variables have finite expectation; however, problem 1 at the end of this section provides an example of a random variable whose expectation does not exist. Definition (4.1) can be extended to the case of mixed random variables through the use of Riemann-Stieltjes integral. Alternatively, the formula given in problem 2 at the end of this section can be used in the general case.

Example 4.1

Consider the problem of searching for a specific name in a table of names. A simple method is to scan the table sequentially, starting from one end, until we either find the name or reach the other end, indicating that the required name is missing from the table. The following is a Pascal program fragment for sequential search:

```
var T: array[0..n] of NAME;
Z: NAME;
I: 0..n;
begin {Z has been initialized elsewhere}
  T[0]:=Z; {T[0] is used as a sentinel or marker}
  I:=n;
  while Z ≠ T[I] do
    I:=I - 1;
  if I > 0 then {found; I points to Z}
   else {not found}.
end
```

In order to analyze the time required for sequential search, let X be the discrete random variable denoting the number of comparisons "$Z \neq T[I]$" made. Clearly, the set of all possible values of X is $\{1, 2, \ldots, n+1\}$, and $X = n+1$ for unsuccessful searches. Since the value of X is fixed for unsuccessful searches, it is more interesting to consider a random variable Y that denotes the number of comparisons on a successful search. The set of all possible values of Y is $\{1, 2, \ldots, n\}$. To compute the average search time for a successful search, we must specify the pmf of Y. In the absence of any specific information, it is natural to assume that Y is uniform over its range; that is:

$$p_Y(i) = \frac{1}{n}, \qquad 1 \leqslant i \leqslant n.$$

Then

$$E[Y] = \sum_{i=1}^{n} i p_Y(i) = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Thus, on the average, approximately half the table needs to be searched. #

Example 4.2

The assumption of uniform distribution, used in Example 4.1, rarely holds in practice. It is possible to collect statistics on access patterns and use empirical distributions to reorganize the table so as to reduce the average search time. Unlike Example 4.1, we now assume for convenience that table search starts from the front. If α_i denotes the access probability for name $T[i]$, then the average successful search time $E[Y] = \sum i\alpha_i$. Then $E[Y]$ is minimized when names in the table are in the order of nonincreasing access probabilities; that is, $\alpha_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_n$. As an example, many tables in practice follow Zipf's law [ZIPF 1949]:

$$\alpha_i = \frac{c}{i}, \qquad 1 \leqslant i \leqslant n,$$

where the constant c is determined from the normalization requirement, $\sum_{i=1}^{n} \alpha_i = 1$.

Thus:

$$c = \frac{1}{\sum_{i=1}^{n} \frac{1}{i}} = \frac{1}{H_n} \simeq \frac{1}{\ln(n)},$$

where H_n is the partial sum of a harmonic series; that is: $H_n = \sum_{i=1}^{n} \frac{1}{i}$.

Now, if the names in the table are ordered as above, then the average search time is

$$E[Y] = \sum_{i=1}^{n} i\alpha_i = \frac{1}{H_n} \sum_{i=1}^{n} 1 = \frac{n}{H_n} \simeq \frac{n}{\ln(n)},$$

which is considerably less than the previous value $(n+1)/2$, for large n. #

Example 4.3

Recall the example of a computer system with five tape drives (Examples 1.1 and 2.2) and let X be the number of available tape drives. Then:

$$\begin{aligned} E[X] &= \sum_{i=0}^{5} i p_X(i) \\ &= 0 \cdot 1/32 + 1 \cdot 5/32 + 2 \cdot 10/32 + 3 \cdot 10/32 + 4 \cdot 5/32 + 5 \cdot 1/32 \\ &= 2.5. \end{aligned}$$

#

The example above illustrates that $E[X]$ need not correspond to a possible value of the random variable X. The expected value denotes the "center" of a probability distribution in the sense of a weighted average, or better, in the sense of a center of gravity.

Example 4.4

Let X be a continuous random variable with an exponential density given by:

$$f(x) = \lambda e^{-\lambda x}, \qquad x > 0.$$

Then

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{0}^{\infty} \lambda x e^{-\lambda x}\, dx.$$