Lecture #18: Probability Distributions
Lecture Presentation

Review of Preparatory Material

Using Discrete Probability Theory to Model and Solve Problems

1. Suppose you have two blue mittens and two grey mittens in a drawer. It is dark so you pick two mittens without being able to see what you are choosing. How likely is it that you pick two mittens with the same colour?

   Sample Space:

What assumptions might you make?

What Probability Distribution corresponds to this?

How likely is it that you pick two mittens with the same colour?

2. Suppose you have a deck of ten cards, After shuffling the deck, you show the top card. You put the card back (so that you have ten cards, once again), shuffling the deck and showing the top card, once again. How likely is it that you show the same card, both times?

   Sample Space:

What assumptions might you make?

What Probability Distribution does this correspond to?

How likely is it that you see the same card, both times?

3. Suppose you roll a die (whose surfaces show the numbers 1, 2, 3, 4, 5 and 6) over and over again, until you see the same number (at the top) two times in a row. How likely is it that you need to roll the die ***more than four times***?

   Sample Space:

What assumptions might you make?

What Probability Distribution does this correspond to?

How likely is it that you need to roll the die *more than four times*?

Sometimes the "Classical" Sample Spaces are Useful — Even If You Don't See Them Right Away

Consider the representation of a ***finite set***

$$\{k_1, k_2, \ldots, k_n\}$$

of n values from a larger set, or "universe", $\mathcal{U}$. A ***hash table with chaining*** is a data structure — often discussed in a course like CPSC 331 — that can be used to represent this kind of set.

For a positive integer m — which we will call the ***table size***, —a ***hash function*** (for the universe U and table size m) is a total function

$$h : \mathcal{U} \to \{0, 1, \ldots, m-1\}$$

mapping each element $k \in U$ to an integer $h(k)$, such that $0 \leq h(k) \leq m-1$. Since the hash table (used to represent the above set $\{k_1, k_2, \ldots, k_n\}$) only depends on the "hash values" $h(k_1), h(k_2), \ldots, h(k_n)$, each ***outcome*** can be represented as a sequence

$$(\alpha_1, \alpha_2, \ldots, \alpha_n)$$

where $\alpha_i = h(k_i) \in \{0, 1, 2, \ldots, m-1\}$ for each integer i such that $1 \leq i \leq n$. That is, the ***sample space***, to be used here, is the set

$$\Omega_{n,m} = \{(\alpha_1, \alpha_2, \ldots, \alpha_n) \mid \alpha_i \in \mathbb{Z} \text{ and } 0 \leq i \leq m-1 \text{ for every integer } i \text{ such that } 1 \leq i \leq k\}.$$

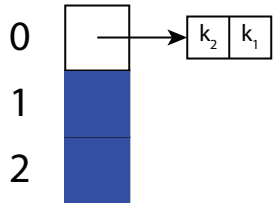
In general, a hash table representing the set $\{k_1, k_2, \ldots, k_n\}$, with table size m, is an ***array*** T with length m. For each integer i such that $0 \leq i \leq m-1$, $T[i]$ is a reference to a *list* of the values

$$S_i = \{k_j \mid 1 \leq j \leq n \text{ and } h(k_j) = i\},$$

where $h : \mathcal{U} \to \{0, 1, 2, \ldots, m-1\}$ is the hash function used to construct this hash table.

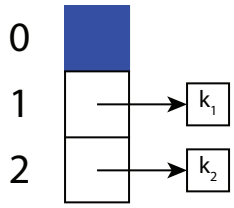
In particular, the hash table is constructed by starting with an "empty" table — and array T such that $T[i]$ is a reference to an empty list, for each integer i such that $0 \leq i \leq m-1$. Then, for $1 \leq j \leq n$ (in increasing order), key k_j is inserted at the *front* of the linked list at position $T[h(k_j)]$.

For example, if $m = 3$ and $n = 2$, then the hash table corresponding to the outcome $(0, 0)$ (where $h(k_1) = h(k_2) = 0$) is as follows.



In this and later pictures, a cell shown in blue stores a reference to an empty list.

On the other hand, if $m = 3$ and $n = 2$, then the hash table corresponding to the outcome $(1, 2)$ (where $h(k_1) = 1$ and $h(k_2) = 2$) is as follows.



If time permits then this "Hash Table" example will be continued in later presentations.

With that noted, an examination of the sample space, described above, should suggest that it is just a "disguised" version of one of the sample spaces for one of the "classical" experiments that are described in the lecture notes. ***Which One?***

Why This is Helpful: It is evidence that we will (possibly) be able to apply analyses for "classical" experiments to solve problems in computer science.