

## EPREUVE D'ANALYSE DES DONNEES ET THEORIE DE L'INFORMATION

Théorie de l'Information

sur 7 points – à rédiger sur feuille séparée

*Documents manuscrits seuls autorisés.**Les intermédiaires des raisonnements ou des calculs doivent figurer sur la copie.  
Les valeurs numériques seront données avec 3 chiffres significatifs.***Exercice 1**

On considère une situation probabiliste de défaut caché et de test, modélisée par un système  $(X, Y)$ , où  $X = \{D, nD\}$  représente la présence ou l'absence de défaut, et  $Y = \{T+, T-\}$  représente les résultats du test (positif ou négatif). Les probabilités sont données par le tableau ci-dessous :

	T+	T-
D	0.07	0.01
nD	0.03	0.89

On cherche à déterminer l'efficacité du test, caractérisée par le quotient

$$r = \frac{I(X; Y)}{H(X)}.$$

Calculer  $H(X)$ ,  $I(X, Y)$ ,  $r$ .

**Exercice 2**

Dans le schéma de cryptographie asymétrique à clé publique (type RSA), donner et commenter le codage de signature qui permet d'authentifier l'expéditeur d'un message.

**Exercice 3**

On considère une source  $S$  qui émet des caractères, que l'on suppose restreints aux majuscules non accentuées et à l'espace (qui tient lieu à la fois de séparateur et de ponctuation). On suppose que la phrase suivante (P. Verlaine) est caractéristique des fréquences des caractères : "UN OISEAU SUR L ARBRE QU ON VOIT CHANTE SA PLAINTÉ" (nota : ce message contient exactement 50 signes : 10 espaces, 5 A, etc. (un ensemble  $A$  de 18 caractères espace compris est représenté)).

1°) Effectuer la statistique des fréquences des caractères de la phrase donnée, et calculer l'entropie  $H(S)$ . Comparer avec l'entropie qu'aurait la source si les 18 caractères de  $A$  étaient émis avec équiprobabilité.

2°) Effectuer le codage binaire de Huffman de  $A$  (on classera par ordre alphabétique les lettres ayant la même fréquence). On montrera les étapes de la construction et on récapitulera en donnant l'arbre de décodage. Calculer l'efficacité de ce code.