une fonction similaire `simulW` permettant d'obtenir m simulations de la variable aléatoire $\lambda_n W_n$.

  (b) Compléter les lignes ci-dessous pour écrire le script qui a permis d'obtenir le graphique présenté.

```
W = simulW(..., ..., ...)
V = simulV(..., ..., ...)
plot2d(..., style = -1)
plot2d(..., style = -2)
```

On justifiera la réponse pour les deux dernière lignes.



La méthode de Monte Carlo

Le terme *méthode de Monte-Carlo* désigne toute méthode visant à calculer une valeur numérique en utilisant des procédés aléatoires, c'est-à-dire des techniques probabilistes.

Notamment, la méthode de Monte-Carlo (**MMC**) est souvent utilisée pour déterminer une valeur approchée d'une intégrale (difficile) à calculer. Pour ce faire, on agit comme suit :

- on fait apparaître l'intégrale en question sous la forme d'une espérance, notamment grâce au théorème de transfert
$$m = \int f(t)\mathrm{d}t = E(f(X))$$
où X suit une loi (usuelle) que l'on sait simuler;
- on approche cette espérance à l'aide de la **loi faible des grands nombres** (LfGN), par l'estimateur de moyenne empirique
$$T_n = \frac{1}{n}\sum_{k=1}^{n} f(X_k),$$
où $(X_1, X_2, ..., X_n)$ est un $n-$échantillon de X et pour n *assez grand.*

☞ La LfGN garantit la convergence. Toutefois, elle ne renseigne pas sur la rapidité de celle-ci ni sur l'erreur d'approximation commise avec cette estimation. C'est pourquoi on fait appel au **TCL** pour obtenir un *intervalle de confiance asymptotique* pour la valeur de cette intégrale/espérance. Plus précisément, l'erreur commise est
$$\varepsilon_n = E(f(X)) - \frac{1}{n}\sum_{k=1}^{n} f(X_k).$$
Le TCL donne alors un intervalle de confiance (asymptotique) au niveau α pour $m = E(f(X))$, construit sur une réalisation de T_n
$$\left[T_n - \frac{\sigma}{\sqrt{n}}t_\alpha; \quad T_n + \frac{\sigma}{\sqrt{n}}t_\alpha\right].$$

où $t_\alpha = \Phi^{-1}(1-\alpha/2)$ et σ^2 est la variance de $f(X)$.

☞ On accède, sous `SciLab` à t_α avec la commande

```
cdfnor("X", 0, 1, 1-alpha/2, alpha/2)
```

☞ Pour $\alpha = 5\%$, on *sait* que $t_\alpha = 1.96$.

Il est assez naturel, si on ne connait pas m, qu'on ne connaisse pas σ non plus; on remplace alors la variance σ^2 par la variance empirique[1] σ_n^2

$$\sigma_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(f(X_k) - T_n\right)^2.$$

Exercice 6. Le but de cet exercice est d'obtenir une estimation de l'intégrale

$$I = \int_0^1 \frac{\mathrm{d}t}{1+t^2}.$$

(1) Déterminer une fonction f et une variable aléatoire X telle que $I = E(f(X))$.

(2) Compléter le programme ci-dessous avant qu'il renvoie une estimation de I par la **MMC** obtenue avec un $n-$échantillon de X.

```
function y=f(x)
    y=.......................
endfunction

function y=I(n)
    U=grand(1, n,  ............, ...............,  ............)
    y=..........................
endfunction
```

(3) Quelle(s) instructions permettent d'avoir une estimation de la variance de $f(X)$?

(4) Donner un intervalle de confiance (asymptotique) au risque 5% pour I, construit dur l'estimation précédente. Quel est, en fonction de n, l'ordre de grandeur de l'erreur d'approximation?

(5) **Révisions.** On propose une autre méthode visant à l'estimation de l'intégrale précédente.[2].

  (a) Rappeler le résultat des *sommes de Riemann.*

  (b) Recopier et compléter le programme ci-contre.

```
function y=est_rect(n)
    y=.........................
    for k=1:n
        y=.....................
     end
     y=........................
endfunction
```

  (c) Le sujet **EDHEC 2008** (notamment proposé sous le nom de Problème 2 ici.) vise notamment à redémontrer ce résultat et donne même une estimation de l'erreur.

$$\left|\int_0^1 f(t)\mathrm{d}t - \frac{1}{n}\sum_{k=0}^{n} f\left(\frac{k}{n}\right)\right| \leq \frac{\max_{t\in[0;1]}|f'(t)|}{2n}.$$

---

[1]on choisit ici l'estimateur non biaisé

[2]Cette méthode s'appelle *méthode des rectangles*

Proposer alors une estimation de I à 10^{-4} près.

Remarque. Si les outils du cours d'ECE n'en permettent pas le calcul, ceux d'ECS permettent d'obtenir
$$I = \frac{\pi}{4} \simeq 0.7853892.$$

Exercice 7. (Extrait de **HEC 2015**) Soit F la fonction définie sur $\mathbb{R}$ par
$$F(x) = \exp\left(e^{-\lambda x}\right).$$

(1) Justifier que F est de classe C^∞ sur $\mathbb{R}$ et que F réalise une bijection de $\mathbb{R}$ sur l'intervalle $]0,1[$.
(2) En déduire que F est la fonction de répartition d'une variable aléatoire T admettant une densité f_T, continue sur $\mathbb{R}$ que l'on déterminera; on dit que T *suit la loi de Gumbel de paramètre* λ.
(3) Établir l'existence de l'espérance $E(T)$ de la variable aléatoire T.
(4) On pose: $Z = -\frac{1}{\lambda}\ln(\lambda X)$. Montrer que les variables aléatoires Z et T sont de même loi.
(5) Justifier l'égalité
$$E(T) = -\frac{1}{\lambda}\int_0^{+\infty} e^{-t}\,\ln t\,\mathrm{d}t.$$
(6) À l'aide de la concavité de la fonction ln sur $\mathbb{R}_+^*$, établir l'inégalité: $E(T) \geq 0$.
(7) On suppose dans cette question que $\lambda = 1$.

   (a) Expliciter la bijection réciproque G de la fonction F.
   (b) On considère le programme `Scilab` suivant:

```
x=linspace(-2,2,400)
y=(exp(-exp(-x)))
plot(x,y)
plot(y,x)
```

      (i) Le réel 0 fait-il partie des nombres renvoyés par la commande `x=linspace(-2,2,400)`?
      (ii) Quel sera le résultat de l'exécution de ce programme?

   (c) Soit U une variable aléatoire suivant la loi uniforme sur l'intervalle $]0,1[$.
   (d) Quelle est la loi de la variable aléatoire $G(U)$?
   (e) Établir l'inégalité: $E(T) \leq 1$.
   (f) Par une méthode de votre choix, écrire en `Scilab` les commandes qui permettent de simuler la loi de T.
   (g) Écrire en `Scilab` les commandes qui permettent de renvoyer une valeur numérique approchée de $E(T)$ en utilisant la méthode de Monte-Carlo.