Podobnie jak w poprzednim przykładzie, odpowiednie warunki szybkiej akceptacji i szybkiej eliminacji mogą znacznie przyśpieszyć działanie algorytmu (patrz p. 3.2.3).

Zwróćmy uwagę na następujące dwa lematy:

LEMAT **3.5.** *Jeżeli punkt losowy $(U,V)$ ma rozkład równomierny na zbiorze $\{(u,v): 0 \le u \le f(u+v)\}$, to zmienna losowa $U+V$ ma rozkład o gęstości proporcjonalnej do $f$.* □

LEMAT **3.6.** *Jeżeli punkt losowy $(U,V)$ ma rozkład równomierny na zbiorze $\{(u,v): 0 \le u \le (f(v/\sqrt{u}))^{2/3}\}$, to zmienna losowa $V/\sqrt{U}$ ma rozkład o gęstości proporcjonalnej do $f$.* □

Te dwa lematy sugerują inne warianty oraz uogólnienia metody ROU.

### 3.1.5. Rozkłady dyskretne

#### Metoda odwracania dystrybuanty

Metoda odwracania dystrybuanty, o której mówiliśmy w p. 3.1.1, w przypadku rozkładów dyskretnych prowadziła do wzoru (3.3). Dla rozkładu prawdopodobieństwa $P\{X = k\} = p_k$, $k = 0,1,2,\ldots$ otrzymujemy prosty algorytm

ALGORYTM **3.19**

$X = 0$, $S = p_0$
*Generuj* $U$ *o rozkładzie równomiernym* $U(0,1)$
*While* $U > S$
*do* $X = X+1$, $S = S + p_X$
*Return* $X$

Łatwe uogólnienie tego algorytmu na przypadek rozkładu

$$P\{X = x_k\} = p_k, \quad k = 0,1,2,\ldots \tag{3.20}$$

pozostawiamy Czytelnikowi.

Liczba $N$ kroków w pętli *while* ma rozkład $P\{N = k\} = p_k$, może się więc okazać, że wartość oczekiwana tej zmiennej losowej jest bardzo duża lub nawet równa $+\infty$. Wynika stąd, że ten prosty algorytm może pracować bardzo wolno. Tak jest np. w niektórych dyskretnych wersjach rozkładu Pareto. Pewne przyśpieszenie algorytmu można uzyskać przez taką permutację ciągu liczb $p_0, p_1, p_2, \ldots$, żeby w nowym ciągu $p_{(0)}, p_{(1)}, p_{(2)}, \ldots$ były spełnione nierówności

$$p_{(0)} \ge p_{(1)} \ge p_{(2)} \ge \cdots$$

Jeżeli dokonamy takiej samej permutacji liczb $x_0, x_1, x_2, \ldots$, to zmienna losowa $Y$ o rozkładzie $P\{Y = x_{(k)}\} = p_{(k)}$, $k = 0,1,2,\ldots$, będzie miała taki sam rozkład jak zmienna losowa $X$, ale średnia liczba kroków w pętli *while* będzie mniejsza.

### Metoda równomiernego rozbicia przedziału $(0,1)$

Metodę odwracania dystrybuanty można interpretować w następujący sposób. Rozbijamy przedział $(0,1)$ na rozłączne podprzedziały o długościach równych $p_0, p_1, p_2, \ldots$ Każdemu podprzedziałowi przyporządkowujemy odpowiednią wartość zmiennej losowej: podprzedziałowi o długości $p_i$ wartość $x_i$. Generujemy liczbę losową $U$ o rozkładzie równomiernym $U(0,1)$ i za wynik generowania zmiennej losowej $X$ przyjmujemy tę wartość $x_i$, która odpowiada podprzedziałowi, do którego wpadło to $U$.

Metoda równomiernego rozbicia przedziału $(0,1)$ polega na podzieleniu przedziału $(0,1)$ na jednakowo długie podprzedziały, sprawdzeniu, do którego z nich wpadło $U$, a następnie na identyfikowaniu wartości generowanej zmiennej losowej $X$ już tylko w tym małym podprzedziale. Oto szczegóły.

Rozważamy zmienną losową $X$ przyjmującą skończoną liczbę różnych wartości
$$\mathrm{P}\{X=k\}=p_k, \qquad k=0,1,\ldots,K \tag{3.21}$$
Dzielimy przedział $(0,1)$ na $K+1$ podprzedziałów $(\frac{i-1}{K+1}, \frac{i}{K+1})$ o jednakowej długości i umawiamy się, że odcinek $(0, \frac{1}{K+1})$ ma numer 1. Zatem zmienna $U$ wpada do podprzedziału o numerze $[(K+1)U+1]$.

Dla danego rozkładu (3.21) tworzymy ciąg $q_i = \sum_{j=0}^{i} p_i$, $i=0,1,\ldots,K$, i przyjmujemy $q_{-1}=0$. Następnie tworzymy pomocniczy ciąg liczb
$$g_i = \max\left\{j\colon q_j < \frac{i}{K+1}\right\}, \qquad i=1,2,\ldots,K+1 \tag{3.22}$$
umawiając się przy tym, że maksimum na zbiorze pustym jest równe zeru. Po tych przygotowaniach algorytm generowania liczb losowych o rozkładzie (3.21) przyjmuje następującą postać:

ALGORYTM **3.20**

*Generuj $U$ o rozkładzie równomiernym* $U(0,1)$
  $X = [(K+1)U+1]$
  $X = g_X + 1$
  *While* $q_{X-1} > U$ *do* $X = X-1$
*Return* $X$

Najważniejsza własność tego algorytmu jest opisana w następującym twierdzeniu:

TWIERDZENIE **3.5.** *Oczekiwana liczba porównań $\{q_{X-1} > U\}$ jest nie większa od 2.*

D o w ó d. Jeżeli w pierwszym kroku algorytmu został wybrany podprzedział o numerze $i$, to liczba porównań $\{q_{X-1} > U\}$ nie będzie większa od liczby różnych wartości $q_j$ w tym podprzedziale, plus jeden. Każdy z $K+1$ podprzedziałów jest wybierany z jednakowym prawdopodobieństwem, więc oczekiwana liczba porównań $\leq$

$$\leq 1 + \frac{1}{K+1} \sum_{i=0}^{K} (\text{liczba różnych } q_j \text{ w } i\text{-tym przedziale}) = 2$$

□

Pomysł metody równomiernego rozbicia przedziału $(0, 1)$ pochodzi z pracy Chena i Asau (1974), a opis tej metody podaliśmy według monografii Devroya (1986). W literaturze anglojęzycznej ta metoda nosi nazwę *method of guide tables.*

### Mieszanina rozkładów dwupunktowych

Mówimy, że rozkład o gęstości $f(x)$ jest *mieszaniną k rozkładów* o gęstościach $f_i(x)$, $i = 0, 1, \ldots, k-1$, jeżeli

$$f(x) = \sum_{i=0}^{k-1} \lambda_i f_i(x)$$

gdzie

$$\sum_{i=0}^{k-1} \lambda_i = 1, \qquad \lambda_i > 0, \qquad i = 0, 1, \ldots, k-1$$

Weźmy pod uwagę zmienną losową $X$ o rozkładzie dyskretnym

$$\mathrm{P}\{X = x_i\} = p_i, \quad i = 0, 1, \ldots, k-1 \tag{3.23}$$

Mówimy, że rozkład (3.23) jest *równomierną mieszaniną rozkładów dwupunktowych,* jeżeli istnieją ciąg par liczb $(u_j, v_j)$, $j = 0, 1, \ldots, k-1$, oraz ciąg liczb $q_j \in (0, 1)$, $j = 0, 1, \ldots, k-1$, takie że

$$p_i = \frac{1}{k} \sum_{j=0}^{k-1} \Big( q_j \mathbf{1}_{\{u_j\}}(x_i) + (1 - q_j) \mathbf{1}_{\{v_j\}}(x_i) \Big), \quad i = 0, 1, \ldots, k-1 \tag{3.24}$$

Generowanie liczb losowych $X$ o rozkładzie dyskretnym (3.23), który jest równomierną mieszaniną rozkładów dwupunktowych, jest bardzo łatwe i szybkie: wystarczy wygenerować indeks $J$ według rozkładu równomiernego na zbiorze $\{0, 1, \ldots, k-1\}$, a następnie wygenerować $X$ według $J$-tego rozkładu dwupunktowego

$$\mathrm{P}\{X = u_J\} = q_J, \quad \mathrm{P}\{X = v_J\} = 1 - q_J$$

TWIERDZENIE **3.6.** *Każdy rozkład dyskretny postaci (3.23) można przedstawić w postaci równomiernej mieszaniny (3.24) rozkładów dwupunktowych.*

D o w ó d (indukcyjny). Dla $k=1$ mamy oczywiście $u_0=x_0$ i $q_0=p_0(=1)$. Przypuśćmy, że teza jest prawdziwa dla rozkładów $(k-1)$-punktowych i rozpatrzmy $k$-punktowy rozkład (3.23).

Niech $i_0$ będzie takim wskaźnikiem, że

$$p_{i_0} = \min_{0\le i<k} p_i$$

Podstawmy $u_0 = x_{i_0}$ oraz $q_0 = kp_{i_0}$ (mamy oczywiście $p_{i_0} \le 1/k$, więc $kp_{i_0} \le 1$). Niech $j_0$ będzie takim wskaźnikiem, że

$$p_{j_0} = \max_{0\le i<k} p_i$$

Mamy oczywiście $p_{j_0} \ge 1/k$, więc $(1-q_0)/k \le p_{j_0}$. Podstawmy $v_0 = x_{j_0}$. Wielkości $(u_0, v_0)$ oraz $q_0$ potraktujmy jako pierwszy składnik sumy (3.24). Liczby

$$p_i' = p_i - \frac{1}{k}\Big(q_0 \mathbf{1}_{\{u_0\}}(x_i) + (1-q_0)\mathbf{1}_{\{v_0\}}(x_i)\Big), \quad 0 \le i < k$$

są nieujemne i sumują się do $1 - \dfrac{1}{k}$, przy czym $p'_{i_0} = 0$. Teraz

$$\begin{pmatrix} x_0 & x_1 & \dots & x_{i_0-1} & x_{i_0+1} & \dots x_{k-1} \\ \tilde{p}_0 & \tilde{p}_1 & \dots & \tilde{p}_{i_0-1} & \tilde{p}_{i_0+1} & \dots \tilde{p}_{k-1} \end{pmatrix}$$

gdzie $\tilde{p}_i = kp_i'/(k-1)$, jest $(k-1)$-punktowym rozkładem prawdopodobieństwa, który na mocy założenia indukcyjnego może być przedstawiony w postaci (3.24), co kończy dowód. □

Algorytm korzysta z trzech następujących bloków liczb: $(q_0, q_1, \dots, q_{k-1})$, $(u_0, u_1, \dots, u_{k-1})$ oraz $(v_0, v_1, \dots, v_{k-1})$ i przebiega według następującego schematu:

ALGORYTM **3.21**

*Generuj* $U_1$ *o rozkładzie równomiernym* $U(0,1)$
*Generuj* $U_2$ *o rozkładzie równomiernym* $U(0,1)$
$I = [kU_1]$
*If* $U_2 \le q_I$ *then* $X = u_I$ *else* $X = v_I$
*Return* $X$

Ten algorytm dwukrotnie odwołuje się do generowania liczb losowych o rozkładzie równomiernym $U(0,1)$; korzystając z następującego lematu, można te dwa odwołania zredukować do jednego.

LEMAT **3.7.** *Jeżeli zmienna losowa $U$ ma rozkład równomierny $U(0,1)$, to zmienna losowa $[kU]$ ma rozkład równomierny na zbiorze $\{0,1,\dots,k-1\}$, zmienna losowa $\{kU\}$ ma rozkład równomierny $U(0,1)$ i te zmienne losowe są niezależne.* □

D o w ó d pozostawiamy Czytelnikowi.