Eulerian and Hamiltonian Graphs

An *Eulerian circuit* is a closed walk in which each edge appears exactly once. A graph is *Eulerian* if it contains an Eulerian circuit. A *Hamiltonian circuit* is a closed walk in which each vertex appears exactly once. A graph is *Hamiltonian* if it contains a Hamiltonian circuit.

To determine whether a graph is Hamiltonian or not is significantly harder than determining whether a graph is Eulerian or not. In this class we study the characterization of Eulerian graphs.

Example. If $\delta(G) \geq 2$ then G contains a cycle.

Solution. Let P be a longest path (actually, any *maximal* path suffices) in G and let u be an endpoint of P. Since P cannot be extended, every neighbor of u is a vertex in P. Since $deg(u) \geq 2$, u has a neighbor $v \in P$ via an edge that is not in P. The edge $\{u, v\}$ completes the cycle with the portion of P from v to u.

Example. Prove that a connected graph G is Eulerian iff every vertex in G has even degree.

Solution. *Necessity:* To prove that "if G is Eulerian then every vertex in G has even degree". Let C denote the Eulerian circuit in G. Each passage of C through a vertex uses two incident edges and the first edge is paired with the last at the first vertex. Hence every vertex has even degree.

Sufficiency: To prove that "if every vertex in G has even degree then G is Eulerian". We will prove this using induction on the number of edges, m.
<u>Base Case:</u> $m = 0$. In this case G has only one vertex and that itself forms a Eulerian circuit.
<u>Induction Hypothesis:</u> Assume that the property holds for any graph G with n vertices and j edges, for all j such that $n - 1 \leq j \leq k$.
<u>Induction Step:</u> We want to prove that the property holds when G has n vertices and $k + 1$ edges. Since G has at least one edge and because G is connected and every vertex of G has even degree, $\delta(G) \geq 2$. From the result of the previous problem, G contains a cycle, say C. Let G' be the graph obtained from G by removing the edges in $E(C)$. Since C has either 0 or 2 edges at every vertex of G, each vertex in G' also has even degree. However, G'

may not be connected. By induction hypothesis, each connected component of G' has an Eulerian circuit. We can now construct an Eulerian circuit of G as follows. Traverse C, but when a component of G' is entered for the first time, we detour along the Eulerian circuit of that component. The circuit ends at the vertex where we began the detour. When we complete the traversal of C, we have completed an Eulerian circuit of G.

Alternative Proof for the Sufficiency Condition: Let G be a graph with all degrees even and let

$$W = v_0 e_0 \ldots e_{l-1} v_l$$

be the longest walk in G using no edge more than once. Since W cannot be extended all edges incident on v_l are part of W. Since all vertices in G have even degree it must be that $v_l = v_0$. Thus W is a closed walk. If W is Eulerian then we are done. Otherwise, there must be an edge in $E[G] \setminus E[W]$ that is incident on some vertex in W. Let this edge be $e = \{u, v_i\}$. Then the walk

$$u e v_i e_i \ldots e_{l-1} v_l e_0 v_0 e_1 \ldots e_{i-1} v_i$$

is longer than W, a contradiction.

Probability

Example. Suppose we flip two fair coins. What is the probability that both tosses give heads given that one of the flips results in heads? What is the probability that both tosses give heads given that the first coin results in heads?

Solution. We consider the following events to answer the question.

> A: event that both flips give heads.
> B: event that one of the flips gives heads.
> C: event that the first coin flip gives heads.

Let's first calculate $\Pr[A|B]$.

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{\Pr[A]}{\Pr[B]} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Similarly we can calculate $\Pr[A|C]$ as follows.

$$\Pr[A|C] = \frac{\Pr[A \cap C]}{\Pr[C]} = \frac{\Pr[A]}{\Pr[C]} = \frac{1/4}{1/2} = \frac{1}{2}.$$

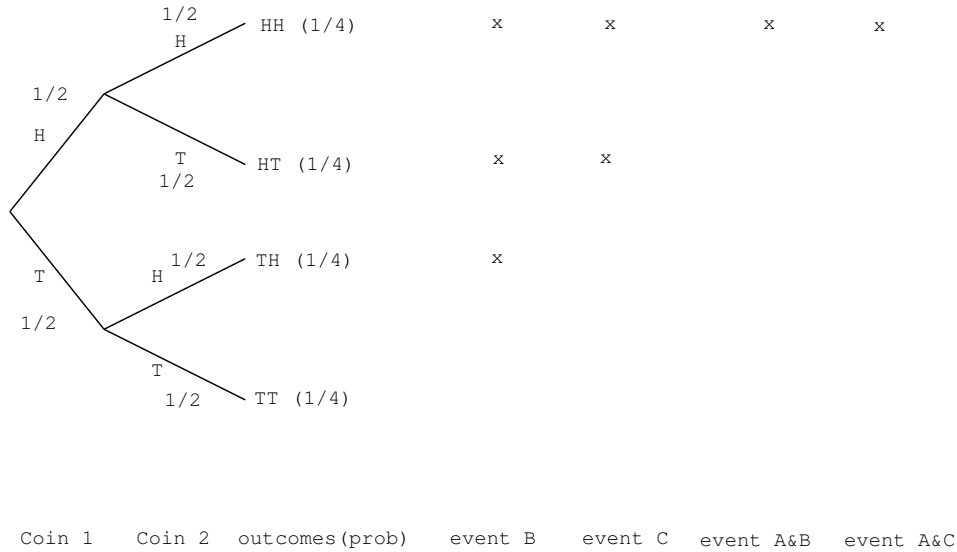The above analysis also follows from the tree diagram in Figure 1.

```
      1/2
            HH (1/4)        x        x           x         x
          H

  1/2
  H
          T
            HT (1/4)        x        x
        1/2


          1/2   TH (1/4)         x
        H
  T

  1/2

        T
      1/2    TT (1/4)


  Coin 1   Coin 2  outcomes(prob)   event B    event C   event A&B   event A&C
```

Figure 1: Tree diagram for the experiment in Example 1.

The Total Probability Theorem. Consider events E and F. Consider a sample point $\omega \in E$. Observe that ω belongs to either F or $\overline{F}$. Thus, the set E is a disjoint union of two sets: $E \cap F$ and $E \cap \overline{F}$. Hence we get

$$
\begin{aligned}
\Pr[E] &= \Pr[E \cap F] + \Pr[E \cap \overline{F}] \\
&= \Pr[F] \times \Pr[E|F] + \Pr[\overline{F}] \times \Pr[E|\overline{F}]
\end{aligned}
$$

In general, if $A_1, A_2, \ldots, A_n$ form a partition of the sample space and if $\forall i, \Pr[A_i] > 0$, then for any event B in the same probability space, we have

$$
\Pr[B] = \sum_{i=1}^{n} \Pr[A_i \cap B] = \sum_{i=1}^{n} \Pr[A_i] \times \Pr[B|A_i]
$$

Example. A medical test for a certain condition has arrived in the market. According to the case studies, when the test is performed on an affected person, the test comes up positive 95% of the times and yields a "false negative" 5% of the times. When the test is performed on a person not suffering from the medical condition the test comes up negative in 99% of the cases and yields a "false positive" in 1% of the cases. If 0.5% of the population actually have the condition, what is the probability that the person has the condition given that the test is positive?

    Solution. We will consider the following events to answer the question.

    C: event that the person tested has the medical condition.
    $\overline{C}$: event that the person tested does not have the condition.
    P: event that the person tested positive.

We are interested in $\Pr[C|P]$. From the definition of conditional probability and the total probability theorem we get

$$
\begin{aligned}
\Pr[C|P] &= \frac{\Pr[C \cap P]}{\Pr[P]} \\
&= \frac{\Pr[C]\Pr[P|C]}{\Pr[P \cap C] + \Pr[P \cap \overline{C}]} \\
&= \frac{\Pr[C]\Pr[P|C]}{\Pr[C]\Pr[P|C] + \Pr[\overline{C}]\Pr[P|\overline{C}]} \\
&= \frac{0.005 \times 0.95}{0.005 \times 0.95 + 0.995 \times 0.01} \\
&= 0.323
\end{aligned}
$$

This result means that 32.3% of the people who are tested positive actually suffer from the condition!

Linearity of Expectation

One of the most important properties of expectation that simplifies its computation is the *linearity of expectation.* By this property, the expectation of the sum of random variables equals the sum of their expectations. This is given formally in the following theorem.

Theorem. For any finite collection of random variables $X_1, X_2, \ldots, X_n$,

$$
\mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i]
$$

Example. Suppose that n people leave their hats at the hat check. If the hats are randomly returned what is the expected number of people that get their own hat back?

Solution. Let X be the random variable that denotes the number of people who get their own hat back. Let $X_i, 1 \leq i \leq n$, be the random variable that is 1 if the ith person gets his/her own hat back and 0 otherwise. Clearly,

$$
X = X_1 + X_2 + X_3 + \ldots + X_n
$$

By linearity of expectation we get

$$
\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} \frac{(n-1)!}{n!} = n \times \frac{1}{n} = 1
$$

Example. The following pseudo-code computes the minimum of n distinct numbers that are stored in an array A. What is the expected number of times that the variable *min* is assigned a value if the array A is a random permutation of the n elements.

FINDMIN(A, n)
1       $min \leftarrow A[1]$
2   **for** $i \leftarrow 2$ **to** n **do**
3   **if** $(A[i] < min)$ **then**
4               $min = A[i]$
5   **return** min

Solution.   Let X be the random variable denoting the number of times that min is assigned a value. We want to calculate $\mathbf{E}[X]$. Let X_i be the random variable that is 1 if min is assigned $A[i]$ and 0 otherwise. Clearly,

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

Using the linearity of expectation we get

$$\mathbf{E}[X] \;=\; \sum_{i=1}^{n} \mathbf{E}[X_i]$$
$$=\; \sum_{i=1}^{n} \Pr[X_i = 1] \tag{1}$$

Note that $\Pr[X_i = 1]$ is the probability that $A[i]$ contains the smallest element among the elements $A[1], A[2], \ldots, A[i]$. Since the smallest of these elements is equally likely to be in any of the first i locations, we have $\Pr[X_i = 1] = \frac{1}{i}$.

Probability Distributions

Tossing a coin is an experiment with exactly two outcomes: heads ("success") with a probability of, say p, and tails ("failure") with a probability of $1 - p$. Such an experiment is called a *Bernoulli trial*. Let Y be a random variable that is 1 if the experiment succeeds and is 0 otherwise. Y is called a *Bernoulli* or an *indicator* random variable. For such a variable we have
$$\mathbf{E}[Y] = p \cdot 1 + (1 - p) \cdot 0 = p = \Pr[Y = 1]$$

Thus for a fair coin if we consider heads as "success" then the expected value of the corresponding indicator random variable is $1/2$.

A sequence of Bernoulli trials means that the trials are independent and each has a probability p of success. We will study two important distributions that arise from Bernoulli trials: the *geometric distribution* and the *binomial distribution*.

The Geometric Distribution

Consider the following question. Suppose we have a biased coin with heads probability p that we flip repeatedly until it lands on heads. What is the distribution of the number of flips? This is an example of a *geometric distribution*. It arises in situations where we

perform a sequence of independent trials until the first success where each trial succeeds with a probability p.

Note that the sample space Ω consists of all sequences that end in H and have exactly one H. That is

$$\Omega = \{H, TH, TTH, TTTH, TTTTH, \ldots\}$$

For any $\omega \in \Omega$ of length i, $\Pr[\omega] = (1-p)^i p$.

Definition.   A *geometric random variable* X with parameter p is given by the following distribution for $i = 1, 2, \ldots$:

$$\Pr[X = i] = (1-p)^{i-1} p$$

We can verify that the geometric random variable admits a valid probability distribution as follows:

$$\sum_{i=1}^{\infty}(1-p)^{i-1}p = p\sum_{i=1}^{\infty}(1-p)^{i-1} = \frac{p}{1-p}\sum_{i=1}^{\infty}(1-p)^i = \frac{p}{1-p} \cdot \frac{1-p}{1-(1-p)} = 1$$

Note that to obtain the second-last term we have used the fact that $\sum_{i=1}^{\infty} c^i = \frac{c}{1-c}$, $|c| < 1$.

Let's now calculate the expectation of a geometric random variable, X. We can do this in several ways. One way is to use the definition of expectation.

$$
\begin{aligned}
\mathbf{E}[X] &= \sum_{i=0}^{\infty} i \Pr[X = i] \\
&= \sum_{i=0}^{\infty} i(1-p)^{i-1}p \\
&= \frac{p}{1-p}\sum_{i=0}^{\infty} i(1-p)^i \\
&= \left(\frac{p}{1-p}\right)\left(\frac{1-p}{(1-(1-p))^2}\right) \quad \left(\because \sum_{i=0}^{\infty} kx^k = \frac{x}{(1-x)^2}, \text{ for } |x| < 1.\right) \\
&= \left(\frac{p}{1-p}\right)\left(\frac{1-p}{p^2}\right) \\
&= \frac{1}{p}
\end{aligned}
$$

Another way to compute the expectation is to note that X is a random variable that takes on non-negative values. From a theorem proved in last class we know that if X takes on only non-negative values then

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} \Pr[X \geq i]$$

Using this result we can calculate the expectation of the geometric random variable X. For the geometric random variable X with parameter p,

$$\Pr[X \geq i] = \sum_{j=i}^{\infty}(1-p)^{j-1}p = (1-p)^{i-1}p\sum_{j=0}^{\infty}(1-p)^{j} = (1-p)^{i-1}p \times \frac{1}{1-(1-p)} = (1-p)^{i-1}$$

Therefore

$$\mathbf{E}[X] = \sum_{i=1}^{\infty}\Pr[X \geq i] = \sum_{i=1}^{\infty}(1-p)^{i-1} = \frac{1}{1-p}\sum_{i=1}^{\infty}(1-p)^{i} = \frac{1}{1-p}\cdot\frac{1-p}{1-(1-p)} = \frac{1}{p}$$

Binomial Distributions

Consider an experiment in which we perform a sequence of n coin flips in which the probability of obtaining heads is p. How many flips result in heads?

If X denotes the number of heads that appear then

$$\Pr[X = j] = \binom{n}{j}p^{j}(1-p)^{n-j}$$

Definition. A *binomial* random variable X with parameters n and p is defined by the following probability distribution on $j = 0, 1, 2, \ldots, n$:

$$\Pr[X = j] = \binom{n}{j}p^{j}(1-p)^{n-j}$$

We can verify that the above is a valid probability distribution using the binomial theorem as follows

$$\sum_{j=1}^{n}\binom{n}{j}p^{j}(1-p)^{n-j} = (p + (1-p))^{n} = 1$$

What is the expectation of a binomial random variable X? We can calculate $\mathbf{E}[X]$ is two

ways. We first calculate it directly from the definition.

$$
\begin{aligned}
\mathbf{E}[X] &= \sum_{j=0}^{n} j \binom{n}{j} p^j (1-p)^{n-j} \\
&= \sum_{j=0}^{n} j \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \\
&= \sum_{j=1}^{n} j \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \\
&= \sum_{j=1}^{n} \frac{n!}{(j-1)!(n-j)!} p^j (1-p)^{n-j} \\
&= np \sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!((n-1)-(j-1))!} p^{j-1} (1-p)^{(n-1)-(j-1)} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!} p^k (1-p)^{(n-1)-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} \\
&= np
\end{aligned}
$$

The last equation follows from the binomial expansion of $(p + (1-p))^{n-1}$.

We can obtain the result in a much simpler way by using the linearity of expectation. Let $X_i, 1 \le i \le n$ be the indicator random variable that is 1 if the ith flip results in heads and is 0 otherwise. We have $X = \sum_{i=1}^{n} X_i$. By the lineartiy of expectation we have

$$
\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} p = np
$$

What is the variance of the binomial random variable X? Since $X = \sum_{i=1}^{n} X_i$, and $X_1, X_2, \ldots, X_n$ are independent we have

$$
\begin{aligned}
\mathrm{Var}[X] &= \sum_{i=1}^{n} \mathrm{Var}[X_i] \\
&= \sum_{i=1}^{n} \mathbf{E}[X_i^2] - \mathbf{E}[X_i]^2 \\
&= \sum_{i=1}^{n} (p - p^2) \\
&= np(1-p)
\end{aligned}
$$

Coupon Collector's Problem.

We are trying to collect n different coupons that can be obtained by buying cereal boxes. The objective is to collect at least one coupon of each of the n types. Assume that each cereal box contains exactly one coupon and any of the n coupons is equally likely to occur. How many cereal boxes do we expect to buy to collect at least one coupon of each type?

Solution.   Let the random variable X denote the number of cereal boxes bought until we have at least one coupon of each type. We want to compute $\mathbf{E}[X]$. Let X_i be the random variable denoting the number of boxes bought to get the ith new coupon. Clearly,

$$X = X_1 + X_2 + X_3 + \ldots + X_n$$

Using the linearity of expectation we have

$$\mathbf{E}[X] = \mathbf{E}[X_1] + \mathbf{E}[X_2] + \mathbf{E}[X_3] + \ldots + \mathbf{E}[X_n] \tag{2}$$

What is the distribution of random variable X_i? Observe that the probability of obtaining the ith new coupon is given by

$$p_i = \frac{n - (i - 1)}{n} = \frac{n - i + 1}{n}$$

Thus the random variable $X_i, 1 \leq i \leq n$ is a geometric random variable with parameter p_i.

$$\mathbf{E}[X_i] = \frac{1}{p_i} = \frac{n}{n - i + 1}$$

Combining this with equation (2) we get

$$\mathbf{E}[X] = \frac{n}{n} + \frac{n}{n - 1} + \frac{n}{n - 1} + \cdots + \frac{n}{2} + \frac{n}{1} = n \sum_{i=1}^{n} \frac{1}{i}$$

The summation $\sum_{i-1}^{n} \frac{1}{i}$ is known as the *harmonic number* $H(n)$ and $H(n) = \ln n + c$, for some constant $c < 1$.

Hence the expected number of boxes needed to collect n coupons is about $nH(n) < n(\ln n + 1)$.