This article describes how the rule of laws in probability helps us survive the vagaries of 'chance as necessity'.

Large is Beautiful

All natural scientists and engineers have to live with random phenomena, be it because of noisy measurements, plain ignorance or fundamental physical principles. The silver lining is that there is a method to the madness, if one is willing to wait long enough – the celebrated 'laws' of probability theory ordain some order in the apparent chaos if we make sufficiently many experiments. Ideally, these should be infinitely many, because it is only in the limit of 'large numbers' that the structure emerges. This article aims to show how, at least in a discrete set-up, a simple counting argument reveals much of this structure. Specifically, the article shows in this discrete set-up how the idea of 'large deviations' arises and how a handle on the probabilities thereof allows one to recover some of the more familiar limit theorems of probability from a different perspective. While the aim is to do this at a fairly intuitive level, the later sections (particularly from section 4 on) will be more palatable for those with some exposure to modern, measure theoretic probability.

Playing Dice with the Universe

To begin at the beginning (It's a very good place to start!), recall your first encounter with probability. It is invariably with a mythical personality who patiently tosses a fair coin (or rolls a fair die) an amazingly large number of times and observes that the fraction of times a particular outcome is observed (head/tail or one/two/
The celebrated law of large numbers says that what's typical will happen almost certainly. What causes this behaviour is the celebrated law of large numbers, which says that what's typical will happen almost certainly.

To be precise, let $X_1, X_2, \ldots$ be the outcomes of independent rollings of a $d$-sided die, marked with values $\{a(1), a(2), \ldots, a(d)\}$ (say). Let $p(i)$ be the probability of the outcome $a(i)$. Given an observed string $X_1, \ldots, X_n$, the a priori probability of observing it is

$$p(X_1)p(X_2) \ldots p(X_n) = p(1)^{k_1}p(2)^{k_2} \ldots p(d)^{k_d},$$

where $k_i$ is the number of times the outcome $a(i)$ occurred in the observed string $X_1, \ldots, X_n$. We can also define 'empirical probabilities' after the $n$-th experiment by: $p_n(i) = k_i/n$, the fraction of times outcome $a(i)$ was observed in the first $n$ trials. (Think of histograms.) Then from above,

$$p(X_1)p(X_2) \ldots p(X_n) = p(1)^{n p_n(1)}p(2)^{n p_n(2)} \ldots p(d)^{n p_n(d)}$$

$$= e^{n \sum_{i=1}^{d} p_n(i) \ln p(i)} e^{-n D(p_n||p) - n S(p_n)},$$

where $S(p_n) = - \sum_i p_n(i) \ln p_n(i)$ is the familiar Shannon entropy of $p_n(\cdot)$ and

$$D(p_n||p) = \sum_i p_n(i) \ln \left( \frac{p_n(i)}{p(i)} \right)$$

is the so called Kullback-Leibler divergence between $p_n$ and $p$, a well-known measure of discrepancy between probability distributions used in statistics and information theory. It is not hard to show that $D(p_n||p) \geq 0$, and is $= 0$ if and only if $p_n = p$. 

.../six) is approximately half or one-sixth, as the case may be. ...