Algorithms and Data Structures

Chapter 5

Catherine Durso

cdurso@cs.du.edu

Chapter 5

Chapter 5 provides an oportunity to review some common concepts from Probability, as well as introducing the technique of using indicator random variables to simplify computation of expected values. The motivation for studying these concepts here comes from the need to do average case analysis for algorithms with widely differing performance on different instances of problems of the same size, and also for the need to design and analyze randomized algorithms.

Functions for Analysis

Example problem:

$HIRE - ASSISTANT(n)$

```
best = 0
    for i in 1 to n
        interview candidate i
        if candidate i is better than candidate best
            best = i
            hire candidate i
```

The number of hires depends on the order in which candidates are interviewed. What is the best case? What is the worst case? If the ranks of the candidates are equally likely to be in any order, what is the expected number of hires?

Randomized Algorithm

An algorithm is *randomized* if its behavior is determined by the input and by values produced by a random number generator.

If the distribution of performances of a randomized algorithm on a problem of size n does not depend on the instance, the expected performance (running time, say) is the average over all sequences of values returned by the random number generator, weighted by the probability of the sequence.

Discrete Probability Spaces

For our purposes, a discrete probability space $(\mathscr{S}, Pr)$ consists of a finite or countably infinite set $\mathscr{S}$ of outcomes and a function $Pr : \mathscr{S} \rightarrow \mathbb{R}$ with the following properties:

- $0 \leq Pr(s) \leq 1$ for all $s \in \mathscr{S}$
- $\sum_{s \in \mathscr{S}} Pr(s) = 1$

Events

In a discrete probability space $(\mathscr{S}, Pr)$ such as described above, an *event* A is simply a subset of $\mathscr{S}$.
The function Pr can be extended to the power set of $\mathscr{S}$ by defining
$Pr(A) = \sum_{s \in A} Pr(s)$.
Now

- $0 \leq Pr\,(A) \leq 1$ for all $A \subseteq \mathscr{S}$
- $Pr\,(\mathscr{S}) = 1$

Example

Many programming languages offer (pseudo) random number generators. In the pseudocode in the text, for integers a and b with $a \leq b$, a call to RANDOM(a, b) returns an integer in $\{a, a+1..b\}$, with each value equally likely.

We can model this with a probability space with

$$\mathscr{S} = \{a, a+1..b\} \text{ and } Pr(x) = \begin{cases} 0 & x \notin \mathscr{S} \\ \frac{1}{b-a+1} & x \in \mathscr{S} \end{cases}.$$

If $a = 0$ and $b = 9$ and $A = \{x | x \text{ is a power of 2}\}$, then $Pr\left(A\right) = Pr\left(\{1, 2, 4, 8\}\right) = 4\left(\frac{1}{9-0+1}\right) = .4$.

Independence

Two events, A and B, are ***independent*** if $Pr\left(A \cap B\right) = Pr\left(A\right) Pr\left(B\right)$. Intuitively, the two events are independent if knowing $s \in A$ does not change the probability that $s \in B$.

For example, suppose we model the result of rolling a die twice and recording the resulting ordered pair of numbers as a probability space with $\mathscr{S} = \{(a,b)\,|a \in \{1,2,3,4,5,6\} \wedge b \in \{1,2,3,4,5,6\}\}$ and $Pr(s) = \frac{1}{36}$. Then the event $A = \{(a,b)\,|a = b\}$ and $B = \{(a,b)\,|b = 6\}$are independent: $Pr\left(A\right) = \frac{6}{36}$ and $Pr\left(B\right) = \frac{6}{36}$, while $Pr\left(A \cap B\right) = \frac{1}{36} = \frac{6}{36} \cdot \frac{6}{36}$.

More Independence

However, $C = \{(a, b) \,|a \geq b\}$ and B are not independent: $Pr\,(C) = \frac{1}{36}\sum_{i=1}^{6} i = \frac{1}{36} \cdot \frac{42}{2} = \frac{21}{36}$, $Pr\,(B) = \frac{6}{36}$, but $Pr\,(C \cap B) = \frac{1}{36}$.

Succesive values of RANDOM (a, b) are usually modeled as independent.

Random Variables

Given a discrete probability space $(\mathscr{S}, Pr)$, a *random variable* on $(\mathscr{S}, Pr)$ is a function $T : \mathscr{S} \to \mathbb{R}$.

If T is a random variable on a discrete probability space $(\mathscr{S}, Pr)$, the *expected value* of T, $E\left[T\right]$, is $\sum_{s \in \mathscr{S}} T\left(s\right) Pr(s)$, a weighted average of $\{T\left(s_1\right), T\left(s_2\right) ...\}$.

Example

Let $\mathscr{S}$ be a standard deck of 52 cards, with face values $\{1, 2..13\}$ in each of four suits. To model a random draw, define $Pr(s) = \frac{1}{52}$. Define $T(s) =$ face value of s.

Then

$E[T] = \sum_{s\in\mathscr{S}} T(s)\, Pr(s)$

$= \sum_{s\in\mathscr{S}} T(s)\, \frac{1}{52}$

$= \sum_{s:T(s)=1} 1 \cdot \frac{1}{52} + \sum_{s:T(s)=2} 2 \cdot \frac{1}{52} + ... \sum_{s:T(s)=13} 13 \cdot \frac{1}{52}$

$= 4 \cdot 1 \cdot \frac{1}{52} + 4 \cdot 2 \cdot \frac{1}{52} + ...4 \cdot 13 \cdot \frac{1}{52}$

$= \frac{4}{52} \sum_{i=1}^{13} i$

$= \frac{1}{13} \frac{13(14)}{2} = 7$

Linearity

A key feature of expected values that the expected value of a sum of random variables on the same space is just the sum of the expected values: if $T_1 : \mathscr{S} \to \mathbb{R}$ and $T_2 : \mathscr{S} \to \mathbb{R}$, $a_1, a_2 \in \mathbb{R}$, and $a_1T_1 + a_2T_2 : \mathscr{S} \to \mathbb{R}$ is defined by $a_1T_1 + a_2T_2(s) = a_1T_1(s) + a_2T_2(s)$, then $E[a_1T_1 + a_2T_2] = a_1E[T_1] + a_2E[T_2]$.

As a consequence, if $X_1, ..X_n$are random variables on $(\mathscr{S}, Pr)$, then $\sum_{i=1}^{n} X_i$ is a random variable on $(\mathscr{S}, Pr)$, and $E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$.

Linearity Example

Model a random draw from a deck of cards as above.
Let the random variable T_1 be the face value, and let T_2 be defined by $T_2(s) = \begin{cases} 1 & s \text{ is a heart} \\ 0 & \text{otherwise} \end{cases}$.

$$E[T_1 + T_2] = \sum_{s\in\mathscr{S}} (T_1 + T_2)(s)\, Pr(s)$$
$$= \sum_{s\in\mathscr{S}} (T_1(s) + T_2(s))\, Pr(s)$$
$$= \sum_{s\in\mathscr{S}} (T_1(s)\, Pr(s) + T_2(s)\, Pr(s))$$
$$= \sum_{s\in\mathscr{S}} T_1(s)\, Pr(s) + \sum_{s\in\mathscr{S}} T_2(s)\, Pr(s)$$
$$= E[T_1] + E[T_2] = 7 + 1 \cdot \frac{13}{52} + 0 \cdot \frac{39}{52} = 7.25$$

Indicator Random Variables

If $A \subseteq \mathscr{S}$,the *indicator random variable* $X_A : \mathscr{S} \to \mathbb{R}$ is defined by $X_A(s) = \begin{cases} 0 & s \notin A \\ 1 & s \in A \end{cases}$.

In the example above, T_2 is the indicator random variable for the set $\{s | s \text{ is a heart}\}$.

Note $E[X_A] = Pr(A)$:
$E[X_A] = \sum_{s \in \mathscr{S}} X_A(s) Pr(s)$
$= \sum_{s \in A} X_A(s) Pr(s) + \sum_{s \in \mathscr{S} - A} X_A(s) Pr(s)$
$= \sum_{s \in A} 1 \cdot Pr(s) + \sum_{s \in \mathscr{S} - A} 0 \cdot Pr(s)$
$= Pr(A)$.

Example

Let $\mathscr{S}$ be the set of permutations of $\{1, 2, 3, 4, 5\}$, each equally likely. Let $T(\sigma)$ be the number of values in the permutation σ that are exactly 1 greater then their predecessor in the permutation. For example $T(45312) = 2$. What is the expected value of T?

Use indicator random variables. Let A_2 be the event that the second element is exactly 1 greater than the first.
$Pr(A_2) = 4 \cdot 3! \frac{1}{5!} = \frac{1}{5}$. Define A_3, A_4 and A_5 similarly. Each event A_i has $Pr(A_i) = \frac{1}{5}$.

The sum $X_{A_2} + X_{A_3} + X_{A_4} + X_{A_5} = T$. But
$E[X_{A_i}] = Pr(A_i) = \frac{1}{5}$ so
$E[T] = E[X_{A_2} + X_{A_3} + X_{A_4} + X_{A_5}] = \sum_{i=2}^{5} E[X_{A_i}] = \frac{4}{5}$.

Hiring Problem, cont.

We can just consider the ranks of the candidates, with 1 the lowest and n the highest. Then $\mathscr{S}$ is the set of all permutations of $\{1, 2, ..n\}$. If each permutation of ranks is equally likely, $Pr(\sigma) = \frac{1}{n!}$.

The expected number of hires is the expected number of positions i in a permutation $\sigma = \sigma_1\sigma_2...\sigma_n$ satisfying $\sigma_i > \sigma_j$ for all $j < i$.

Hiring Events

Set A_i to be the event that the i^{th} candidate is hired, i.e. $\sigma_i > \sigma_j$ for all $j < i$.

$Pr\left(A_i\right) = \left(\begin{array}{c} n \\ i \end{array}\right) (i-1)!\,(n-i)!\frac{1}{n!} = \frac{1}{i}$ because we can have any choice of i ranks in the first i positions, but the largest must be in the i^{th} position.

Set X_{A_i} to be the indicator random variable for A_i.

Then the number of hires in a permutation σ is $\sum_{i=1}^{n} X_{A_i}(\sigma)$. Conclude that the expected number of hires is $\sum_{i=1}^{n} E\left[X_{A_i}\right] = \sum_{i=1}^{n} \frac{1}{i}$.

The appendix equation A.7 gives that $\sum_{i=1}^{n} \frac{1}{i} = \ln(n) + O(1)$.

Technical Aside

$\sum_{i=1}^{n} \frac{1}{i} = \ln(n) + O(1)$ comes from $\sum_{i=2}^{n} \frac{1}{i} \leq \int_{x=1}^{n} \frac{1}{x} dx = \ln n \leq \sum_{i=1}^{n-1} \frac{1}{i}$. (Draw the graph, and consider the upper sum and the lower sum for rectangles of width 1.) Thus $\sum_{i=1}^{n} \frac{1}{i} \leq \ln n + 1$, from

the left side of the inequality, and $\sum_{i=1}^{n} \frac{1}{i} \geq \ln n + \frac{1}{n}$, from the right.

Conclude $\ln n + \frac{1}{n} \leq \sum_{i=1}^{n} \frac{1}{i} \leq \ln n + 1$. This is much fewer than the worst case.

Randomized Algorithm

The calculations above only apply if we can model all rank orders as equally likely. We can guarantee this by permuting the input order in such a way that all permutations are equally likely.

$RANDOMIZED - HIRE - ASSISTANT(n)$

1 randomly permute the list of candidates

2 $HIRE - ASSISTANT(n)$

One common application of randomization in algorithms is to arrange that no particular input gives worst-case behavior. In that way, a regrettable relationship between the distribution of the inputs and the behavior of the algorithm can be avoided.

Permutation

Permutation can be accomplished using a random number generator.

$RANDOMIZE-IN-PLACE(A)$

$n = A.length$
for $i = 1$ to n
  swap $A[i]$ with $A[RANDOM(i, n)]$

What is the running time?
Claim that this routine yields each possible permutation of A with equal probability.

Uniform Permutation

Loop invariant: Just prior to the ith iteration of the for-loop, $A[1..i-1]$ contains each possible permutation of length $i-1$ of the elements of A with probability $\frac{(n-i+1)!}{n!}$.

Initialization

Here $i = 1$, so the claim is that the empty array $A[1..0]$ contains all 0-permutations with probability $\frac{n!}{n!} = 1$. This is true.

Maintenance

At stage i, the probability of selecting a particular value in $A[i..n]$ to swap into $A[i]$ is independent of the probability that $A[1..i-1]$ is any particular $i-$permutation $\langle x_1, x_2..x_{i-1} \rangle$.

Therefore the probability of producing a given $i-$permutation $\langle x_1, x_2..x_{i-1}, x_i \rangle$at the ith pass through the loop is the probability of entering the loop with the permutation $\langle x_1, x_2..x_{i-1} \rangle$, $\frac{(n-i+1)!}{n!}$, times the probability that the swap in line 3 places x_i in the ith position, $\frac{1}{n-i+1}$. This probability is $\frac{(n-i)!}{n!}$. When i is incremented, the loop invariant will remain correct.

Termination

At termination, $i = n + 1$, so the loop invariant implies that all $n-$permutations of the elements of A are produced with probability $\frac{1}{n!}$, as required.

Combinatorial Argument

Alternatively, demonstrate that all permutations are equally likely to be generated by noting that there are $n!$ possible sequences of results of the calls to RANDOM(i, n), $< r_1, r_2..r_n >$ because there are $n - i + 1$ possible outcomes for r_i. Each sequence occurs with probability $n!$ because call is independent and each value of r_i is equally likely.

Finally, each sequence produces a distinct permutation, because the permutations generated by sequences $< r_1, r_2..r_n >$ and $< s_1, s_2..s_n >$ will differ in the first position i in which $r_i \neq s_i$. Thus $n!$ distinct permutations are generated, each with probability $\frac{1}{n!}$, as required.

RANDOM(a,b)

If you have just a random number generator that returns values between 0 and some $max - int$, how would you generate a uniform pick from $\{a, a+1..b\}$?

Motivating Example

Suppose $max - int = 7$, so the call to $rand()$, say, returns a value in $\{0, 1, 2, 3, 4, 5, 6, 7\}$. You want a value in $\{17, 18, 19\}$, $\{a..b\}$ with $a = 17$, $b = 19$. A value in $\{0, 1, 2\} = \{0..b - a\}$ would do: just add a to the result.

We need to associate an the desired return values $0, 1$, and 2 with disjoint events in $\{0, 1, 2, 3, 4, 5, 6, 7\}$ having equal probability, hence equal numbers of elements, since $rand()$ returns each value in $\{0, 1, 2, 3, 4, 5, 6, 7\}$ with equal probability. In trying to produce 3 or $b - a + 1$ events of equal size, the most we can have in each event is $\lfloor \frac{8}{3} \rfloor = 2$. If we return $\left\lfloor \frac{rand()}{2} \right\rfloor$ if it is in $\{0, 1, 2\}$ and otherwise discard the result and repeat, we return 0 when $rand() \in \{0.1\}$, 1 when $rand() \in \{2.3\}$, and 2 when $rand() \in \{4.5\}$, and otherwise try again.

What does this suggest for $max - int$ and $b - a$?