Additional Proofs from Week 2

Math 394

1 Inclusion-Exclusion Formula By Induction

1.1 The Induction Principle

The book mentions the possibility of proving the inclusion-exclusion formula by an induction argument. Recall how the induction principle works.

> We consider the sequence natural numbers $1, 2, 3, \ldots, n, \ldots$ and a proposition involving a natural number (such as the number of items involved in the proposition), $p(n)$. If we can show that

- $p(1)$ is true
- If we assume that $p(n)$ is true, then $p(n+1)$ is also true

> then the proposition holds for all n. As a simple example, the famed formula $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$ can be proved by induction (besides the constructive proof ascribed to Gauss):

- $\sum_{k=1}^{1} k = 1 = \frac{1 \cdot 2}{2}$
- Assume the proposition is true for n, then

$$\sum_{k=1}^{n+1} k = \sum_{k=1}^{n} k + n + 1 = \frac{n(n+1)}{2} + n + 1 = \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+2)(n+1)}{2}$$

The principle formalizes a natural "and so on" argument: if it is true for $n = 1$ it is true for $n = 2$, but then it is true for $n = 3$, and so on...

1.2 Proof

If $n = 1$, the inclusion-exclusion formula is trivial. Suppose it is true for n subsets. Then,

$$P\left[\bigcup_{k=1}^{n+1} E_k\right] = P\left[\left(\bigcup_{k=1}^{n} E_k\right)\bigcup E_{n+1}\right] =$$

$$= P\left[\bigcup_{k=1}^{n} E_k\right] + P\left[E_{n+1}\right] - P\left[\left(\bigcup_{k=1}^{n} E_k\right)\bigcap E_{n+1}\right] =$$

$$= \sum_{k=1}^{n+1} P\left[E_k\right] + (-1)^{n-1}\sum_{k=2}^{n}\sum_{i_1<1_2<\ldots<i_k} P\left[E_{i_1}\bigcap E_{i_2}\bigcap\cdots\bigcap E_{i_k}\right] - P\left[\bigcup_{k=1}^{n}\left(E_k\bigcap E_{n+1}\right)\right] \tag{1}$$

We have used the basic formulas of Boolean algebras, as well as the inclusion-exclusion formula for n sets. Now, we apply the same inclusion exclusion formula to the last probability, which involves n sets, and thus is valid by the inductive assumption. This results in

$$P\left[\bigcup_{k=1}^{n}\left(E_k\bigcap E_{n+1}\right)\right] =$$

$$= \sum_{k=1}^{n} P\left[E_k\bigcap E_{n+1}\right] + (-1)^{n-1}\sum_{k=2}^{n}\sum_{i_1<i_2<\ldots<i_k} P\left[E_{i_1}\bigcap E_{i_2}\bigcap\cdots\bigcap E_{i_k}\bigcap E_{n+1}\right]$$

Plugging this into (1) we see that subtracting the last expressions adds precisely the "missing terms" needed to express the inclusion-exclusion principle for $E_1, E_2, \ldots, E_{n+1}$.

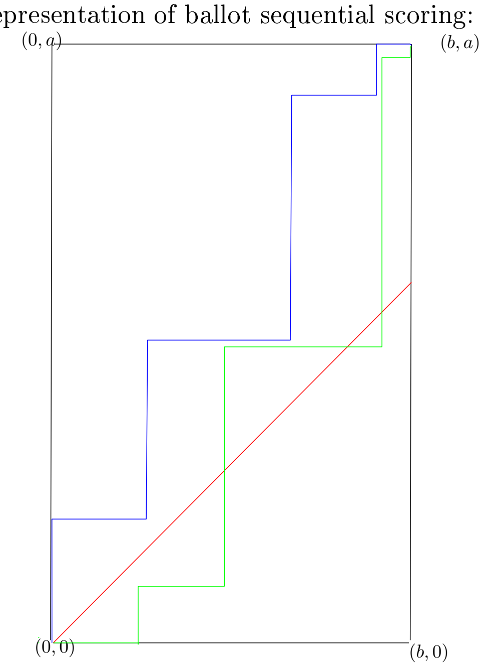
2 The Voting Problem

Suppose we have an election with two candidates, A and B. Suppose A earns a votes, and B b votes, and $a > b$. Suppose that ballots are counted in "random order". What is the probability that candidate A be ahead of B at all times during the count.

Here "random order" means that, all possible orderings of the $n = a + b$ ballots are equally likely. Note that a given sequence of ballots is a sequence of $a + b$ symbols that are either A or B, such that the total number of A's is a (and, consequently, the total number of B's has to be $n - a = b$). How many such sequences are there? In total, we have as many sequences as the ways in which we can pick a (or b) items out of n. As we know, that's

$$\binom{n}{a} = \binom{n}{b} = \binom{a+b}{a} = \binom{a+b}{b} \tag{2}$$

It is not as easy to spot the number of ways in which the ballots for A will always stay ahead of those for B. To this purpose, we try the following representation of ballot sequential scoring:



We represent ballots for A on a vertical axis, and for B on the horizontal. Starting at $(0, 0)$ (no ballots counted), we move one step horizontal or vertical, according to the name on the next ballot. Inevitably, we will end at the point with coordinates (b, a). In the picture, the red line indicates the points of coordinates (u, u), meaning that both candidates have the same number of ballots. The blue path is an example of a sequence of ballot recording where A is always ahead of B: it never touches the red line. The green path is an example of a sequence that does not comply with this requirement.

Since our basic outcomes (the atoms in our algebra, which we assume to have all the same probability) correspond to one of these paths, we have a representation of our sample space as the collection of "staircase" paths joining $(0,0)$ to (b,a)[1]. We need to count the paths that satisfy our condition – let's call the event E – but there does not seem to be a very obvious way to do this.

To get around the problem, we aim at the complement of E, the paths that do touch the red diagonal. Even this does not seem a trivial task, so we divide E^c in two subsets $E^c = F \bigcup G$, where F is the set of "bad paths" that start with a vote for A as first ballot, and G is the set of "bad" paths that start with a vote for B.

Looking at the paths in G, it is clear that their number is the number of paths that start at $(1,0)$ and end at (b,a). But their number is then easily seen to be given by a formula similar to (2), where now we are looking at $n-1$ ballots, split into a for A and $b-1$ for B (we "erased" the first ballot, which, for a path in G, was a vote for B). Hence, their number is given by

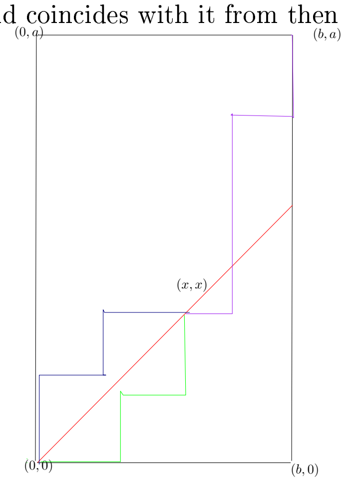
$$\binom{a+b-1}{a} \tag{3}$$

The next, somewhat surprising, step is the recognition that the number of paths in F is equal to the number of paths in G, and it is based on a technique that has wide-ranging scope in the applications of probability. The technique is called the "Reflection Principle"[2], and it works like this: we will realize that there is one path in F for every path in G, and vice-versa, creating a 1-1 correspondence between these two sets that, being finite sets, have to have the same number of elements.

[1] We will meet other examples where the sample space can be represented as a collection of paths. It turns out that this type of sample space is extremely helpful in dealing with all sorts of problems, in particular when we consider random experiments involving time – like "random motions" – which are a very significant area in probability.

[2] It can also be seen as a first simple example of another technique with even more wide-ranging scope, called "coupling" (coupling of two paths in a probability space whose elements are paths)

To this end, consider a path, say, in G, and follow it up the first time it hits the diagonal (it has to, not being in E). Consider the path in F that is the mirror image of this path when reflected around the diagonal up to this step, and coincides with it from then on:



In the picture, a path in F is drawn in dark blue up to its first hit of the diagonal, at position (x,x). The green path is its reflection, the first part of a path in G. Form (x,x), the two paths coincide. It is clear that this connection establishes a 1-1 mapping between F and G, which have consequently the same number of elements.

Wrapping up, the number of paths in A will be given by the total number of paths, (2), minus twice the number of paths in G (that is, the total number of paths in F and G), given by (3). Dividing by (2) gives us our probability:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\#\Omega - 2\#C}{\#\Omega} = 1 - 2\frac{\#C}{\#\Omega} = 1 - 2\frac{\binom{a+b-1}{a}}{\binom{a+b}{a}} =$$

$$1 - 2\frac{(a+b-1)!a!b!}{a!(b-1)!(a+b)!} = 1 - 2\frac{b}{a+b} = \frac{a+b-2b}{a+b} = \frac{a-b}{a+b}$$

that is the difference in the candidates respective percentages.