# ANNALES

## 2005

I

**Préliminaire:** RANDOM est une variable aléatoire qui suit la loi diffuse uniforme sur l'intervalle (0,1) . $a_0$ et $b_0$ sont donnés tels que $a_0 < b_0$ .
quelle est la loi suivie par $X=a_0 + (b_0-a_0) * \text{RANDOM}$ ?
a et b sont 2 réels tels que : $a_0 \leq a < b \leq b_0$ quelle est la loi de (X sachant $a \leq X \leq b$) ?
**exercice:** c est un réel tel que $0 < c < 1$
On dispose d'une fonction RAND qui lors de N appels successifs donne des valeurs r qui sont une réalisation d'un échantillon de taille N de RANDOM .
D est la variable aléatoire image de R par l'application $r \rightarrow d$ définie par l'algorithme

```
r=a0 + (b0 – a0)*RAND;  a=a0 ; b=b0 ;
pour k=1 à N faire:
  m=(1-c)*a + c*b ;
  si m ≤ r alors a=m sinon b=m ;
  fin_de_boucle_k
d=b-a;
fin.
```

(conseil : éventuellement, essayer c=1/2 pour mieux comprendre cet algorithme)
**1 dans le cas ou N=1** :
quelle est la loi de D ? quelle est son espérance E(D), sa variance ?
quelle valeur de c donne E(D) minimum ? dans ce cas particulier que peut-on dire de très spécial sur la variable aléatoire D
**2 dans le cas ou N est >1** : mêmes questions (ici vous pouvez vous dispenser de calculer la variance)
**3 maintenant N=4**: c=1/4 calculer la probabilité que D soit plus petit que pour c=1/2

II

Une urne contient des boules blanches et des noires . On appelle tirage au hasard avec remise : un tirage au cours duquel la probabilité de tirer une boule d'une certaine catégorie = le nombre de boules de cette catégorie divisé par le nombre total de boules contenues dans l'urne (avec remise signifie que ces nombres ne varient pas au cours des tirages successifs).
**1** au cours de 8 tirages au hasard avec remise on n'a obtenu qu'une blanche ; peut-on rejeter avec le seuil 0.05 l'hypothèse H : "il y a le même nombre de boules blanches et noires dans l'urne" ?
**2** au cours de 1024 tirages au hasard avec remise on a obtenu 477 blanches ; peut-on rejeter avec le même seuil 0.05 l'hypothèse H ?

2004

Tous les algorithmes demandés seront (naturellement) écrits en L.D.A , clairs et bien commentés .

E est un ensemble d'éléments avec une relation d'ordre total, notée classiquement $\leq$ .
L'algorithme Rand_Tri reçoit en entrée une suite de N éléments de E et retourne en sortie la suite triée suivant l'ordre croissant .

Principe de l'algorithme Rand_Tri : un élément de la suite (appelé : pivot) est choisit au ''hasard'' (c'est à dire que chaque élément de la suite a la même probabilité d'être choisi) ; on compare le pivot avec tous les éléments de la suite , ceux qui sont $\leq$ au pivot sont placés avant le pivot et les autres après le pivot ; on obtient ainsi 2 sous suites (éventuellement l'une des deux peut être vide) . On itère ce procédé sur chaque sous suite ainsi formée, tant qu'elles comportent plus d'un élément . Lorsque l'algorithme s'arrête la suite de départ est triée suivant l'ordre croissant .
On s'intéresse à la variable aléatoire N_C nombre de comparaisons effectuées au cours de l'exécution de l'algorithme .

Question préliminaire : le nombre de comparaisons effectuées au cours de l'exécution de l'algorithme dépend naturellement de l'ordre initial de la suite d'entrée, qui n'est pas connu, notons L la loi de l'ordre de la suite d'entrée . Montrer qu'en fait la loi de N_C ne dépend pas de la loi L ; c'est à dire précisément, que : pour tout entier k
Prob( N_C = k , connaissant l'ordre d'entrée ) = Prob( N_C = k )

## I Calcul de l'espérance de N_C ,notée E(N_C)

1) on note $x_i$ i=1…N les éléments de la suite en sortie de l'algorithme (c.à.d.. triée suivant l'ordre croissant) ; soient i et j tels que $1\leq i < j \leq N$ ;
dans le cas où $j>i+1$, il existe des éléments $x_k$ entre $x_i$ et $x_j$ ; montrer que la Probabilité qu'au cours de l'exécution de l'algorithme $x_i$ ou $x_j$ aient été choisis comme pivot avant l'un quelconque des $x_k$ pour $i<k<j$ , vaut $2/(j-i+1)$ . En déduire la Probabilité qu'au cours de l'exécution de l'algorithme, il y ait eu une comparaison entre $x_i$ et $x_j$ .
Dans le cas où $j=i+1$ , la méthode de calcul ci-dessus n'a plus de sens, cependant, vérifier directement que la formule que vous avez trouvée pour la Probabilité est encore valable .

2) on note $X_{i,j}$ la variable aléatoire caractéristique de l'événement : ''au cours de l'exécution de l'algorithme, il y a eu une comparaison entre $x_i$ et $x_j$ '' .Calculer l'espérance de $X_{i,j}$ et sa variance . Trouver une relation **très simple** entre N_C et les $X_{i,j}$ .A l'aide de cette relation trouver une formule qui donne E(N_C) en fonction de N (vous ne chercherez pas à simplifier cette formule qui contient une double sommation)

3) écrire un algorithme qui calcule E(N_C) .

II Estimation de la variance de N_C , notée Var(N_C)

Maintenant , on suppose que N le nombre d'éléments des suites triées , est un '' grand nombre '' du point de vue des statistiques .

1) les variables aléatoires $X_{i,j}$ sont elles deux à deux indépendantes ?
Pour calculer Var(N_C) , pourquoi ne peut-on utiliser la relation entre N_C et les $X_{i,j}$ d'une manière analogue à ce qui a été fait dans I 2) ?

2) on suppose disponible une fonction appelée rand_tri qui lance l'exécution de l'algorithme Rand_Tri et retourne le nombre de comparaisons effectuées au cours de cette exécution ; écrire un algorithme (après avoir justifié son principe) qui donne une estimation de Var(N_C) .

3) vous supposerez que le paramètre taille_échantillon (que vous avez du introduire dans l'algorithme écrit en 2) ) est supérieur à 100 , compléter votre algorithme pour qu'il donne aussi un intervalle de confiance de la forme classique (0 , b) pour Var(N_C) , au seuil de 0.05 .

4) on considère la condition : Probabilité de
( valeur absolue de [ESTIMATEUR de Var(N_C) - Var(N_C) ] / Var(N_C) < 0.01 )
> 0.95
écrire un algorithme (en le justifiant) qui détermine la valeur minimum du paramètre taille_échantillon qui permet d'obtenir la condition ci dessus .
Ecrire une phrase simple qui traduit intuitivement la condition considérée .