4.6 The Alias Method for Generating Discrete Random Variables

In this section we study a technique for generating discrete random variables which, although requiring some setup time, is very fast to implement.

In what follows, the quantities $\mathbf{P}, \mathbf{P}^{(k)}, \mathbf{Q}^{(k)}, k \leq n-1$, represent probability mass functions on the integers $1, 2, \ldots, n$—that is, they are n-vectors of nonnegative numbers summing to 1. In addition, the vector $\mathbf{P}^{(k)}$ has at most k nonzero components, and each of the $\mathbf{Q}^{(k)}$ has at most two nonzero components. We show that any probability mass function $\mathbf{P}$ can be represented as an equally weighted mixture of $n-1$ probability mass functions $\mathbf{Q}$ (each having at most two nonzero components). That is, we show, for suitably defined $\mathbf{Q}^{(1)}, \ldots, \mathbf{Q}^{(n-1)}$, that $\mathbf{P}$ can be expressed as

$$\mathbf{P} = \frac{1}{n-1}\sum_{k=1}^{n-1} \mathbf{Q}^{(k)} \tag{4.4}$$

As a prelude to presenting the method for obtaining this representation, we need the following simple lemma whose proof is left as an exercise.

Lemma Let $P = \{P_i, i = 1, \ldots, n\}$ denote a probability mass function. Then

(a) there exists an i, $1 \leq i \leq n$, *such that* $P_i < 1/(n-1)$, and
(b) for this i there exists a j, $j \neq i$, such that $P_i + P_j \geq 1/(n-1)$.

Before presenting the general technique for obtaining the representation (4.4), let us illustrate it by an example.

Example 4h Consider the three-point distribution $\mathbf{P}$ with $P_1 = \frac{7}{16}$, $P_2 = \frac{1}{2}$, $P_3 = \frac{1}{16}$. We start by choosing i and j satisfying the conditions of the preceding lemma. Since $P_3 < \frac{1}{2}$ and $P_3 + P_2 \geq \frac{1}{2}$, we can work with $i = 3$ and $j = 2$. We now define a two-point mass function $\mathbf{Q}^{(1)}$, putting all its weight on 3 and 2 and such that $\mathbf{P}$ is expressible as an equally weighted mixture between $\mathbf{Q}^{(1)}$ and a second two-point mass function $\mathbf{Q}^{(2)}$. In addition, all the mass of point 3 is contained in $\mathbf{Q}^{(1)}$. As we have

$$P_j = \frac{1}{2}(Q_j^{(1)} + Q_j^{(2)}), \quad j = 1, 2, 3 \tag{4.5}$$

and $Q_3^{(2)}$ is supposed to equal 0, we must therefore take

$$Q_3^{(1)} = 2P_3 = \frac{1}{8}, \qquad Q_2^{(1)} = 1 - Q_3^{(1)} = \frac{7}{8}, \qquad Q_1^{(1)} = 0$$

To satisfy (10.2), we must then set

$$Q_3^{(2)} = 0, \qquad Q_2^{(2)} = 2P_2 - \frac{7}{8} = \frac{1}{8}, \qquad Q_1^{(2)} = 2P_1 = \frac{7}{8}$$

Hence we have the desired representation in this case. Suppose now that the original distribution was the following four-point mass function:

$$P_1 = \frac{7}{16}, \qquad P_2 = \frac{1}{4}, \qquad P_3 = \frac{1}{8}, \qquad P_4 = \frac{3}{16}$$

Now $P_3 < \frac{1}{3}$ and $P_3 + P_1 \geq \frac{1}{3}$. Hence our initial two-point mass function —$\mathbf{Q}^{(1)}$ — concentrates on points 3 and 1 (giving no weight to 2 and 4). Because the final representation gives weight $\frac{1}{3}$ to $\mathbf{Q}^{(1)}$ and in addition the other $\mathbf{Q}^{(j)}$, $j = 2, 3$, do not give any mass to the value 3, we must have that

$$\frac{1}{3}Q_3^{(1)} = P_3 = \frac{1}{8}$$

Hence

$$Q_3^{(1)} = \frac{3}{8}, \qquad Q_1^{(1)} = 1 - \frac{3}{8} = \frac{5}{8}$$

Also, we can write

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

where $\mathbf{P}^{(3)}$, to satisfy the above, must be the vector

$$\begin{aligned}
\mathbf{P}_1^{(3)} &= \frac{3}{2}\left(P_1 - \frac{1}{3}Q_1^{(1)}\right) = \frac{11}{32} \\
\mathbf{P}_2^{(3)} &= \frac{3}{2}P_2 = \frac{3}{8} \\
\mathbf{P}_3^{(3)} &= 0 \\
\mathbf{P}_4^{(3)} &= \frac{3}{2}P_4 = \frac{9}{32}
\end{aligned}$$

Note that $\mathbf{P}^{(3)}$ gives no mass to the value 3. We can now express the mass function $\mathbf{P}^{(3)}$ as an equally weighted mixture of two-point mass functions $\mathbf{Q}^{(2)}$ and $\mathbf{Q}^{(3)}$, and we end up with

$$\begin{aligned}
\mathbf{P} &= \frac{1}{3}\mathbf{Q}^{(1)} + \frac{2}{3}\left(\frac{1}{2}\mathbf{Q}^{(2)} + \frac{1}{2}\mathbf{Q}^{(3)}\right) \\
&= \frac{1}{3}(\mathbf{Q}^{(1)} + \mathbf{Q}^{(2)} + \mathbf{Q}^{(3)})
\end{aligned}$$

(We leave it as an exercise for the reader to fill in the details.) $\square$

The above example outlines the following general procedure for writing the n-point mass function $\mathbf{P}$ in the form (4.4), where each of the $\mathbf{Q}^{(i)}$ are mass functions giving all their mass to at most two points. To start, we choose i and j satisfying the conditions of the lemma. We now define the mass function $\mathbf{Q}^{(1)}$ concentrating on the points i and j and which contain all the mass for point i by noting that in the representation (4.4) $Q_i^{(k)} = 0$ for $k = 2, \ldots, n-1$, implying that

$$Q_i^{(1)} = (n-1)P_i \qquad \text{and so } Q_j^{(1)} = 1 - (n-1)P_i$$

Writing

$$\mathbf{P} = \frac{1}{n-1}\mathbf{Q}^{(1)} + \frac{n-2}{n-1}\mathbf{P}^{(n-1)} \tag{4.6}$$

where $\mathbf{P}^{(n-1)}$ represents the remaining mass, we see that

$$\begin{aligned}
P_i^{(n-1)} &= 0 \\
P_j^{(n-1)} &= \frac{n-1}{n-2}\left(P_j - \frac{1}{n-1}Q_j^{(1)}\right) = \frac{n-1}{n-2}\left(P_i + P_j - \frac{1}{n-1}\right) \\
P_k^{(n-1)} &= \frac{n-1}{n-2}P_k, \quad k \neq i \text{ or } j
\end{aligned}$$

That the above is indeed a probability mass function is easily checked—for example, the nonnegativity of $P_j^{(n-1)}$ follows from the fact that j was chosen so that $P_i + P_j \geq 1/(n-1)$.

We may now repeat the above procedure on the $(n-1)$ point probability mass function $\mathbf{P}^{(n-1)}$ to obtain

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

and thus from (4.6) we have

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

We now repeat the procedure on $P^{(n-2)}$ and so on until we finally obtain

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

In this way we are able to represent $\mathbf{P}$ as an equally weighted mixture of $n-1$ two-point mass functions. We can now easily simulate from $\mathbf{P}$ by first generating a random integer N equally likely to be either $1, 2, \ldots, n-1$. If the resulting value N is such that $\mathbf{Q}^{(N)}$ puts positive weight only on the points i_N and j_N, we can set X equal to i_N if a second random number is less than $\mathbf{Q}_{i_N}^{(N)}$ and equal to j_N otherwise. The random variable X will have probability mass function $\mathbf{P}$. That is, we have the following procedure for simulating from $\mathbf{P}$.

STEP 1: Generate U_1 and set $N = 1 + \text{Int}[(n-1)U_1]$.
STEP 2: Generate U_2 and set

$$X = \begin{cases} i_N & \text{if } U_2 < Q_{i_N}^{(N)} \\ j_N & \text{otherwise} \end{cases}$$

Remarks

1. The above is called the alias method because by a renumbering of the $\mathbf{Q}$'s we can always arrange things so that for each k, $Q_k^{(k)} > 0$. (That is, we can arrange things so that the kth two-point mass function gives positive weight to the value k.) Hence, the procedure calls for simulating N, equally likely to be $1, 2, \ldots, n-1$, and then if $N = k$ it either accepts k as the value of X, or it accepts for the value of X the "alias" of k (namely, the other value that $\mathbf{Q}^{(k)}$ gives positive weight).
2. Actually, it is not necessary to generate a new random number in Step 2. Because $N-1$ is the integer part of $(n-1)U_1$, it follows that the remainder $(n-1)U_1 - (N-1)$ is independent of N_1 and is uniformly distributed on (0, 1). Hence, rather than generating a new random number U_2 in Step 2, we can use $(n-1)U_1 - (N-1)$. □

4.7 Generating Random Vectors

A random vector $X_1, \ldots, X_n$ can be simulated by sequentially generating the X_i. That is, first generate X_1; then generate X_2 from its conditional distribution given the generated value of X_1; then generate X_3 from its conditional distribution given the generated values of X_1 and X_2; and so on. This is illustrated in Example 4i, which shows how to simulate a random vector having a multinomial distribution.

Example 4i Consider n independent trials, each of which results in one of the outcomes $1, 2, \ldots, r$ with respective probabilities $p_1, p_2, \ldots, p_r$, $\sum_{i=1}^r p_i = 1$. If X_i denotes the number of trials that result in outcome i, then the random vector $(X_1, \ldots, X_r)$ is said to be a multinomial random vector. Its joint probability mass function is given by

$$P\{X_i = x_i, i = 1, \ldots, r\} = \frac{n!}{x_1! \cdots x_r!} p_1^{x_1} \cdots p_r^{x_r}, \quad \sum_{i=1}^r x_i = n$$

The best way to simulate such a random vector depends on the relative sizes of r and n. If r is large relative to n, so that many of the outcomes do not occur on any of the trials, then it is probably best to simulate the random variables by generating