

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

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

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

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

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

- (b) Démontrer que

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

où α a été définie à la question 10b.

Le nombre N étant fixé, que vaut $\lim_{c \rightarrow +\infty} \mathbf{P}\left(|T - N \ln N| \geq cN\right)$?

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

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

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 :

- Écrire la procédure `Init` permettant de définir une variable `Jeu` correspondant à la configuration initiale du paquet de cartes.
- 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 $[[0,31]]$.
- Que fait la fonction `T` ?
- 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 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 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 S_N et $n \in \mathbb{N}^*$. On note $\bar{A}$ l'événement contraire de A .

- (a) Exprimer $\mu_n(\bar{A}) - \pi(\bar{A})$ 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 $\lim_{n \rightarrow +\infty} d(\mu_n, \pi)$.

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 $k \in \llbracket 1, N \rrbracket$, 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 ,
- $S = S_1 + S_2 + \dots + S_N$, 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 $j \in \llbracket 1, N \rrbracket$, 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 $(S_k)_{k \in \llbracket 1, N \rrbracket}$ sont indépendantes.

- 17) Déterminer la loi de S_1 .
- 18) Déterminer pour tout entier $k \in \llbracket 2, N \rrbracket$ 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 $B_1^m, B_2^m, \dots, B_N^m$.
- (b) Que vaut $\mathbf{P}(B_j^m)$ pour tout entier $j \in \llbracket 1, N \rrbracket$?
- (c) On rappelle que pour tout entier $n \geq 2$ et pour toute famille d'événements $A_1, \dots, A_n$, on a

$$\text{l'inégalité : } \mathbf{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbf{P}(A_i) \quad \text{En déduire } \mathbf{P}(S > m) \leq N \left(1 - \frac{1}{N}\right)^m.$$

- 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?