
6.2.1 The Union Bound and Extension

The **union bound** or **Boole's inequality** [13] is applicable when you need to show that the probability of union of some events is less than some value. Remember that for any two events A and B we have

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &\leq P(A) + P(B). \end{aligned}$$

Similarly, for three events A , B , and C , we can write

$$\begin{aligned} P(A \cup B \cup C) &= P((A \cup B) \cup C) \\ &\leq P(A \cup B) + P(C) \\ &\leq P(A) + P(B) + P(C). \end{aligned}$$

In general, using induction we prove the following

The Union Bound

For any events A_1, A_2, \dots, A_n , we have

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i). \quad (6.2)$$

The union bound is a very simple but useful result. It is used frequently in different applications. Here, we look at one application in the area of *random graphs*. Random graphs are widely used when analyzing social networks, wireless networks, and the internet. A simple model for random graphs is the Erdős-Rényi model $G(n, p)$ [11,12]. In this model, we have n nodes in the graph. In social networking context, each node might represent a person. Every pair of nodes are connected by an edge with probability p . The occurrence of each edge in the graph is independent from other edges in the graph. Figure 6.1 shows an example of a randomly generated graph using this model. Here, $n = 5$ and p was chosen to be $\frac{1}{2}$.

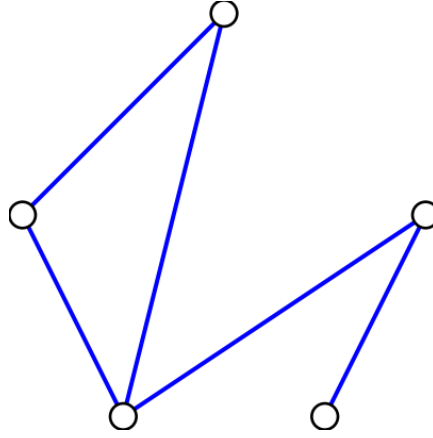


Fig.6.1 - An example of a randomly generated graph based on the $G(n, p)$ model. Here $n = 5$ and p was chosen to be $\frac{1}{2}$.

The question we are interested in here is the probability that there exists an isolated node in the graph [11,12]. An isolated node is a node that is not connected to any other nodes in the graph. In a wireless networking context, an isolated node is a node that cannot communicate with any other node in the network.

Example 6.17

Let B_n be the event that a graph randomly generated according to $G(n, p)$ model has at least one isolated node. Show that

$$P(B_n) \leq n(1 - p)^{n-1}.$$

And conclude that for any $\epsilon > 0$, if $p = p_n = (1 + \epsilon) \frac{\ln(n)}{n}$ then

$$\lim_{n \rightarrow \infty} P(B_n) = 0.$$

Solution

There are n nodes in the network. Let's call them Node 1, Node 2, ..., Node n . Let A_i be the event that the i th node is isolated. Then we have

$$B_n = \bigcup_{i=1}^n A_i.$$

Thus, using the union bound we conclude that

$$P(B_n) = P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i).$$

By symmetry, for all i, j , we have $P(A_i) = P(A_j)$, so

$$P(B_n) \leq nP(A_1).$$

Thus, we only need to find $P(A_1)$. The event A_1 occurs if Node 1 is not connected to any of the other $n - 1$ nodes. Since the connections are independent, we conclude that

$$P(A_1) = (1 - p)^{n-1}.$$

Therefore, we obtain

$$P(B_n) \leq n(1 - p)^{n-1},$$

which is the desired result. To prove the limit result, we use

$$\lim_{x \rightarrow \infty} \left(1 + \frac{c}{x}\right)^x = e^c, \text{ for any constant } c \in \mathbb{R}.$$

So, we obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} P(B_n) &\leq \lim_{n \rightarrow \infty} n(1 - p_n)^{n-1} \\ &= \lim_{n \rightarrow \infty} n \left[1 - (1 + \epsilon) \frac{\ln n}{n}\right]^{n-1} \\ &= \lim_{n \rightarrow \infty} n \left[1 - \frac{1+\epsilon}{\frac{n}{\ln n}}\right]^{n-1} \\ &= \lim_{n \rightarrow \infty} n \left(\left[1 - \frac{1+\epsilon}{\frac{n}{\ln n}}\right]^{\frac{n}{\ln n}}\right)^{\frac{(n-1) \ln n}{n}} \\ &= \lim_{n \rightarrow \infty} n e^{-(1+\epsilon) \ln n} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n^\epsilon} \\ &= 0. \end{aligned}$$

But since $P(B_n) \geq 0$, we conclude

$$\lim_{n \rightarrow \infty} P(B_n) = 0.$$

It is an interesting exercise to calculate $P(B_n)$ exactly using the inclusion-exclusion principle:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right).$$

In fact, the union bound states that the probability of union of some events is smaller than the first term in the inclusion-exclusion formula. We can in fact extend the union bound to obtain lower and upper bounds on the probability of union of events. These bounds are known as **Bonferroni inequalities** [13]. The idea is very simple. Start writing the inclusion-exclusion formula. If you stop at the first term, you obtain an upper bound on the probability of union. If you stop at the second term, you obtain a lower bound. If you stop at the third term, you obtain an upper bound, etc. So in general if you write an odd number of terms, you get an upper bound and if you write an even number of terms, you get a lower bound.

Generalization of the Union Bound: Bonferroni Inequalities

For any events A_1, A_2, \dots, A_n , we have

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &\leq \sum_{i=1}^n P(A_i); \\ P\left(\bigcup_{i=1}^n A_i\right) &\geq \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j); \\ P\left(\bigcup_{i=1}^n A_i\right) &\leq \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k). \\ &\vdots \end{aligned}$$

Example 6.18

Let B_n be the event that a graph randomly generated according to $G(n, p)$ model has at least one isolated node. Show that

$$P(B_n) \geq n(1-p)^{n-1} - \binom{n}{2}(1-p)^{2n-3}.$$

Solution

Similar to Example 6.17, let A_i be the event that the i th node is isolated. Then we have

$$B_n = \bigcup_{i=1}^n A_i.$$

Thus, using two terms in the inclusion-exclusion principle, we obtain

$$P(B_n) = P\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j).$$

By symmetry, we obtain

$$\begin{aligned} \sum_{i=1}^n P(A_i) &= nP(A_1), \\ \sum_{i < j} P(A_i \cap A_j) &= \binom{n}{2} P(A_1 \cap A_2). \end{aligned}$$

Thus, we conclude

$$P(B_n) \geq nP(A_1) - \binom{n}{2} P(A_1 \cap A_2).$$

In Example 6.17 we found

$$P(A_1) = (1 - p)^{n-1}.$$

Similarly, we obtain

$$P(A_1 \cap A_2) = (1 - p)^{2(n-2)+1} = (1 - p)^{2n-3}.$$

The reason for this is that $A_1 \cap A_2$ is the event that Nodes 1 and 2 are isolated. There are $2(n - 2)$ potential edges from the rest of the graph two Nodes 1 and 2, and there also is a potential edge from Node 1 to Node 2. These edges exist independently from each other and with probability p_n . We conclude

$$P(B_n) \geq n(1 - p)^{n-1} - \binom{n}{2} (1 - p)^{2n-3},$$

which is the desired result.

Expected Value of the Number of Events: It is interesting to note that the union bound formula is also equal to the expected value of the number of occurred events.

To see this, let A_1, A_2, \dots, A_n be any events. Define the indicator random variables X_1, X_2, \dots, X_n as

$$X_i = \begin{cases} 1 & \text{if } A_i \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

If we define $X = X_1 + X_2 + X_3 + \dots + X_n$, then X shows the number of events that actually occur. We then have

$$\begin{aligned} EX &= EX_1 + EX_2 + EX_3 + \dots + EX_n && \text{(by linearity of expectation)} \\ &= P(A_1) + P(A_2) + \dots + P(A_n), \end{aligned}$$

which is indeed the righthand-side of the union bound. For example, from this we can conclude that the expected number of isolated nodes in a graph randomly generated according to $G(n, p)$ is equal to

$$EX = n(1 - p)^{n-1}.$$