21

PIE:

The Principal of Inclusion and Exclusion

(This section optional)

The law of addition of Section 20 is a probabilistic interpretation of the principal of inclusion and exclusion. In general, the problem of counting things is to count each object once and only once. Suppose we have N objects, some of which have property a and some which have property b and some of which have both properties. We will denote by N_a the number of objects with property a (some of which may also have property b). N_b is defined similarly with N_{ab} being the number of objects with both properties a **and** b. V_{ab} denotes the number of objects with property a **or** b. Then it follows that $V_{ab} = N_a + N_b - N_{ab}$.

The rationale is precisely the same as for the law of addition. Since the N_{ab} objects are counted by both N_a and N_b we must subtract N_{ab} once.

By applying the formula $V_{ab} = N_a + N_b - N_{ab}$ recursively we can solve for V_{abc}. Doing this we get $V_{abc} = N_a + N_b + N_c - N_{ab} - N_{ac} - N_{bc} + N_{abc}$. Going further we can derive the general solution for $V_{abc...n}$:

$$V_{abc...n} = N_a + N_b + N_c + \ldots + N_n$$
$$- N_{ab} - N_{ac} - N_{an} - N_{mn} + N_{abc} + \ldots + N_{lmn} - N_{abcd} - \ldots$$

The Principal of Inclusion and Exclusion

The general pattern of the formula is that for each set of properties, s, we add or subtract N_s. If the number of properties in s is odd, we add N_s, otherwise we subtract N_s. If we denote by N_0 the number of objects with none of the properties, we have $\mathbf{N_0 = N - V_{abc...n}}$. A proof of the formula of inclusion and exclusion is given in the next paragraph for those of you who are hardy. The rest of you should feel free to skip to the following examples.

Consider an arbitrary object, call it *X*, out of our set of N objects. Suppose that this object has k properties (out of however many properties we have). To prove that the formula given by the principal of inclusion and exclusion is correct, we need only to show that X is counted once and only once by the formula. X is counted k (= $\binom{k}{1}$) times by terms of the form N_z. It is then counted $\binom{k}{2}$ times by terms of the form N_{yz}. In general it is counted $\binom{k}{j}$ times by terms with j subscripts. Hence the total number of times it is counted is $\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - ... \pm \binom{k}{k} = h$. However, 1 − h is the binomial expansion of $(1 - 1)^k = 0$. That is, 1 − h = 0, which means that X is counted h = 1 times.

<u>Example</u> Suppose we flip a coin 5 times. We would like to know how many ways can we do this such that there is a sequence of **exactly** 3 heads in a row (4 heads in a row doesn't count). Let N_i be the number of sequences of tosses with a sequence of exactly 3 heads beginning on the i'th throw. We want to calculate V_{12345}. However, this is equal to V_{123} since a sequence of 3 heads cannot start on the 4'th or 5'th throws. Applying the PIE formula we get $V_{123} = N_1 + N_2 + N_3 - N_{12} - N_{13} - N_{23} + N_{123}$. Only the first 3 terms can be non-zero since the other terms are self-contradictory. (Since for example we can't have sequences of heads **beginning** on the first **and** the second throws.) N_1 is the number of sequences beginning with exactly 3 heads. The fourth throw must be tails and this leaves two possibilities for the fifth throw. Hence, $N_1 = 2$. Similarly, if there is a sequence of exactly 3 heads beginning with the second toss then both the first and fifth tosses must be tails. Hence, $N_2 = 1$. Lastly if such a sequence of heads begins on the third throw, the second throw must be tails and the first throw can be either heads or tails. This gives $N_3 = 2$. Hence, $V_{12345} = V_{123} = 5$.

Example Consider the previous example except that we want to know the number of sequences with 3 heads in a row **or more**. Now we want N_i to be the number of sequences of tosses with a sequence of 3 or more heads beginning on the i'th throw. As before, $V_{12345} = V_{123}$ and $V_{123} = N_1 + N_2 + N_3 - N_{12} - N_{13} - N_{23} + N_{123}$. Also as before, only the first three terms can be non-zero. This time $N_1 = 4$ since the fourth and fifth tosses can each be heads or tails. $N_2 = 2$ since the fifth throw can be heads or tails. Similarly, $N_3 = 2$ because the first throw can be heads or tails (but the second throw must be tails). $V_{12345} = V_{123} = 8$.

Example Consider the same problems as in the previous two examples but for eight tosses of the coin. In the first case, we will count the number of sequences containing **exactly** three heads in a sequence. For this we get $V_{123456} = N_1 + N_2 + N_3 + N_4 + N_5 + N_6 - N_{15} - N_{16} - N_{26}$. These terms are 2^4, 2^3, 2^3, 2^3, 2^3, 2^4, 1, 1, and 1 respectively. The total is 61. In the second case, we will count the number of sequences containing three heads **or more** in a sequence. Again, we get $V_{123456} = N_1 + N_2 + N_3 + N_4 + N_5 + N_6 - N_{15} - N_{16} - N_{26}$. This time however the terms are 2^5, 2^4, 2^4, 2^4, 2^4, 2^4, 2, 2, and 1 respectively for a total of 107.

Great Example: The Problem of Derangements

This is a classic application of PIE. Be warned however, that the result depends on a (single) formula from calculus. Suppose that a hundred professors attend a conference. Suppose also that they are required to leave their cattle prods in the cloakroom. However, the cloakroom attendant does not have stubs for cattle prods, so he hands the prods back randomly. What is the probability that not one professor gets back his (or her) own prod? We let N represent the number of ways the prods can be handed back; that is N = 100!. If we can count the number of ways, N_0, that the prods can be handed back without any professor getting his own, then the probability of no one getting back his own is $\frac{N_0}{N} = \frac{N_0}{100!}$. Let N_{abc} be the number of ways that the prods can be handed back without persons a, b, or c getting back their own prods. Then, the number of ways, N_0, that no one gets back his own prod is $N - N_1 - N_2 - N_3 - \ldots - N_{100} + N_{1,2} + N_{1,3} + \ldots + N_{99,100} - \ldots\ldots$. This can be simplified by binomial coefficients to:

$$100! - \binom{100}{1}(100-1)! + \binom{100}{2}(100-2)! - \binom{100}{3}(100-3)! + \ldots + \binom{100}{100}(100-100)!$$

which in turn simplifies to:

$100! - 100! + \frac{100!}{2!} - \frac{100!}{3!} + \frac{100!}{4!} - \frac{100!}{5!} + \ldots + \frac{100!}{100!}$. Dividing this by 100! we get $\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} - \ldots + \frac{1}{100!}$. This is the first 99 terms of the Taylor series for $\frac{1}{e}$ (where e is the base of the natural logarithms or roughly 2.71828). In general if we have n professors, the probability that a professor gets his own prod back is $\frac{1}{n}$. The probability that all n professors get back their own prods is $\frac{1}{n!}$.

But for n over about 5, the probability that no one gets back his own prod is roughly $\frac{1}{e}$ and is therefore is independent of n.