Soit n un entier naturel non-nul.
On effectue une série illimitée de tirages d'une boule avec remise dans une urne contenant n boules numérotées de 1 à n.
Pour tout entier k non-nul, on note X_k la variable aléatoire égale au numéro de la boule obtenue au $k^{\text{ème}}$ tirage.
Pour tout entier naturel k non-nul, on note S_k la somme des numéros des boules obtenues lors des k premiers tirages : $S_k = \sum\limits_{i=1}^{k} X_i$.
On considère enfin la variable aléatoire T_n égale au nombre de tirages nécessaires pour que, pour la première fois, la somme des numéros des boules obtenues soit supérieurs ou égale à n.
Exemple : avec $n = 10$, si les numéros obtenus aux cinq premiers tirages sont dans cet ordre ?????????? alors ....

Partie A

1. *(Pour $k \in \mathbb{N}^*$, déterminer la loi de X_k, ainsi que son espérance.)*
La variable X_k modélise le numéro de la boule tirée au $k^{\text{ème}}$ tirage.
Les valeurs possibles sont : $X_k(\Omega) = [\![1, n]\!]$.
Il y a équiprobabilité, donc X_k est de loi uniforme : $X_k \hookrightarrow \mathcal{U}\big([\![1, n]\!]\big)$.

2. **a)** *(Déteminer $T_n(\Omega)$.)*
La variable T_n modélise le nombre de tirages nécessaires pour que le « score cumulé » S_k atteigne n.
les tirages sont indépendants *(avec remise)*, et chacun apporte un score $X_k \in [\![1, n]\!]$.
Il faut donc entre 1 et n tirages que le score cumulé atteigne n, soit : $T_n(\Omega) = [\![1, n]\!]$.

b) *(Calculer $\mathbb{P}(T_n = 1)$.)*
On a l'égalité d'événements : $[T_n = 1] = [S_1 = n] = [X_1 = n]$.
Ainsi : $\mathbb{P}(T_n = 1) = \mathbb{P}(X_1 = n) = \frac{1}{n}$. *(car $X_1 \hookrightarrow \mathcal{U}([\![1, n]\!])$.)*

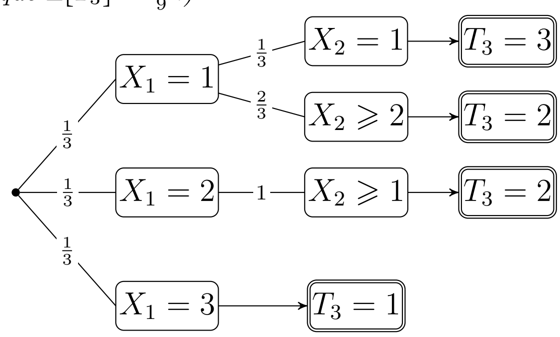
c) *(Montrer que : $\mathbb{P}(T_n = n) = \left(\frac{1}{n}\right)^{n-1}$.)*
On a l'égalité d'événements : $[T_n = n] = [S_{n-1} < n] \cap \overbrace{[S_n \geqslant n]}^{\text{automatique}}$.
Pour que la somme S_{n-1} soit $\leqslant n-1$, il faut que les $(n-1)$ termes soient tous égaux à 1, soit : $[S_{n-1} < n] = \bigcap\limits_{k=1}^{n-1} [X_k = 1]$.
Par indépendance : $\mathbb{P}(T_n = n) = \mathbb{P}\Big(\bigcap\limits_{k=1}^{n-1} [X_k = 1]\Big) = \prod\limits_{k=1}^{n-1} \overbrace{\mathbb{P}(X_k = 1)}^{\frac{1}{n}} = \left(\frac{1}{n}\right)^{n-1}$

3. *(Dans cette question, $n = 2$. Déterminer la loi de T_2.)*
On a $n = 2$, donc : $T_2(\Omega) = \{1, 2\}$.
On vient de trouver la moitié de la loi : $\mathbb{P}(T_2 = 1) = \frac{1}{2}$.
Ainsi, on trouve toute la loi ci-contre.

valeur i	1	2
proba. $\mathbb{P}(T_2 = i)$	$\frac{1}{2}$	$\frac{1}{2}$

4. *(Dans cette question, $n = 3$. Donner la loi de T_3. Vérifier que $\mathbb{E}[T_3] = \frac{16}{9}$.)*
On présente les tirages consécutifs par un arbre de probabilités.
On trouve en appliquant la formule des probabilités totales :

valeur i	1	2	3
proba. $\mathbb{P}(T_2 = i)$	$\frac{1}{3}$	$\frac{5}{9}$	$\frac{1}{9}$



Partie B

5. *(Déterminer $S_k(\Omega)$ pour tout $k \in \mathbb{N}^*$.)*

Après k tirages indépendants, tous à valeurs $[\![1, n]\!]$, on a : $k \leqslant S_k \leqslant k \cdot n$.

Ainsi, le score cumulé est un entier : $S_k(\Omega) = [\![k, k \cdot n]\!]$.

6. Soit $k \in [\![1, n-1]\!]$.

a) *(Exprimer S_{k+1} en fonction de S_k et de X_{k+1}.)*

On a la relation de Chasles entre les sommes : $S_{k+1} = \sum_{i=1}^{k+1} X_i = \overbrace{\sum_{i=1}^{k} X_i}^{S_k} + X_{k+1}$.

b) *(En utilisant un système complet d'événements lié à la variable aléatoire S_k, démontrer alors que : $\forall i \in [\![k+1, n]\!], \quad \mathbb{P}(S_{k+1} = i) = \frac{1}{n} \cdot \sum_{j=k}^{i-1} \mathbb{P}(S_k = j)$.)*

On conditionne par la valeur de S_k. *(formule des probabilités totales)*

Il vient :
$$\begin{aligned}\mathbb{P}(S_{k+1} = i) &= \sum_{i \in S_k(\Omega)} \mathbb{P}\Big([S_{k+1} = i] \cap [S_k = j]\Big) \\ &= \sum_{i=k}^{k \cdot n} \mathbb{P}\Big([X_{k+1} = i - j] \cap [S_k = j]\Big) \\ &= \sum_{i=k}^{k \cdot n} \mathbb{P}(X_{k+1} = i - j) \cdot \mathbb{P}(S_k = j)\Big) \qquad \textit{(indépendance)}\end{aligned}$$

Or : $X_{k+1} \hookrightarrow \mathcal{U}\Big([\![1, n]\!]\Big)$,

donc, on a : $\mathbb{P}(X_{k+1} = i - j) = \begin{cases} \frac{1}{n} & \text{si } 1 \leqslant i - j \leqslant n \Leftrightarrow \underbrace{\overbrace{i - n}^{\leqslant 0} \leqslant j}_{\text{automatique}} \leqslant i - 1 \\ 0 & \text{sinon.} \end{cases}$

car on a fait l'hypothèse : $i \in [\![k+1, n]\!]$.

Ainsi, dans la somme, restent les termes non-nuls : $j \in S_k(\Omega)$ tels que $j \leqslant i - 1$.

Ces termes valent tous $\frac{1}{n}$, il reste donc bien : $\mathbb{P}(S_{k+1} = i) = \frac{1}{n} \cdot \sum_{j=k}^{i-1} \mathbb{P}(S_k = j)$.

7. a) *(Pour $k \in \mathbb{N}^*$ et $j \in \mathbb{N}^*$, rappeler la formule du triangle de Pascal entre $\binom{j-1}{k-1}$, $\binom{j-1}{k}$ et $\binom{j}{k}$.)*

On a : $\binom{j-1}{k-1} + \binom{j-1}{k} = \binom{j}{k}$.

b) *(En déduire que pour tout $k \in \mathbb{N}^*$ et pour tout entier naturel $i \geqslant k+1$: $\sum_{j=k}^{i-1} \binom{j-1}{k-1} = \binom{i-1}{k}$.)*

*(C'est la « **Hockey Stick Formula** »)* On utilise la formule de Pascal précédente.

On peut procéder par récurrence, mais le plus rapide est le télescopage :

On trouve alors : $\sum_{j=k}^{i-1} \binom{j-1}{k-1} = \sum_{j=k}^{i-1} \left[\binom{j}{k} - \binom{j-1}{k}\right] = \binom{i-1}{k} - \underbrace{\binom{k-1}{k}}_{0}$.

Pour tout entier $k \in [\![1, n]\!]$, on note $\mathcal{H}_k$ la proposition :

$$\text{« } \forall i \in [\![k, n]\!] \quad \mathbb{P}(S_k = i) = \frac{1}{n^k} \cdot \binom{i-1}{k-1} \text{ ».} \qquad (\mathcal{H}_k)$$

c) *(Démontrer par récurrence, pour tout entier $k \in [\![1, n]\!]$, que $\mathcal{H}_k$ est vraie.)*

- **Hypothèse de récurrence**

  Pour $k \in [\![1, n]\!]$, on considère l'hypothèse de récurrence :
  $$\forall i \in [\![k, n]\!] \quad \mathbb{P}(S_k = i) = \frac{1}{n^k} \cdot \binom{i-1}{k-1} \qquad (\mathcal{H}_k)$$

- **Initialisation** On a bien : $\forall k \in [\![1, n]\!], \quad \mathbb{P}(S_1 = i) = \mathbb{P}(X_1 = i) = \frac{1}{n} \quad (\mathcal{H}_k)$

  (car $\binom{i-1}{k-1} = 1$)

- **Hérédité** Soit $k \in [\![1, n-1]\!]$ un entier.
  On suppose $(\mathcal{H}_k)$ soit : $\forall i \in [\![k, n]\!] \quad \mathbb{P}(S_k = i) = \frac{1}{n^k} \cdot \binom{i-1}{k-1}$
  D'après la question **6.b)**,

$$\text{on a } \forall i \in [\![k+1, n]\!], \quad \mathbb{P}(S_{k+1} = i) = \frac{1}{n} \cdot \sum_{j=k}^{n-1} \overbrace{\frac{1}{n^k} \cdot \binom{j-1}{k-1}}^{\mathbb{P}(S_k=j)}.$$

$$= \frac{1}{n^{k+1}} \cdot \underbrace{\sum_{j=k}^{n-1} \binom{j-1}{k-1}}_{\text{q}^{\text{n}} \text{ préc.}}$$

  On trouve donc bien : $\quad \forall i \in [\![k+1, n]\!], \quad \mathbb{P}(S_{k+1} = i) = \frac{1}{n^{k+1}} \cdot \binom{i-1}{k}. \qquad (\mathcal{H}_{k+1})$

- **Conclusion**
  On a montré que l'hypothèse de récurrence $(\mathcal{H}_k)_{k \in [\![1,n]\!]}$ est
  - initialisée
  - héréditaire

  On a donc bien : $\quad \forall k \in [\![1, n]\!], \ \forall i \in [\![k, n]\!] : \qquad \mathbb{P}(S_k = i) = \frac{1}{n^k} \cdot \binom{i-1}{k-1} \qquad (\mathcal{H}_k)$

8. a) *(Soit $k \in [\![1, n-1]\!]$. Comparer les événements : $\quad [T_n > k] \quad$ et $\quad [S_k \leqslant n-1]$.)*

On a : $[T_n > k]$ = « le score cumulé n'a pas attaint n aux tirages $[\![1, k]\!]$ »
= « le score cumulé n'a pas atteint n au $k^{\text{ème}}$ tirage »
$= [S_k < n] = [S_k \leqslant n-1]$.

b) *(En déduire que : $\quad \forall n \in [\![0, n-1]\!], \quad \mathbb{P}(T_n > k) = \frac{1}{n^k} \cdot \binom{n-1}{k}$.)*

On a donc : $\quad \mathbb{P}(T_n > k) = \mathbb{P}(S_k \leqslant n-1) = \sum_{i=k}^{n-1} \mathbb{P}(S_k = i)$.

Calculons cette somme par l'expression de $\mathbb{P}(S_k = i)$ *(formule $\mathcal{H}_k$)* puis la formule de la crosse de hockey **7.a)**. Il vient : $\displaystyle\sum_{i=k}^{n-1} \mathbb{P}(S_k = i) = \frac{1}{n^k} \cdot \sum_{i=1}^{n-1} \binom{i-1}{k-1} = \frac{1}{n^k} \cdot \binom{n-1}{k}$.

Ainsi, on trouve bien : $\quad \mathbb{P}(T_n > k) = \frac{1}{n^k} \cdot \binom{n-1}{k}$.

9. *(Démontrer que $\mathbb{E}[T_n] = \sum_{k=0}^{n} \mathbb{P}(T_n > k)$, puis que $\mathbb{E}[T_n] = (1 + \frac{1}{n})^{n-1}$.)*

- **Transformation d'Abel pour $\mathbb{E}[T_n]$**

  On développe l'espérance : $\quad \mathbb{E}[T_n] = \displaystyle\sum_{k=1}^{n} k \cdot \mathbb{P}(T_n = k)$

$$= \sum_{k=1}^{n} k \cdot \big[\mathbb{P}(T_n > k-1) - \mathbb{P}(T_n > k)\big]$$

$$= \sum_{k=1}^{n} k \cdot \mathbb{P}(T_n > k-1) - \sum_{k=1}^{n} k \cdot \mathbb{P}(T_n > k)$$

  On procède sur la première somme au changement d'indice $i = k - 1$.

  Il vient : $\quad \mathbb{E}[T_n] = \displaystyle\sum_{i=0}^{n-1} (i+1) \cdot \mathbb{P}(T_n > i) - \sum_{k=1}^{n} k \cdot \mathbb{P}(T_n > k)$

  On met les deux sommes au même ensemble d'indice $[\![0, n]\!]$.

  Il vient : $\quad \displaystyle\sum_{k=1}^{n} k \cdot \mathbb{P}(T_n > k) = \sum_{k=0}^{n} k \cdot \mathbb{P}(T_n > k)$

$$\text{et : } \sum_{i=0}^{n-1} (i+1) \cdot \mathbb{P}(T_n > i) = \sum_{i=0}^{n} (i+1) \cdot \mathbb{P}(T_n > i) - (n+1) \cdot \underbrace{\mathbb{P}(T_n > n)}_{0 \quad \text{car } T_n(\Omega) = [\![1,n]\!]}$$

$$= \sum_{k=0}^{n} (k+1) \cdot \mathbb{P}(T_n > i).$$

Enfin, on regroupe pour conclure : $\mathbb{E}[T_n] = \sum\limits_{k=0}^{n} (k+1) \cdot \mathbb{P}(T_n > k) - \sum\limits_{k=0}^{n} k \cdot \mathbb{P}(T_n > k)$

$= \sum\limits_{k=0}^{n} (\underbrace{k+1-k}_{1}) \cdot \mathbb{P}(T_n > k)$.

On a ainsi obtenu : $\mathbb{E}[T_n] = \sum\limits_{k=0}^{n} \mathbb{P}(T_n > k)$.

- **Calcul de l'espérance**

On remplace $\mathbb{P}(T_n > k)$, et on reconnaît la formule du binôme de Newton.

Ainsi : $\mathbb{E}[T_n] = \sum\limits_{k=0}^{n} \mathbb{P}(T_n > k) = \sum\limits_{k=0}^{n-1} \mathbb{P}(T_n > k)$

$= \sum\limits_{k=0}^{n-1} \binom{n-1}{k} \cdot \frac{1}{n^k}$

$= \sum\limits_{k=0}^{n-1} \binom{n-1}{k} \cdot a^k \cdot b^{n-1-k} = (a+b)^{n-1}$

pour $a = \frac{1}{n}$ et $b = 1$.

Ainsi, on a bien trouvé : $\mathbb{E}[T_n] = \left(1 + \frac{1}{n}\right)^{n-1}$.

10. *(Calculer* $\lim\limits_{n\to\infty} \mathbb{E}[T_n]$.*)*

On écrit la puissance comme une exponentielle pour lever l'indétermination.

Soit : $\mathbb{E}[T_n] = \left(1 + \frac{1}{n}\right)^{n-1} = \exp\left[(n-1) \cdot \ln\left(1 + \frac{1}{n}\right)\right]$.

On connaît l'équivalent quand $h \to 0$ pour le logarithme : $\ln(1+h) \sim h$.

Ainsi, pour $n \to \infty$, avec $h = \frac{1}{n} \to 0$, il vient : $(n-1) \cdot \ln\left(1 + \frac{1}{n}\right) \sim \frac{n-1}{n} \to 1$.

Par continuité de l'exponentielle : $\mathbb{E}[T_n] = \exp\Big[\underbrace{(n-1) \cdot \ln\left(1 + \frac{1}{n}\right)}_{\to 1}\Big] \xrightarrow[n\to\infty]{} \exp(1) = \mathrm{e}$.

Partie C

Dans cette partie, on fait varier l'entier n et on étudie la convergence en loi de la suite de variables $(T_n)_{n\geqslant 1}$ obtenue.

11. Soit Y une variable aléatoire à valeurs dans N^* telle que : $\forall k \in \mathbb{N}^*$, $\mathbb{P}(Y = k) = \frac{k-1}{k!}$.

a) *(Vérifier par le calcul que* $\sum\limits_{k=1}^{+\infty} \mathbb{P}(Y = k) = 1$.*)*

Calculons les sommes partielles : $\Sigma_N = \sum\limits_{k=1}^{N} \mathbb{P}(Y = k) = \sum\limits_{k=1}^{N} \frac{k-1}{k!} = \sum\limits_{k=1}^{N} \left(\frac{1}{(k-1)!} - \frac{1}{k!}\right)$.

Par sommation télescopique, il reste : $\Sigma_N = \sum\limits_{k=1}^{N} \left(\frac{1}{(k-1)!} - \frac{1}{k!}\right) = \frac{1}{0!} - \underbrace{\frac{1}{N!}}_{\to 0}$.

Ainsi : $\sum\limits_{k=1}^{+\infty} \mathbb{P}(Y = k) = \lim\limits_{N\to\infty} \Sigma_N = \frac{1}{0!} = 1$.

b) *(Montrer que Y admet une espérance et calculer cette espérance.)*

Sous réserve de convergence, on a : $\mathbb{E}[Y] = \sum\limits_{k=1}^{+\infty} k \cdot \mathbb{P}(Y = k)$

$= \sum\limits_{k=1}^{+\infty} \frac{k(k-1)}{k!} = \sum\limits_{k=2}^{+\infty} \frac{1}{(k-2)!}$

On fait le changement d'indice $i = k - 2$, et on reconnaît une série exponentielle.

Ainsi, il y a convergence de la série, et : $\mathbb{E}[Y] = \sum\limits_{i=0}^{+\infty} \frac{1}{i!} = \mathrm{e}$.

12. *(Pour tout entier naturel k non-nul, démontrer que : $\lim\limits_{n\to\infty} \mathbb{P}(T_n > k) = \frac{1}{k!}$.)*

On développe : $\mathbb{P}(T_n > k) = \frac{1}{n^k} \cdot \binom{n-1}{k} = \frac{1}{n^k} \cdot \frac{(n-1)!}{k!(n-1-k)!} = \frac{1}{k!} \cdot \frac{(n-1)!}{n^k \cdot (n-1-k)!}$.

On étudie le produit : $\frac{(n-1)!}{n^k \cdot (n-1-k)!} = \dfrac{\overbrace{(n-1) \times \cdots \times (n-k)}^{k \text{ facteurs}}}{\underbrace{n^k}_{k \text{ facteurs}}} = \underbrace{\frac{n-1}{n} \times \ldots \times \frac{n-k}{n}}_{k \text{ facteurs}}$.

Chacun des facteurs tend vers 1 quand $n \to \infty$. Il y a k facteurs *(et k est fixé !)*.

Ainsi le produit tend vers 1. Il reste donc : $\lim\limits_{n\to\infty} \mathbb{P}(T_n > k) = \frac{1}{k!}$.

13. *(Démontrer alors que $(T_n)_{n\geqslant 1}$ converge en loi vers la variable aléatoire Y.)*

Pour $k \in \mathbb{N}$, on a : $\mathbb{P}(T_n = k) = \mathbb{P}(T_n > k-1) - \mathbb{P}(T_n > k)$.

On passe à la limite $n \to \infty$. Il vient :

$\lim\limits_{n\to\infty} \mathbb{P}(T_n = k) = \lim\limits_{n\to\infty} \mathbb{P}(T_n > k-1) - \lim\limits_{n\to\infty} \mathbb{P}(T_n > k) = \frac{1}{(k-1)!} - \frac{1}{k!} =$.

On réduit au même dénominateur, et il vient bien : $\lim\limits_{n\to\infty} \mathbb{P}(T_n = k) = \frac{k-1}{k!} = \mathbb{P}(Y = k)$.

Ainsi la suite (T_n) convergence en loi vers *(la loi de)* Y.

On rappelle qu'en langage `Scilab`, l'instruction `grand(1,1,'uin',1,n)` renvoie un entier aléatoire de $[\![1, n]\!]$.

14. *(Compléter la fonction ci-dessous, qui prend en argument le nombre n de boules contenues dans l'urne, afin qu'elle simule la variable aléatoire T_n.)*

```
function y=T(n)
    S=.................
    y=.................
    while .............
        tirage = grand(1,1,'uin',1,n)
        S=S+tirage
        y=.............
    end
endfunction
```

```
function y=T(n)
    S = 0
    y = 0
    while (S<n)
        tirage = grand(1,1,'uin',1,n)
        S = S + tirage
        y = y + 1
    end
endfunction
```

15. On suppose déclarée la fonction précédente, et on écrit le script ci-dessous :

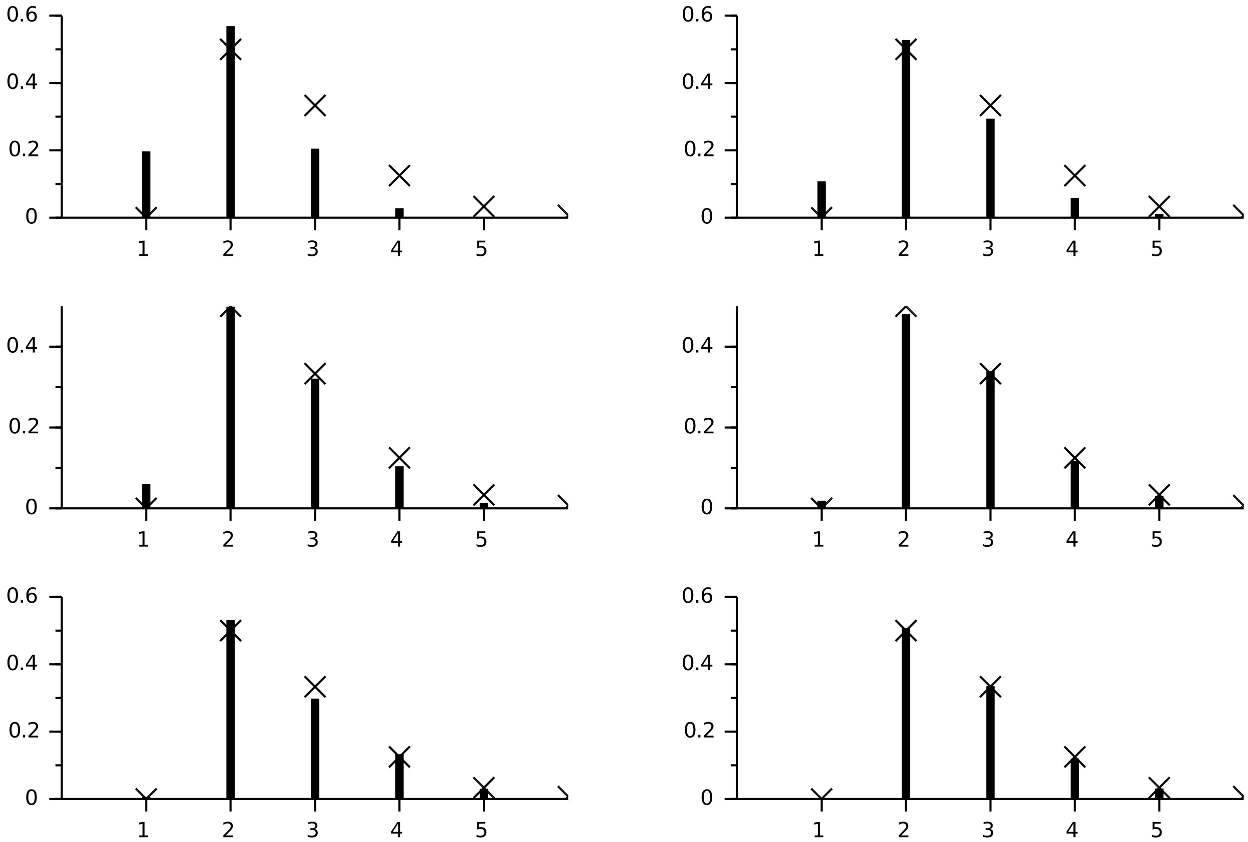
```
function y=freqT(n)
    y = zeros(1,n)
    for i = 1:100000
        k = T(n)
        y(k) = y(k) + 1
    end
    y = y / 100000
endfunction

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.

(de gauche à droite puis de haut en bas pour `n = 5, 10, 20, 50, 100, 1000`*)*



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 ?)*

- `freqT` renvoie les **fréquences empiriques**, mesurées en lançant un grand nombre de fois la fonction `T`, et en comptant.
  Les **fréquences empiriques** sont représentées par les **barres verticales grasses**.
- `loitheoY` renvoie les **probabilités théoriques de la loi limite**.
  C'est donc la suite : $\mathbb{P}(Y = k) = \frac{k-1}{k!}$.
  Les **probabilités théoriques** sont représentées par les **croix obliques**.

b) *(Expliquer en quoi cette succession de graphiques permet d'illustrer le résultat de la question* ***13.****)*
On observe le résultat attendu :

quand n est **grand**,
les **fréquences empiriques** pour T_n
sont **proches**
des **probabilités théoriques** pour la loi limite Y.