Basic Combinatorics for Probability

Guy Lebanon

In this note we review basic combinatorics as it applies to probability theory (see [1] for more information). Although we are not concerned with probability in this note, we sometimes mention it under the assumption that all configurations are equally likely. Thus if the sample space Ω is of size n and every configuration or point is equally likely the probability of an event $A \subset \Omega$ is $p(A) = |A|/n$. Computing probabilities thus becomes a problem of counting $|A|$ and $|\Omega|$.

Proposition 1. *Given n_1 elements $a_1, \ldots, a_{n_1}$, n_2 elements $b_1, \ldots, b_{n_2}$ etc. up to n_r elements $x_1, \ldots, x_{n_r}$ it is possible to form $\prod_{j=1}^{r} n_j$ ordered r-tuplets containing one element of each kind. In particular there are n^r ways to form ordered r-tuplets over n elements. Equivalently, there are n^r ways to place r distinct balls into n cells (choose one cell for each ball).*

Proposition 2. *For a population of n elements and a prescribed sample size r there exist n^r different (ordered) samples with replacement and $(n)_r \stackrel{\text{def}}{=} n(n-1)\cdots(n-r+1)$ samples without replacement. As a corollary, the number of different orderings of n elements is $n! \stackrel{\text{def}}{=} n(n-1)(n-2)\cdots 1$.*

Example: the birthdays of r people form a sample of size r from the population of all days in the year. Assuming each 365 year days are equally likely the probability that all r birthdays are different equals

$$p = \frac{|A|}{|\Omega|} = \frac{(365)_r}{365^r} = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\cdots\left(1 - \frac{r-1}{365}\right).$$

For a small p we can approximate by neglecting all cross products $p \approx 1-(1+2+\cdots+(r-1))/365 = 1-r(r-1)/730$ (for example $r = 10$ yields the approximation $0.877 \approx 0.883 = p$ - note how close it is to 1). For large r we use the approximation $\log(1-x) \approx -x$ (for small x) to get $\log p \approx -(1+2+\cdots)/365 = -r(r-1)/730$ (for example for $r = 30$ we get $0.30307 \approx 0.294 = p$).

Proposition 3. *A population of n elements has $\binom{n}{r} \stackrel{\text{def}}{=} (n)_r/r! = n!/(r!(n-r)!)$ different sub-populations of size r. Equivalently there are $n!/(r!(n-r)!)$ ways to select r elements out of n distinct elements without importance to their orderings.*

Proof. There are $(n)_r$ ways to select r elements out of the n elements with ordering. Since there are $r!$ possible orderings of the selection we need to further divide $(n)_r$ by $r!$ (in order to ignore these orderings). □

Example: There are $|\Omega| = \binom{52}{5}$ different hands at poker. The probability that a hand has five different face values (assuming all hands are equally likely) is $4^5 \binom{13}{5} / \binom{52}{5} \approx 0.507$ (face values are chosen in $\binom{13}{5}$ ways and there are four suits possible for each of the five face values).

Example: The number of sequences of length $p+q$ containing p alphas and q betas is $\binom{p+q}{p}$ (choose p among $p+q$ sequence positions and assign them to alphas).

Proposition 4. *The number of ways to deposit n distinct objects into k bins with r_i objects in bin i (r_i are non-negative integers summing to n) is $n!/(r_1!\cdots r_k!)$ (ordering of bins is important but within each bin the ordering is not important).*

Proof. Repeated use of the previous proposition shows that the number may be written as

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\binom{n-r_1-r_2}{r_3}\cdots\binom{n-r_1-\cdots-r_{k-2}}{r_{k-1}}.$$

□

Example: A throw of twelve dice can result in 6^{12} different outcomes which we consider equally likely. The event that each face appears twice can occur in as many ways as twelve dice can be arranged in six groups of two each. The probability of that event is therefore $12!/(2^6 \cdot 6^{12}) \approx 0.0034$.

The statements thus far were mostly concerned with placing r distinct balls into n distinct cells (with or without orderings). In some cases it is desirable to treat the balls as indistinguishable.

Proposition 5. *The number of ways to place r indistinguishable objects into n distinct cells is $A_{r,n} \stackrel{\text{def}}{=} \binom{n+r-1}{r} = \binom{n+r-1}{n-1}$. The number of ways to do so if no cell remains empty is $\binom{r-1}{n-1}$.*

Proof. We represent a configuration as a sequence of indistinguishable objects and cell boundaries i.e., $|**|*||**|$ corresponds to 2 objects in cell 1, 1 in cell 2 and 2 in cell 4. In the first case, the number equals the number of ways to choose the r interior cell boundaries from the total number of symbols in the interior $r+n-1$ (once you select the cell boundaries the remaining positions becomes the indistinguishable objects). In the second case we have an added constraints that no two cell boundaries may be adjacent. The number then corresponds to selections of the positions of the $n-1$ interior cell boundaries out of $r-1$ potential positions in between two indistinguishable objects. □

Example: There are $\binom{n+r-1}{r}$ different partial derivatives of order r.

Consider a population of n elements with n_1 of them red and $n_2 = n - n_1$ black. The probability that a draw of r elements (without replacement) has k red elements is $q_k = \binom{n_1}{k}\binom{n-n_1}{r-k}/\binom{n}{r}$ (number of ways to get a configuration of k red from n_1 and $r-k$ non-red from $n-n_1$ divided by the total number of configurations). This probability is commonly known as the hypergeometric distribution and differs from the binomial distribution in that the sampling in this case is without replacement (the two become equivalent as $r \ll n$).

Example: In industrial quality control items are sampled from a collection of items and inspected for defects. The probability of sampling k defected items out of r items is given by the hypergeometric distribution. The total population size n as well as k and r are known but the number of defected items in the population n_1 is unknown. The latter may be estimated by maximizing the likelihood of the sample, and may be given confidence interval using standard statistical estimation[1].

Example: (capture-recapture) Consider the following experiment in an attempt to estimate the number of fish in a lake. First, n_1 fish are captures, marked, and released. At a later time, r fish are caught with k of them bearing the mark of the original capture. Assuming the size of the population of fish is n, the probability of getting k marked fish in the second capture follows the hypergeometric distribution with the corresponding parameters. In this case, n_1, r, k are known but n is unknown. We can estimate n or construct confidence intervals using the likelihood (probability of observed data as a function of the unknown parameter n)[1]. For example, if $k = 100$, $n_1 = r = 1000$ we have 93% confidence interval that $n \in (8500, 12000)$.

References

[1] W. Feller. *An Introduction to Probability Theory and its Application.* John Wiley and Sons, third edition, 1968.

[1] In this case we have but a single experiment. More accurate estimators may be constructed by repeated the process multiple times and maximizing the likelihood of n_1 over the set of experiments.