Vocabulaire et notations :
Une carte située au sommet de la pile est dite *en position n°1*, celle qui se trouve immédiatement en dessous est dite *en position n°2*, etc. Ainsi une carte située en position n°N désigne la carte située en bas de la pile. On prendra garde à bien distinguer la position d'une carte dans le paquet du numéro qu'elle porte.

Partons d'un tas de cartes rangées initialement dans l'ordre suivant :

pour tout i élément de $[\![1,N]\!]$, la carte C_i se trouve en position i.

Ainsi, à l'instant initial, la carte C_1 se trouve sur le dessus du paquet alors que C_N se trouve donc tout en dessous du paquet.

Pour k élément de $[\![1,N]\!]$, on appelle *insertion à la k-ième place* l'opération qui consiste à prendre la carte située au-dessus du paquet et à l'insérer entre la k-ième et la $(k+1)$-ième place. Une insertion à la première place ne change pas l'ordre des cartes. Une insertion à la N-ième place consiste à faire glisser la carte située au-dessus du paquet pour la mettre sous le paquet.

Le *battage par insertions* du jeu de cartes consiste à effectuer une suite d'insertions aléatoires, en choisissant, à chaque instant, au hasard uniformément dans $\{1,\cdots,N\}$ la place à laquelle l'insertion a lieu, indépendamment des insertions précédentes.
Les instants successifs d'insertions seront notées $1,2,\ldots,n,\ldots$, l'instant initial est $n=0$.

Notations. Nous notons :

- T_1 le premier instant où la carte située sur le dessus du paquet est glissée en dernière position, c'est-à-dire le premier instant où la carte C_N se trouve remontée de la position N à la position $N-1$,
- T_2 le premier instant où la carte C_N se trouve remontée en position $N-2$,
- et plus généralement, pour i dans $[\![1,N-1]\!]$, T_i le premier instant où la carte C_N atteint la position $N-i$.
- On posera également $\Delta_1 = T_1$ et $\forall\, i \in [\![2,N-1]\!]$, $\Delta_i = T_i - T_{i-1}$.
- Enfin, on notera $T = T_{N-1}+1$.

On admet que les conditions de l'expérience permettent de faire l'hypothèse que les variables aléatoires $(\Delta_i)_{i\in[\![1,N-1]\!]}$ sont indépendantes.

Description d'un exemple. Dans le tableau ci-dessous, nous décrivons les résultats d'une expérience faite sur un paquet de $N=4$ cartes. La première ligne du tableau indique les instants n, la deuxième ligne indique les positions d'insertions, et dans la dernière ligne figure la configuration du paquet à l'instant n.

	instant n	0	1	2	3	4	5	6	7
	insertion en place k		3	2	4	1	3	4	2
Configuration	position 1	C_1	C_2	C_3	C_2	C_2	C_1	C_4	C_2
du	position 2	C_2	C_3	C_2	C_1	C_1	C_4	C_2	C_4
paquet	position 3	C_3	C_1	C_1	C_4	C_4	C_2	C_3	C_3
	position 4	C_4	C_4	C_4	C_3	C_3	C_3	C_1	C_1

Pour cette expérience, on a les résultats $T_1(\omega)=3$, $T_2(\omega)=5$ et $T_3(\omega)=6$ et $T(\omega)=7$.

Partie 1 - Description et premiers résultats

1) Justifier que $\forall i \in [\![2,N-1]\!]$ $T_i = \Delta_1 + \Delta_2 + \cdots + \Delta_i$.
   Que représente l'intervalle de temps Δ_i ?
2) Loi de Δ_1.
   Déterminer pour tout entier $n \in \mathbb{N}$, $\mathbf{P}(\Delta_1 > n)$ et reconnaître la loi de Δ_1.
3) Soit $i \in [\![2,N-1]\!]$. Loi de Δ_i.

   (a) Établir que pour tout entier $n \in \mathbb{N}$, on a $\mathbf{P}(\Delta_i > n) = \left(\frac{N-i}{N}\right)^n$. En déduire que Δ_i suit une loi usuelle que l'on précisera.

(b) En déduire $E(\Delta_i) = \dfrac{N}{i}$ et $V(\Delta_i) = N\dfrac{N-i}{i^2}$.

4) Loi de T_2. Soit $n \geq 2$.

(a) Démontrer que $\mathbf{P}(T_2 = n) = \displaystyle\sum_{k=1}^{n-1} \mathbf{P}(\Delta_2 = n-k)\mathbf{P}(\Delta_1 = k)$.

(b) Justifier que $\displaystyle\sum_{k=1}^{n-1} \left(\frac{1-1/N}{1-2/N}\right)^k = N\left(1-\frac{1}{N}\right)\left[\left(\frac{1-1/N}{1-2/N}\right)^{n-1} - 1\right]$.

(c) En déduire que l'on a : $\mathbf{P}(T_2 = n) = \frac{2}{N}\left[\left(1-\frac{1}{N}\right)^{n-1} - \left(1-\frac{2}{N}\right)^{n-1}\right]$.

5) À l'instant T_2, la carte C_N est située en position $N-2$ et deux cartes se trouvent sous elle qui ont été insérées aux instants T_1 et T_2.
Que valent alors les probabilités, qu'à l'instant T_2 :

(a) la carte insérée à l'instant T_1 soit en place $N-1$ et celle insérée à l'instant T_2 en place N ?

(b) la carte insérée à l'instant T_2 soit en place $N-1$ et celle insérée à l'instant T_1 en place N ?

6) À l'instant T_3, la carte C_N est située en position $N-3$ et trois cartes, insérées aux instants T_1, T_2 et T_3, se trouvent sous elle. On note alors, pour $i \in \{1,2,3\}$, a_i la position de la carte ayant été insérée à l'instant T_i.

(a) Combien y a-t-il de résultats possibles pour le triplet (a_1,a_2,a_3) ?

(b) Quelques exemples. Donner les probabilités qu'à l'instant T_3 :

i) on obtienne $(a_1,a_2,a_3) = (N-2,N-1,N)$?

ii) on obtienne $(a_1,a_2,a_3) = (N-2,N,N-1)$?

7) Justifier la phrase suivante :

"À partir de l'instant T, toutes les configurations du jeu de cartes sont équiprobables."

On retiendra que si on arrête le battage des cartes par insertion exactement à l'instant T, on a un paquet convenablement mélangé. *Cependant le temps T étant aléatoire, il n'est pas possible d'arrêter de battre les cartes à cet instant précis, à moins de marquer la carte C_N bien sûr !*

Partie 2 - Estimation du nombre d'insertions pour bien mélanger les cartes

Notations : on introduit les suites $(H_n)_{n\geq 1}$ et $(u_n)_{n\in\mathbb{N}^*}$ définies par :

$$\forall n \geq 1 \quad H_n = \sum_{k=1}^{n} \frac{1}{k} \text{ et } u_n = H_n - \ln(n)$$

8) Espérance et variance de T

Justifier que $E(T) = NH_N$ et que $V(T) = N^2\left(\displaystyle\sum_{k=1}^{N}\frac{1}{k^2}\right) - NH_N$.

9) Étude de la suite (u_n)

(a) Montrer que pour tout entier $k \geq 1$, on a $\dfrac{1}{k+1} \leq \displaystyle\int_k^{k+1} \frac{\mathrm{d}t}{t} \leq \frac{1}{k}$.

(b) En déduire successivement :

i) la décroissance de la suite (u_n),

ii) l'encadrement : $\forall n \in \mathbb{N}^*$, $\ln(n+1) \leq H_n \leq \ln(n) + 1$.

(c) Déduire de ce qui précède que la suite (u_n) est convergente et que sa limite, notée γ appartient à $[0,1]$.

10) (a) Établir que $E(T) \underset{N\to+\infty}{\sim} N\ln(N)$ et $E(T) = N\ln(N) + N\gamma + o(N)$.

(b) Quelle est la nature de la suite $\left(\dfrac{V(T)}{N^2}\right)_{N\in\mathbb{N}^*}$? *(on prendra garde au fait que $V(T)$ dépend de N).*

Justifier qu'il existe une constante α, strictement positive, telle que

$$V(T) \underset{N\to\infty}{\sim} \alpha N^2 \text{ et } V(T) \leq \alpha N^2$$

11) Écart à la moyenne

On rappelle l'inégalité de Bienaymé-Chebychev valable pour une variable aléatoire X admettant une espérance et une variance :

$$\forall \varepsilon > 0 \quad \mathbf{P}(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$$

Soit N fixé et une constante c strictement plus grande que 1.

(a) Justifier que $\forall \omega \in \Omega,\ |T(\omega) - N\ln(N)| \leq |T(\omega) - E(T)| + N$.
Comparer par une inclusion les événements suivants

$$\Big(|T - N\ln(N)| \geq cN\Big) \text{ et } \Big(|T - E(T)| \geq N(c-1)\Big)$$

(b) Démontrer que

$$\mathbf{P}\Big(|T - N\ln(N)| \geq cN\Big) \leq \frac{\alpha}{(c-1)^2}$$

où α a été définie à la question 10b.
Le nombre N étant fixé, que vaut $\lim\limits_{c\to+\infty} \mathbf{P}\Big(|T - N\ln N| \geq cN\Big)$?

12) Démontrer aussi que pour tout $\varepsilon > 0$:

$$\lim_{N\to+\infty} \mathbf{P}\Big(|T - N\ln(N)| \geq \varepsilon N\ln(N)\Big) = 0$$

On peut traduire ces résultats en disant que l'événement : "T s'écarte de $N\ln(N)$ de manière significative" est un événement asymptotiquement rare.
Pour information, pour un paquet de 32 cartes, on donne $32\ln(32) \simeq 110$ *et pour un paquet de 52 cartes,* $52\ln(52) \simeq 205$.

13) Simulation informatique. Dans cette question on considère un jeu de $N = 32$ cartes.
MODÉLISATION : On définit en PASCAL le `TYPE Paquet=ARRAY[1..32] OF INTEGER;` Le paquet de 32 cartes est représenté par une variable `Jeu` de `TYPE Paquet` rempli initialement d'entiers entre 1 et 32; donc, initialement, `Jeu[i]` contient i, c'est à dire que la carte C_i est en position i. Au cours des insertions, `Jeu[i]` désigne le numéro de la carte en position numéro i. Par exemple, `Jeu[i]=10` signifie que la carte C_{10} est en position i.
On indique à la fin de cette question un extrait de programme à compléter en suivant les questions suivantes :

(a) Écrire la procédure `Init` permettant de définir une variable `Jeu` correspondant à la configuration initiale du paquet de cartes.
(b) Compléter la procédure `Insertion` qui simule une opération d'insertion. On rappelle que la fonction `RANDOM(32)` permet de tirer un nombre entier au hasard dans l'intervalle $[\![0,31]\!]$.
(c) Que fait la fonction `T`?
(d) Écrire le programme principal permettant de calculer et d'afficher la moyenne des valeurs prises par la fonction `T` sur 100 expériences et compléter la ligne de déclaration de variables.

Extrait du programme.

```
PROGRAM ESSEC2011;
TYPE Paquet=ARRAY[1..32] OF INTEGER;
VAR Jeu :Paquet;
    ...............    (à compléter)

PROCEDURE Init( .............)


PROCEDURE Insertion(VAR Jeu :Paquet);
VAR i,k,cartedessus :INTEGER;
BEGIN
   k := .......    (position où on va insérer la carte du dessus)
   cartedessus :=Jeu[1];
   IF k>1 THEN FOR i :=1 TO k-1 DO Jeu[i] := ....
   Jeu[k] := ....
END;

FUNCTION T(Jeu :Paquet) :INTEGER;
VAR n :INTEGER;
BEGIN
   Init(Jeu);
   n :=0;
   WHILE Jeu[1]<>32 DO
   BEGIN
      Insertion(Jeu);
      n :=n+1
   END;
   T :=n
END;

BEGIN  { programme principal }

END.
```

Partie 3 - Distance variationnelle à la loi uniforme

Notations:

• On note π l'équiprobabilité sur $\mathcal{S}_N$, c'est-à-dire l'application de $\mathcal{P}(\mathcal{S}_N)$ dans [0,1] telle que :

$$\forall A \subset \mathcal{S}_N \quad \pi(A) = \frac{\text{card}(A)}{N!}; \text{ en particulier, } \forall \sigma \in \mathcal{S}_N \quad \pi(\{\sigma\}) = \frac{1}{N!}$$

• On note également μ_n la probabilité sur $\mathcal{S}_N$ définie comme suit :
pour chaque configuration σ de $\mathcal{S}_N$, $\mu_n(\{\sigma\})$ désigne la probabilité qu'à l'instant n le tas de cartes se trouve dans la configuration σ.
On a alors pour pour toute partie A de $\mathcal{S}_N$, $\mu_n(A) = \sum_{\sigma \in A} \mu_n(\{\sigma\})$.
On peut mesurer la qualité du mélange à un instant donné n en estimant l'écart entre μ_n et π. Une *distance* d entre ces probabilités est définie de la manière suivante :

$$d(\mu_n,\pi) = \max\left\{ |\mu_n(A) - \pi(A)| \,,\, A \subset \mathcal{S}_N \right\}$$

14) Soient A une partie de $\mathcal{S}_N$, $n \in \mathbb{N}^*$ et E_n l'événement : "à l'instant n le paquet de cartes se trouve dans une configuration qui appartient à la partie A."

(a) Expliquer, en utilisant la question 7, l'égalité suivante : $\mathbf{P}_{(T\leq n)}(E_n) = \pi(A)$.
En déduire $\mathbf{P}(E_n \cap (T \leq n)) = \pi(A)\mathbf{P}(T \leq n)$.

(b) Établir que $\mathbf{P}(E_n \cap (T > n)) \leq \mathbf{P}(T > n)$.

(c) Montrer que

$$\mu_n(A) \leq \pi(A) + \mathbf{P}(T > n)$$

15) Soit A une partie de $\mathcal{S}_N$ et $n \in \mathbb{N}^*$. On note $\overline{A}$ l'événement contraire de A.

(a) Exprimer $\mu_n(\overline{A}) - \pi(\overline{A})$ en fonction de $\mu_n(A)$ et $\pi(A)$.

(b) Déduire des questions précédentes la majoration :

$$|\mu_n(A) - \pi(A)| \leq \mathbf{P}(T > n)$$

16) Montrer que $\forall n \in \mathbb{N}^*$, $0 \leq d(\mu_n,\pi) \leq \mathbf{P}(T > n)$. Déterminer la limite $\lim\limits_{n\to+\infty} d(\mu_n,\pi)$.

Partie 4- Une majoration de $\mathbf{P}(T > n)$

Dans cette partie, nous nous intéressons provisoirement à un collectionneur de timbres. Celui-ci reçoit chaque jour une lettre affranchie avec un timbre choisi au hasard uniformément parmi les N timbres en vigueur. On étudie ici le nombre de jours que doit attendre le collectionneur pour posséder la collection complète des N timbres. Le jour 0 il n'a aucun timbre.

On note alors :

- pour tout entier $k \in [\![1,N]\!]$, S_k le nombre aléatoire de jours que doit attendre le collectionneur pour que le nombre de timbres différents qu'il possède passe de $k-1$ à k,
- $S = S_1 + S_2 + \cdots + S_N$, soit la variable aléatoire correspondant au nombre de jours à attendre pour posséder la collection complète des N timbres,
- en supposant les N timbres en vigueur numérotés de 1 à N, pour tout $j \in [\![1,N]\!]$, B_j^m l'événement "le jour m, le collectionneur n'a toujours pas reçu de lettre affranchie avec le timbre numéro j."

On admet que les variables aléatoires $(S_k)_{k\in[\![1,N]\!]}$ sont indépendantes.

17) Déterminer la loi de S_1.

18) Déterminer pour tout entier $k \in [\![2,N]\!]$ la loi de la variable S_k.

19) En déduire que la variable S suit la même loi de probabilité que la variable T étudiée dans les parties précédentes.
Ce résultat sera utilisé pour estimer la quantité $\mathbf{P}(T > n)$.

20) Soit $m \in \mathbb{N}^*$.

(a) Exprimer l'événement $(S > m)$ à l'aide des événements B_1^m, B_2^m, $\ldots$, B_N^m.

(b) Que vaut $\mathbf{P}(B_j^m)$ pour tout entier $j \in [\![1,N]\!]$?

(c) On rappelle que pour tout entier $n \geq 2$ et pour toute famille d'événements $A_1,\ldots,A_n$, on a l'inégalité : $\mathbf{P}(\bigcup\limits_{i=1}^{n} A_i) \leq \sum\limits_{i=1}^{n} \mathbf{P}(A_i)$ En déduire $\mathbf{P}(S > m) \leq N\left(1 - \dfrac{1}{N}\right)^m$.

21) (a) Montrer que $\ln(1+x) \leq x$ pour tout $x \in]-1,+\infty[$.

(b) Déduire des résultats précédents la majoration

$$\forall m \in \mathbb{N},\ \mathbf{P}(T > m) \leq Ne^{-\frac{m}{N}}$$

22) On reprend les notations introduites dans la partie précédente.

(a) Soit $c > 0$ fixé. Montrer que pour n entier supérieur ou égal à $N\ln N + cN$ on a : $d(\mu_n,\pi) \leq e^{-c}$.

(b) *Application numérique.* On estime qu'une distance en variation à la loi uniforme de 0,2 est acceptable.
Avec un jeu de 32 cartes, combien de *battages par insertions* doit-on faire pour considérer le paquet mélangé de façon acceptable?