Week 4 Recap

Repeated identical independent experiments

We flip a biased coin (it comes Heads with probability p) repeatedly. We roll a fair die repeatedly, looking for one or six. We repeat a given random experiment many times, independently, and look for an event A which we call "a success," we assume that this event A has probability p to occur in any such given experiment. For any such setup, it is convenient to think of the probability space as $\Omega = \{0,1\}^{\{}1,2,\dots\}$ where 1 represent "success." This is a big probability space and we cannot compute the probability of all subsets of it. However, we can compute the probability of any set of the form $A_{y_1,\dots,y_k} = \{\omega = (x_i)_1^\infty : x_i = y_i, 1 \le i \le k\}$ where $y_1,\dots,y_k$ are given values in $\{0,1\}$. Namely, assume that m of the y_i are 1 and $k-m$ are 0. Then

$$P(A_{y_1,\dots,y_k}) = p^k(1-p)^{n-k}.$$

The set $\mathcal{F}$ of those subsets of Ω for which we can compute the probability is obtained by taking all the sets of the form \(A_{y_1,\dots,y_k\}\), and countable unions, countable intersection, complements, and repeating such operations any finite number of times. In computing the probability of $P(A_{y_1,\dots,y_k})$ we have used a simple finite probability space (because we only have to consider the first k trials of the repeated experiment) and we have used the independence hypothesis. This setup leads to a number of interesting questions:

- What is the probability that we see k success in the first n trials?
- What is the probability that the first success occurs during the k-th trials?
- Fix a positive integer r. What is the probability that the r-th success occurs at the k trials?

The first question leads to the binomial distribution, $\mathrm{Bin}(n,p)$ on \{0,\dots,n\} with parameter p. The second leads the geometric distribution, $\mathrm{Geom}(p)$, on $\{1,\dots,\}$ with parameter p. The third defines the Negative Binomial distribution, $\mathrm{NegBin}(r,p)$, on $\{r, r+1,\dots\}$ with parameters r and p. The negative binomial distribution with parameter $r=1$ is the same as the geometric distribution (same parameter p for both).

You should always be able to recover the formula for these distribution from the description of the problem they relate to.

The formula for the negative binomial is

$$P(r\text{-th success occurs at } k\text{-th trials}) = \binom{k-1}{r-}(1-p)^{r-1}p, \quad k = r, r+1, \dots$$

Can you verify that this is a probability distribution on $\{r, r+1, r+2, \dots\}$?

The multinomial distribution: The multinomial distribution with parameters $n, p_1, \ldots, p_k$, where n, k are a positive integers and the p_is are positive reals summing to 1, occurs when an experiments has k disjoint possible outcomes $A_1, \ldots, A_k$ (instead of success/failure) with respective probability $p_1, \ldots, p_k$. We repeat this experiment, independently, n times and ask what is the probability to see outcomes A_i exactly n_i, $1 \leq I \leq k$ during those n trials. Here, we assume that $n_1 + \cdots + n_k = n$.

The hypergeometric distribution: Imagine sampling individuals from a population of N individuals containing two types a, b of individuals, N_a individuals of type a, N_b individuals of type b. We sample n individuals from the population without replacement and ask: what is the probability that the sample contains exactly \k\) individual of type a? The answer is

$$P(k) = \frac{\binom{N_a}{k}\binom{N_b}{n-k}}{\binom{N}{n}}.$$

Note that the values of k receiving positive probability are $\max\{0, n - N_b\} \leq k \leq \min\{n, N_a\}$.