NOMBRE AU HASARD (C1, C12)
(12 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Les notions courantes de « **hasard** » de « **nombre au hasard** » peuvent se concrétiser à partir de divers **processus de génération de nombres au hasard**. Le principe consiste à engendrer, par un procédé quelconque, une **suite** de nombres, indépendants et équiprobables (**suite iid**), à valeurs dans l'ensemble $N_9 = \{0, 1,..., 9\}$.

(i) Le **modèle générateur** est donc un **modèle puissance** tq $(N_9, \mathscr{P}(N_9), P)^{\otimes \mathbf{N}}$ dans lequel P est une **mesure de probabilité** (**mesure discrète**) définie sur $\mathscr{P}(N_9)$ selon :

(1) $P(\{x\}) = 10^{-1}, \quad \forall x \in N_9$,

ie, en notant δ_x la **loi de DIRAC** placée au point $x \in N_9$:

(2) $P = 10^{-1} \Sigma_{x=0}^{9} \delta_x$ (**loi uniforme discrète** $\mathscr{U}(N_9)$).

La **suite de nombres au hasard** précédente est une **suite uniforme**, ie uniformément distribuée, ou équidistribuée, d'un tirage à l'autre.

On peut aussi définir des tirages de **nombres au hasard non uniformes**, ie avec probabilités inégales : eg avec $P = \Sigma_{x=0}^{9} p_x \delta_x$, où $p \in S_{10}$ (**simplexe** de $\mathbf{R}^{10}$) (cf **changement de variable aléatoire**, **lemme d'uniformisation des lois**).

Très généralement, l'expression de **« tirage au hasard »** désigne le *« tirage selon un hasard uniforme »* défini ci-dessus.

(ii) Plus généralement, un entier $k \in \mathbf{N}^*$ étant donné, on appelle **tirage au hasard (uniforme) de k nombres** une épreuve consistant à observer un nombre entier $x \in \{0, 1,..., 10^k - 1\} \subset \mathbf{N}$ généré par la loi discrète uniforme suivante :

(3) $P = \Sigma_{x=0}^{d-1} p_x . \delta_x$, où $d = 10^k$ et $p \in S_d$ (**simplexe** de $\mathbf{R}^d$).

(iii) De façon encore plus générale, on dit qu'un **processus stochastique** $X = \{(\Omega, \mathscr{T}, P), (\mathscr{X}, \mathscr{B}), (X_t)_{t \in T}\}$ est un **processus générateur** de K nombres au hasard de k chiffres chacun et avec probabilités inégales ssi :

(a) $\mathscr{X} = \mathscr{X}_0^K$, avec $\mathscr{X}_0 = \{0, 1,..., 10^k - 1\}$, $\mathscr{B}_0$ est la **tribu discrète** de $\mathscr{X}_0$ et $\mathscr{B} = \mathscr{B}_0^{\otimes K}$;

(b) $T = \{1 ,..., K\} = N_K^*$;

(c) $X = (X_t)_{t \in T}$ est un **processus purement aléatoire** (ie ici indépendant équidistribué), ou **processus iid**, dont la **variable parente** $\xi : \Omega \mapsto \mathcal{X}_0$ suit une **loi discrète** P^ξ définie sur $\mathcal{B}_0$ selon :

(4) $P^\xi = \Sigma_{x \in \mathcal{X}0}\ p_x \,.\, \delta_x$,

où $\mathcal{X}0$ désigne $\mathcal{X}_0$.

Souvent, P^ξ est la **loi uniforme**, ie $p_x = 10^{-k}$, $\forall\ x \in \mathcal{X}_0$.

(iv) En pratique, la **génération de tables de nombres au hasard** s'effectue à l'aide de divers procédés :

(a) des *procédés physiques* : eg procédés vibratoires (optiques, acoustiques, etc) ;

(b) des *procédés mathématiques* dérivés de la théorie des nombres. Ces procédés fournissent seulement des **nombres « pseudo-aléatoires »**. En effet, la méthode de construction est, dans son principe, déterministe : eg **méthode de LEHMER**, **méthode de NEUMANN**.

Une **table de nombres au hasard** est donc une suite uniforme de **variables aléatoires**. Des exemples classiques (ou « historiques ») de tables de nombres au hasard sont les suivants :

(a) calcul numérique : **table de R.A. FISHER - F. YATES**, obtenue par extraction des décimales de rangs 15 à 19 d'une table de logarithmes (méthode « révisée ») ;

(b) démographie : **table de L.H.C. TIPPETT**, obtenue à partir des résultats d'un recensement de la population du Royaume-Uni ;

(c) mécanisme optique : **table de M.G. KENDALL - B. BABINGTON-SMITH**, obtenue par rotation et éclairement irrégulier d'un disque divisé en 10 secteurs identiques ;

(d) mécanisme acoustique : **table de la RAND Corporation**, obtenue par « écrêtage » d'un bruit de fond ;

(e) sociologie : **table de A. HALD**, obtenue à partir des résultats de la Loterie Nationale du Danemark.

Les tables de nombres au hasard non distribués uniformément, ie de nombres distribués selon des **lp** particulières, se déduisent généralement des tables uniformes (cf **uniformisation des lois**).

(v) La **qualité d'une table** dépend surtout du caractère uniforme (au sens probabiliste du terme, ie du caractère iid) de la suite de nombres obtenue. Divers **tests d'hypothèses** permettent de s'assurer du caractère équiprobable et « aléatoire » de telles suites : eg **tests d'adéquation** à la loi uniforme, **tests d'indépendance**, test d'absence de **lacunes** (« trous » dans les suites obtenues), voire **test des séquences**, etc (cf aussi **aléatoire**, **échantillon aléatoire**, **processus purement aléatoire**).

La plus petite taille (longueur de suite) N, ie le plus grand exposant de la puissance tensorielle $(N_9 , \mathscr{P}(N_9), \mathscr{U}(N_9))^{\otimes N}$, est appelée **période « statistique »** de la suite considérée. Plus N est grand pr aux nécessités pratiques (eg tailles des populations à sonder, etc), meilleure est la table.

Les capacités de calcul actuelles des ordinateurs facilitent la construction des tables ainsi que les tests relatifs à leur qualité.

Les deux principaux critères de qualité d'une table de nombres au hasard sont donc (a) l'indépendance de la suite et (b) son équidistribution (suite iid).

(vi) Les tables de nombres au hasard peuvent être utilisées dans des contextes variés :

(a) en **théorie des sondages**, pour le choix des unités à tirer dans les **bases** de sondage ;

(b) dans diverses **simulations**, eg dans l'étude de **phénomènes** aléatoires complexes (simulation de **systèmes aléatoires** complexes) ;

(c) pour analyser les propriétés statistiques de certaines **procédures statistiques** (eg estimation, tests), notamment leurs **propriétés asymptotiques** ;

(d) pour la résolution même de problèmes mathématiques (eg en calcul numérique), lorsque le calcul explicite des solutions est difficile ou impossible à pratiquer avec les méthodes classiques. C'est l'objet des **méthodes de MONTE CARLO** : calcul d'intégrales multiples, des solutions d'équations différentielles ou intégrales, etc.