Simulation par la Méthode des alias ou Méthode de Walker:

Prof. Mohamed El Merouani

2018/2019

Supposons que nous disposons de procédés efficients pour échantillonner à partir des v.a. X_1 et X_2, avec fonctions de masse $\{p_i\}$ et $\{q_i\}$, respectivement ; on veut générer une v.a. X de fonction de masse :

$$P(X_i = i) = \alpha p_i + (1 - \alpha) q_i$$

avec $\alpha \in (0, 1)$. Pour ce faire, on peut échantillonner à partir de X, en générant d'abord un nombre $U \sim \mathcal{U}(0, 1)$, et après, en générant à partir de X_1 si $U < \alpha$ et à partir de X_2 si $U > \alpha$.

Méthode de composition :

Exemple :

On veut échantillonner à partir de X de fonction de masse :

i	0	1	2	3	4	5
p_i	0, 12	0, 12	0, 12	0, 12	0, 32	0, 20

La fonction de masse de X peut s'écrire comme composition des fonctions de masse de deux variables uniformes X_1 et X_2 données, respectivement (voir table en-bas), avec $\alpha = 0,6$, c'est-à-dire

$$p_i = 0,6p_i^1 + 0,4p_i^2$$

i	0	1	2	3	4	5
p_i^1	0, 20	0, 20	0, 20	0, 20	0, 20	0, 0
p_i^2	0, 0	0, 0	0, 0	0, 0	0, 50	0, 50

L'algorithme sera, alors :

Générer $U_1, U_2 \sim \mathcal{U}(0,1)$
si $U_1 < 0,6$, sortir $X = ent(5U_2)$
sinon, sortir $X = ent(2U_2) + 4$

La méthode des alias (Walker, 1977) permet de générer de manière efficiente des variables discrètes avec support fini. Elle inclue une phase de prétraitement, mais après elle est rapide et d'implementation facile. Supposons que nous voulons générer à partir d'une variable X de fonction de masse $P = \{p_i, i = 1, \ldots, n\}$. La méthode des alias est, en fait, une méthode de composition basée sur la représentation de P comme une mixture :

$$P = \frac{1}{n-1} \sum_{k=1}^{n-1} Q^{(k)} \tag{1}$$

où $Q^{(k)}$ est une distribution concentrée sur, en total, deux points de $\{1, \ldots, n\}$. La démonstration de cette décomposition est basée sur le résultat suivant :

Méthode des alias ou méthode de Walker :

Lemme :

Soit $P = \{p_i, i = 1, \dots, n\}$ une fonction de masse de probabilité. Alors :

1. il existe i, $1 \leq i \leq n$, tel que $p_i < \frac{1}{(n-1)}$.
2. pour tel i, il existe $j \neq i$ tel que $p_i + p_j \geq \frac{1}{(n-1)}$.

Démonstration :
Par l'absurde, si on n'a pas le 1^{er} point du lemme, alors $\sum_{l=1}^{n} p_l > 1$.
Si on n'a pas le $2^{\text{ème}}$ point, $\sum_{l=1}^{n} p_l < 1$. $\blacksquare$
Avant de donner une idée de la démonstration de (1), illustrons le procèdé par un exemple.
$P^{(k)}, k \leq n-1$, désignera les distributions auxiliaires avec masse, en total, en $n-k$ points de $\{1, \dots, n\}$.

Considérons la loi discrète P concentrée en $\{1, 2, 3\}$, avec probabilités :

$$p_1 = \frac{3}{8}, \quad p_2 = \frac{1}{2}, \quad p_3 = \frac{1}{8}$$

Nous désirons faire

$$P = \frac{1}{2}Q^{(1)} + \frac{1}{2}Q^{(2)}$$

avec $Q^{(1)}, Q^{(2)}$ bi-ponctuelles, c-à-d

$$p_l = \frac{1}{2}q_l^{(1)} + \frac{1}{2}q_l^{(2)}, l = 1, 2, 3$$

et un des $q_l^{(i)}$ est nul.
En appliquant le lemme, nous avons

$$p_3 < \frac{1}{2}, p_3 + p_2 > \frac{1}{2},$$

avec lequel nous faisons $i = 3, j = 2$. Choisissons $Q^{(1)}$ de facon qu'il se concentre en 2 et 3, et pour que 3 prend toute sa masse de P à partir de $Q^{(1)}$, c-à-d

$$p_3 = \frac{1}{2}q_3^{(1)}, q_2^{(1)} = 1 - q_3^{(1)}$$

avec quoi on a

$$q_3^{(1)} = \frac{1}{4}, q_2^{(1)} = \frac{3}{4}, q_1^{(1)} = 0$$

Dans ce cas, on obtient immédiatemment $Q^{(2)}$, alors

$$p_2 = \frac{1}{2}q_2^{(1)} + \frac{1}{2}q_2^{(2)} \Longrightarrow q_2^{(2)} = \frac{1}{4}$$

$$q_1^{(2)} = 1 - q_2^{(2)} \Longrightarrow q_1^{(2)} = \frac{3}{4}$$

et $q_3^{(2)} = 0$. Donc,

$$Q^{(1)} = \left(0, \frac{3}{4}, \frac{1}{4}\right), Q^{(2)} = \left(\frac{3}{4}, \frac{1}{4}, 0\right).$$

Méthode des alias ou méthode de Walker :

Le procédé se généralise pour representer une masse de n composante de la forme (1).
Nous choisissons i_1, j_1 qui satisfont les conditions du lemme.
On définit $Q^{(1)}$ de manière qu'il se concentre en i_1 et j_1, et qu'il proportionne toute la masse pour le point i_1, en remarquant qu'à partir de la représentation (1), on doit avoir $q_{i_1}^{(k)} = 0$, pour $k = 2, \ldots, n-1$, et

$$q_{i_1}^{(1)} = (n-1)p_{i_1}$$

$$q_{j_1}^{(1)} = 1 - (n-1)p_{i_1}.$$

Alors,

$$P = \frac{1}{n-1}Q^{(1)} + \frac{n-2}{n-1}P^{(n-1)}$$

avec

$$p_{i_1}^{(n-1)} = 0$$

$$p_{j_1}^{(n-1)} = \frac{n-1}{n-2}\left(p_{i_1} + p_{j_1} - \frac{1}{n-1}\right)$$

$$p_k^{(n-1)} = \frac{n-1}{n-2} p_k, \qquad k \neq i_1 \text{ ou } j_1$$

$P^{(n-1)}$ se concentre, en total, en $n-1$ points et on peut trouver i_2, j_2 suivant le lemme, et $Q^{(2)}$ concentré en i_2, j_2 et qui donne toute la masse (de $P^{(1)}$) à i_2.

On a donc,

$$P^{(n-1)} = \frac{1}{n-2} Q^{(2)} + \frac{n-3}{n-2} P^{(n-2)}$$

avec quoi

$$P = \frac{1}{n-1} Q^{(1)} + \frac{1}{n-1} Q^{(2)} + \frac{n-3}{n-1} P^{(n-2)}$$

En répétant le processus, on arrive à la décomposition

$$P = \frac{1}{n-1} Q^{(1)} + \frac{1}{n-1} Q^{(2)} + \cdots + \frac{1}{n-1} Q^{(n-1)}$$

Ainsi, P est une mixture également pondérée de $n-1$ masses bi-ponctuelles avec les $Q^{(k)}$ qui donnent la masse à i_k et j_k.

Pour générer de X, il suffit générer uniformément un entier N de $\{1, \ldots, n-1\}$, générer un nombre aléatoire et sortir avec i_N, si ce nombre est inférieur à $q_{i_N}^{(N)}$, et avec j_N dans autre cas.
La simulation est, alors,

Générer $U_1, U_2 \sim \mathcal{U}(0,1)$
Faire $N = 1 + ent((n-1)U_1)$
Sortir

$$X = \begin{cases} i_N & \text{si } U_2 < q_{iN}^{(N)} \\ j_N & \text{autrement} \end{cases}$$

Méthode des alias ou méthode de Walker :

Remarques :

• La décomposition peut être présentée comme une table :

1	2	$\cdots$	$n-1$
i_1	i_2	$\cdots$	i_{n-1}
j_1	j_2	$\cdots$	j_{n-1}
$q_{i_1}^{(1)}$	$q_{i_2}^{(2)}$	$\cdots$	$q_{i_{n-1}}^{(n-1)}$

• La méthode s'appelle des alias parce que, par renumérotation des Q's, nous pouvons obtenir pour chaque k que $Q_k^{(k)} > 0$. Ainsi, le procédé simulera N uniformément (en $1, 2, \cdots, n-1$) et si $N = k$, ou bien on accepte la valeur de k comme X, ou bien celui de l'"alias" de k (c'est-à-dire, l'autre valeur à laquelle $Q^{(k)}$ donne une valeur peut-être positive).

Remarques :

• En effet, il n'est pas nécessaire de générer deux nombres aléatoires uniformes U_1, U_2. Comme $N-1 = ent((n-1)U_1)$, la partie fractionnaire de $(n-1)U_1 - (N-1)$ est indépendante de N et elle a une distribution uniforme sur $(0,1)$. Ainsi, au lieu de générer un nouveau nombre aléatoire en U_2, on peut utiliser

$$(n-1)U_1 - (N-1) = (n-1)U_1 - ent((n-1)U_1).$$

Dans l'exemple, nous avons obtenu la décomposition

1	$3/8$	$=$	$1/2(0+3/4)$
2	$1/2$	$=$	$1/2(3/4+1/4)$
3	$1/8$	$=$	$1/2(1/4+0)$

et de là

	1	2
i	3	2
j	2	1
q_i	$1/4$	$1/4$

Méthode des alias ou méthode de Walker :

Exemple :

Une paire d'itérations de l'algorithme peuvent être :

Générer $U_1 = 0.089013$

$N = 1 + ent(2U_1) = 1$

$U_2 = 2U_1 - (1 - 1) = 0.178026 < 1/4$

Sortir 3

Générer $U_1 = 0.089013$

$N = 1 + ent(2U_1) = 2$

$U_2 = 2U_1 - (2 - 1) = 0.7872 > 1/4$

Sortir 1