

UE ALGO6 — TD2 — Séance 3 : aléatoire et simulation

`java.util.Random` implémente les trois méthodes suivantes :

- **boolean** `nextBoolean()` //retourne une valeur booléenne aléatoire
- **double** `nextDouble()` //retourne un double entre 0.0 et 1.0
- **int** `nextInt(n)` //retourne un int entre 0 inclus et n exclu

Exercice 1. Comment simuler :

- un lancé de dé à 6 faces non pipé ;
- un tirage à pile ou face pipé (i.e. pile avec la probabilité p) ;
- autour d'une roue est écrit «ananas» (Chacune des lettres occupe $\frac{1}{6}$ de la roue). La roue tourne et s'arrête sur une lettre. Simuler le tirage aléatoire des lettres.

Génération de textes

Une manière simple de générer un texte aléatoirement est de considérer que le n^{e} mot du texte ne dépend que du $(n - 1)^{\text{e}}$ mot. Cette manière de faire est utilisée par des comptes twitter : «*Repasant par les rues effrontément en haillons, il composait une étrange musique amoureuse*» est un exemple de *tweet* publié par le *singe* (<http://twitter.com/singe9>).

Pour cela, il faut d'abord fixer les probabilités $P(m, m')$ qu'un mot m soit suivi du mot m' . Ceci se fait à l'aide d'un texte «d'apprentissage».

Exercice 2. Soit le texte d'apprentissage suivant :

We shall fight in France, we shall fight on the seas and oceans, we shall fight in the air. We shall fight on the beaches, we shall fight on the landing grounds, we shall fight in the fields and in the streets, we shall fight in the hills.

Sachant que $P(i, j)$ se calcule à l'aide du nombre d'occurrence du mot i (n_i) et du nombre d'occurrence de la séquence ij (n_{ij}) Quelle est la valeur que ce texte permet de donner aux probabilités suivantes :

- $P(\text{we}, \text{shall})$
- $P(\text{fight}, \text{in})$, $P(\text{fight}, \text{on})$ et $P(\text{fight}, j) \forall j \notin \{\text{in}, \text{on}\}$
- $P(\text{in}, j) \forall j$
- $P(i, j) \forall j, i$

Exercice 3. Proposer un algorithme qui calcule les $P(i, j)$ en une passe sur le texte.

Exercice 4. Proposer un algorithme qui génère un texte conforme aux $P(i, j)$.

Exercice 5. Pour aller plus loin il faut introduire de la *nouveauté*, on pourra utiliser le lissage de *Good-Turing*¹.

1. *Dixit Wikipedia* : Good-Turing frequency estimation was developed by Alan Turing and his assistant I. J. Good as part of their efforts at Bletchley Park to crack German ciphers for the Enigma machine during World War II.