Séance 1
Générateurs de nombres aléatoires

Dans cette première séance d'exercices, nous allons apprendre à générer des nombres aléatoires. Bien que les nombres que nous allons générer par ordinateur ne soient jamais réellement aléatoires, puisqu'ils seront produits à partir d'algorithmes bien définis, il est possible de faire en sorte qu'ils apparaissent aléatoires à quelqu'un qui ne connaît pas l'algorithme.

1. Algorithme de génération d'une variable aléatoire uniforme : Un premier simple algorithme de génération de nombres aléatoires (appelé générateur congruentiel linéaire) peut être construit de la manière suivante. En choisissant un entier initial I_0, la formule suivante permet de générer une suite de nombres "aléatoires" R_j :

$$I_j = \text{mod}\ (aI_{j-1} + c, m)$$
$$R_j = \frac{I_j}{m},$$

où a, c et m sont des nombres entiers. De cette façon, en choisissant judicieusement les valeurs des paramètres, on obtient une série de nombres R_j distribués uniformément sur l'intervalle $[0, 1[$.

a) Construisez un générateur de nombres aléatoires uniformes sur $[0, 1[$ en suivant la méthode décrite ci-dessus. Utilisez $I_0 = 4711$, $a = 205$, $c = 29573$ et $m = 139968$. Observez la moyenne donnée par `ROOT`, est-ce attendu ? Observez l'écart type (RMS) donné par `ROOT`, est-ce attendu ?

b) Comparez la corrélation entre deux nombres aléatoires consécutifs. Que remarquez-vous ?

c) Dessinez l'histogramme en incluant les barres d'erreurs à l'aide de l'option `"e"` de la méthode `Draw()`. Justifiez leur taille.

2. Génération d'une variable aléatoire uniforme à l'aide des fonctions prédéfinies de `ROOT` : Il existe bien entendu des algorithmes bien plus puissants que le simple calcul modulaire que nous venons d'utiliser. `ROOT` propose un algorithme, que nous allons comparer avec notre première méthode.

a) Au moyen de la classe `TRandom3` de `ROOT`, refaites l'exercice précédent.

b) **[Devoir]** Construisez un générateur de nombres aléatoires uniformes sur l'intervalle $[-1\,000, 1\,000[$, utilisant la classe `TRandom3`, avec une seed à 123456. Générez 10 000 nombres aléatoires dans un histogramme de 100 bins allant de $-1\,000$ à $1\,000$. Observez la moyenne et l'écart type donnés par `ROOT`. Il y a de grandes chances pour que la moyenne de votre échantillon ne soit pas 0 (moyenne de la distribution théorique). Quelle est la distribution suivie par l'estimateur moyenne (difficile de manière exacte) ? Quelle est la p-value (two-tailed) de notre observation de la moyenne ?

3. Génération d'une loi normale : Maintenant que nous sommes en mesure de générer des nombres aléatoires suivant une loi uniforme, nous pouvons commencer à générer des distributions un peu plus intéressantes. Le premier exemple sera celui de la loi Normale centrée réduite : $\mathcal{N}(0, 1)$. Il a été montré que la manière optimale de simuler cette distribution est donnée par la *méthode de Box-Muller* dont voici le principe :
Si U_1 et U_2 sont des variables aléatoires indépendantes qui suivent une loi uniforme sur $[0, 1[$, alors les variables aléatoires :

$$T_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$$
$$T_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$$

suivent toutes les deux une distribution normale centrée réduite et sont indépendantes.

a) Écrivez un programme qui génère des nombres aléatoires suivant une loi $\mathcal{N}(\mu, \sigma^2)$. Utilisez la classe `TRandom3` pour générer les nombres U_1 et U_2.

b) Utilisez les outils d'ajustement (fit) pour ajuster une gaussienne sur la distribution générée. Vérifiez que les paramètres obtenus sont en accord avec vos valeurs de μ et σ^2.

c) [**Devoir**] Dans un histogramme de 100 bins entre 0 et 20, générez 10 000 points suivant $\mathcal{N}(\mu = 10, \sigma^2 = 16)$, avec une seed de 123456 pour le générateur. Combien d'événements sont dans l'underflow ? Combien d'événements sont dans l'overflow ? Combien d'événements auriez-vous prédits dans l'underflow et dans l'overflow ? Que vaut la moyenne échantillon ? Que vaut le RMS ? Connaissant la vraie distribution utilisée pour générer les nombres aléatoires, quelle est la p-value (two-tailed) de la moyenne observée ?

d) [**Devoir**] Fittez l'histogramme avec une gaussienne. Quel est le χ^2 observé ? Quel est le nombre de degrés de liberté ? Quelle est la p-value ? Est-ce satisfaisant ?