Probabilistic Analysis II -
Probabilistic Analysis for Hashing

Expectation

X is a random variable.

The expectation of X is:

$$E(X) = \sum_I \text{Prob}(X = I)\ I.$$

- Linearity of expectation:

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

- Conditional expectation:

$$E(X) = E(X \mid A)\ \text{Prob}(A) + E(X \mid \text{Not } A)\ (1 - \text{Prob}(A)).$$

- **Formula for non-negative random variables ($X \in \{0, 1, 2, 3, \ldots\}$):**

$$E(X) = \sum_{i=1}^{\infty} \text{Prob}(X \geq i).$$

Formula for Non-negative Random Variables

Assume X is a random variable taking only values $\{0, 1, 2, 3, \ldots\}$.

By definition:

$$E(X) = \sum_{i=0}^{\infty} Pr(X = i) \times i.$$

Theorem. If X is a random variable taking only values $\{0, 1, 2, 3, \ldots\}$, then

$$E(X) = \sum_{i=1}^{\infty} Pr(X \geq i).$$

Formula for Non-negative Variables: Proof 1

Theorem. If $X \in \{0, 1, 2, 3, \ldots\}$, then $E(X) = \sum_{i=1}^{\infty} \text{Prob}(X \geq i)$.

Proof.

$$
\begin{aligned}
E(X) &= \sum_{i=0}^{\infty} \text{Prob}(X = i) \times i \\
&= Pr(X = 1) + 2Pr(X = 2) + 3Pr(X = 3) + 4Pr(X = 4) + \\
&= Pr(X = 1) + Pr(X = 2) + Pr(X = 3) + Pr(X = 4) + \ldots \\
&\qquad\qquad + Pr(X = 2) + Pr(X = 3) + Pr(X = 4) + \ldots \\
&\qquad\qquad\qquad\qquad + Pr(X = 3) + Pr(X = 4) + \ldots \\
&\qquad\qquad\qquad\qquad\qquad\qquad + Pr(X = 4) + \ldots
\end{aligned}
$$

$$
= \begin{array}{l} \text{Prob}(X \geq 1) + \\ \text{Prob}(X \geq 2) + \\ \text{Prob}(X \geq 3) + \quad \ldots \end{array} \qquad = \sum_{i=1}^{\infty} \text{Prob}(X \geq i).
$$

$\square$

Formula for Non-negative Variables: Proof 2

Theorem. If X is a random variable taking only values $\{0, 1, 2, 3, \ldots\}$, then

$$E(X) = \sum_{i=1}^{\infty} Pr(X \geq i).$$

Proof. Let X_i be a random variable where $X_i = \begin{cases} 1 & \text{if } X \geq i, \\ 0 & \text{if } X < i. \end{cases}$

$$X = \sum_{i=1}^{\infty} X_i.$$

$$E(X_i) = \text{Prob}(X \geq i) \times 1 + \text{Prob}(X < i) \times 0 = \text{Prob}(X \geq i).$$

$$\begin{aligned} E(X) &= E(\sum_{i=1}^{\infty} X_i) = \sum_{i=1}^{\infty} E(X_i) \qquad \text{by linearity of expectation} \\ &= \sum_{i=1}^{\infty} \text{Prob}(X \geq i). \end{aligned}$$

$\square$

Columbus Casino

```
procedure ColumbusCasino()
repeat
    c ← CoinFlip();
    if (c = heads) then
        Print "I win";
    else
        Print "I quit";
    end
until (c = tails);
```

Columbus Casino: Analysis

X = Number of heads.

Running time $= cX + c$.
(Last c term is the time for the last coin flip which is a tail.)

Expected running time $= E(cX + c) = cE(X) + c$.

Use formula $E(X) = \sum_{i=1}^{\infty} Pr(X \geq i)$.

$$E(X) = \sum_{i=1}^{\infty} Pr(X \geq i) = \sum_{i=1}^{\infty} (1/2)^i = 1.$$

Expected running time $= cE(X) + c = c + c = 2c$.

Columbus Casino 2

```
procedure ColumbusCasinoII()
repeat
    d ← RollDie();                    /* Pay 85 cents to roll die */
    if (d = 6) then
        Print "I quit";
    else
        Print "I win one dollar";
    end
until (d = 6);
```

Columbus Casino II: Analysis

X = Number of heads.

Running time $= cX + c$.
(Last c term is the time for the last roll where $d = 6$.)

Expected running time $= E(cX + c) = cE(X) + c$.

Use formula $E(X) = \sum_{i=1}^{\infty} Pr(X \geq i)$.

$$E(X) = \sum_{i=1}^{\infty} Pr(X \geq i) = \sum_{i=1}^{\infty} (5/6)^i = (5/6) + (5/6)^2 + (5/6)^3 + \ldots$$

$$= (5/6)(1 + (5/6) + (5/6)^2 + \ldots) = (5/6)\frac{1}{1 - (5/6)} = 5.$$

Expected running time $= cE(X) + c = 5c + c = 6c$.

Example

```
function Func10(A[ ], n)
for i ← 1 to n do
    (d1, d2) ← RollTwoDice();
    if ((d1 = 1) and (d2 = 1)) then
        return (A[i]);
    end
end
return (A[n]);
```

Columbus Casino 3

procedure `ColumbusCasinoIII`(m)

B is a bin of m red balls and $2m$ green balls;

repeat

  $\beta \leftarrow$ select (and remove) a random ball from B;

  if (β is red) **then**

    Print "I quit";

  else

    Print "I win one dollar";

  end

until (β is red);

Expectation: Review

X is a random variable.

The expectation of X is: $E(X) = \sum_I \text{Prob}(X = I)\, I$.

- Linearity of expectation:

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

$$E(\sum_{i=1}^{k} X_k) = \sum_{i=1}^{k} E(X_k)$$

- Conditional expectation:

$$E(X) = E(X \mid A)\ \text{Prob}(A) + E(X \mid \text{Not } A)\ (1 - \text{Prob}(A)).$$

- Formula for non-negative random variables ($X \in \{0, 1, 2, 3, \ldots\}$):

$$E(X) = \sum_{i=1}^{\infty} \text{Prob}(X \geq i).$$