Class Notes 2-1-2019

- **Independence of Discrete Random Variables**:
  Discrete random variables defined on the same Ω are independent if and only if

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

  for all choices $x_1, \ldots, x_n$ of possible values of $X_1, \ldots, X_n$.

- **Three Fair Coin Flips:**
  Ω consists of all 8 3-tuples, like $(\mathtt{H}, \mathtt{T}, \mathtt{H})$, etc. Let $X_1 = 1$ if the i^{th} flip is heads, and $X_i = 0$ otherwise. We check just one example:

$$P(X_1 = 1, X_2 = 0, X_3 = 1) = P((\mathtt{H}, \mathtt{T}, \mathtt{H})) = \frac{1}{8} = \frac{1}{2}\frac{1}{2}\frac{1}{2} = P(X_1 = 1)P(X_2 = 0)P(X_3 = 1)$$

  where for example we have

$$P(X_1 = 1) = P(\{(\mathtt{H}, \mathtt{H}, \mathtt{H}), (\mathtt{H}, \mathtt{H}, \mathtt{T}), (\mathtt{H}, \mathtt{T}, \mathtt{H}), (\mathtt{H}, \mathtt{T}, \mathtt{T})\}) = \frac{4}{8} = \frac{1}{2}$$

- **Sampling with and without replacement:**
  Randomly select one by one k elements from $\{1, 2, \ldots, n\}$. X_i denotes the number on the i^{th} draw, $i = 1, \ldots, k$.

  - *with replacement:* Here $\Omega = \{1, \ldots, n\}^k$, all n^k k-tuples filled with any numbers from $\{1, 2, \ldots, n\}$. For any $x \in \{1, 2, \ldots, n\}$ we have

$$P(X_i = x) = \frac{n^{k-1}}{n^k} = \frac{1}{n}$$

    since $X_i = x$ for all n^{k-1} k-tuples that have x in position i and anything in the other $k - 1$ positions, anything being any number from $\{1, 2, \ldots, n\}$. Thus for any selections $x_1, \ldots, x_k \in \{1, 2, \ldots, n\}$ we have

$$P(X_1 = x_1, \ldots X_k = x_k) = \frac{1}{n^k} = \prod_{i=1}^{k} \frac{1}{n} = \prod_{i=1}^{k} P(X_i = x_i)$$

    i.e., $X_1, \ldots, X_k$ are (mutually) independent.

  - *without replacement:* Here Ω consists of all $(n)_k = n(n-1)\cdots(n-k+1)$ k-tuples filled with distinct elements from $\{1, 2, \ldots, n\}$. We must have $k \le n$.
    Again we have but by different counting, fixing x in position i

$$P(X_i = x) = \frac{(n-1)_{k-1}}{(n)_k} = \frac{1}{n}$$

    For any selections of distinct $x_1, \ldots, x_k \in \{1, 2, \ldots, n\}$ we have

$$P(X_1 = x_1, \ldots X_k = x_k) = \frac{1}{(n)_k} \neq \prod_{i=1}^{k} \frac{1}{n} = \prod_{i=1}^{k} P(X_i = x_i)$$

    Thus $X_1, \ldots, X_k$ are not independent, not even pairwise.

- **Independent Trials:**
  The simplest case is the repeated trials of an experiment with two possible outcomes, say success or failure with probability p and $1-p$ respectively. The trials are assumed to be independent and are repeated n times, like flipping a biased coin. As sample space we take

$$\Omega = \{\omega = (s_1, \ldots, s_n) : \text{each } s_i = 0 \text{ or } 1\}$$

  We assign probabilities as follows:

$$P(\omega) = p^k(1-p)^{n-k} \quad \text{where } k \text{ is the number of 1's in } \omega$$

  These probabilities add to 1 when summed over all $\omega \in \Omega$.

- **Bernoulli Distribution:**
  The Bernoulli random variable records the outcome 1 or 0 of a single trial. The random variable X has the Bernoulli distribution with success probability p, $0 \le p \le 1$ if it has possible values 0 or 1 with $P(X=1) = p$ and $P(X=0) = 1-p$. In shorthand we write $X \sim \text{Ber}(p)$. A sequence of n independent Bernoulli trials with success probability p gives rise to independent Bernoulli random variables $X_1, \ldots X_n$. For example, we then have

$$P(X_1 = 0, X_2 = 1, X_3 = X_4 = 0, X_5 = 1, X_6 = 0) = p^2(1-p)^4$$

- **Binomial Distribution:**
  Most often we are not so much interested in the pattern of zeros and ones as they might occur in a sequence of n independent Bernoulli trials but more so in the number of successes $S_n = X_1 + \ldots + X_n$ in such trials.
  There are $\binom{n}{k}$ n-tuple patterns with exactly k 1's and $(n-k)$ 0's (combinatorics), each pattern has probability $p^k(1-p)^{n-k}$ and they are all mutually exclusive. This gives us

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

  S_n is said to have a binomial distribution with parameters $n \ge 1$ and $p \in [0,1]$.
  In shorthand we write $S_n \sim \text{Bin}(n,p)$. These probabilities add to one via the binomial theorem (google Binomial Theorem for its history):

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$$

  If $n = 1$ then $\text{Bin}(n,p) = \text{Ber}(p)$.

- **Dice Rolls:**
  Roll a die 5 times and count the number S_5 of sixes. What is the chance of getting two or three sixes? $S_5 \sim \text{Bin}(5, 1/6)$ and

$$\begin{aligned} P(S_5 = 2 \cup S_5 = 3) = P(S_5 = 2) + P(S_5 = 3) &= \binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 + \binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 \\ &= 10\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2\left(\frac{5}{6}+\frac{1}{6}\right) = \frac{250}{1296} \approx .193 \end{aligned}$$

- **Geometric Distribution:**
  To properly introduce this distribution we need to consider a sample space consisting of all countably infinite sequences ω of zeros and ones. Again define the Bernoulli random variable $X_i(\omega) = 1$ if the i^{th} position of ω holds a 1, otherwise $X_i(\omega) = 0$. When $0 < p < 1$ (the interesting case) we will always have $P(\omega) = 0$ since the product of infinitely many $p's$ and $(1-p)'s$ is zero. However, the set of all ω's for which a specified finite segment of length n holds a specific pattern of zeros and ones has the usual probability $p^k(1-p)^{n-k}$, where k is the number of ones in the pattern. There is an uncountable infinite number of such ω's with such a specified finite segment. Thus all these uncountable infinite number of zero probability add up to something sensible, namely $p^k(1-p)^{n-k}$, which is just the probability of seeing such a pattern in such a segment of length n, regardless what happens in all the other trials outside that segment. Think of this as being analogous to the length of an interval made up of uncountable infinite many points with length zero. A fully rigorous treatment of such issues, the Kolmogorove extension theorem, can be found in a course on measure theory or probability theory at the next level after this.

  In the context of an infinite number of independent Bernoulli trials we can consider the number N of trials needed to see the first success. Clearly,

  $$P(N = k) = P(X_1 = 0, \ldots, X_{k-1} = 0, X_k = 1) = (1-p)^{k-1}p$$

  A random variable X has a *geometric distribution* with parameter $0 < p \leq 1$ if the possible values of X are $1, 2, 3, \ldots$ and $P(X = k) = (1-p)^{k-1}p$ for $k = 1, 2, 3, \ldots$. In shorthand we write $X \sim \text{Geo}(p)$.
  Clearly these probabilities add to one, using the geometric series.