Discrete Probability Cheat Sheet
Network Science MATH50007

Luca Cocconi

1. **Sample space:** the set of all possible values that a given random variable can take up. For example, if you are rolling a die and denoting X_i the random variable corresponding to the outcome of the roll, the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. In this case, Ω is a finite set of cardinality $|\Omega| = 6$.

2. **Discrete probability distribution:** the probability $P(X_i = x)$ that a random variable X_i with a countable sample space takes up a particular value x. A valid probability distribution needs to be normalised, i.e. $\sum_{x\in\Omega} P(X_i = x) = 1$. If all outcomes are equiprobable we have $P(X_i = x) = 1/|\Omega|$. For the example of a die roll, $P(X_i = 1) = P(X_i = 2) = P(X_i = 3) = P(X_i = 4) = P(X_i = 5) = P(X_i = 6) = 1/6$.

3. **Expectation:** the weighted average of the value taken up by a random variable. It is denoted $\langle X_i \rangle$ and given by $\langle X_i \rangle = \sum_{x\in\Omega} xP(X_i = x)$. If all outcomes are equiprobable this is just the arithmetic mean of the elements in the sample space. For the example of a die roll, $\langle X_i \rangle = (1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$.

   The variance of X_i is $Var(X_i) = \langle (X_i - \mu)^2 \rangle$ where $\mu = \langle X_i \rangle$, and this can be rearranged as, $Var(X_i) = \langle X_i^2 \rangle - \langle X_i \rangle^2$.

4. **Linearity of expectation:** The expectation of a sum in the sum of the expectations, so $\langle X_1 + X_2 + ... + X_N \rangle = \langle X_1 \rangle + \langle X_2 \rangle + ... + \langle X_N \rangle$. Additionally, constant can be pulled in and out of expectations, so $\langle aX_i \rangle = a\langle X_i \rangle$.

5. **Bernoulli trial:** A random variable with only two possible outcomes, for example $\Omega = \{\textit{success}, \textit{failure}\}$ or $\Omega = \{\textit{true}, \textit{false}\}$, for which the probability of each is independent of the number of attempts. A coin toss is a typical example, with $P(X_i = \textit{head}) = P(X_i = \textit{tails}) = 1/2$. In general the two outcomes are not equiprobable.

6. **Binomial distribution:** The probability that n out of N Bernoulli trials are successful. Assuming that the success probability for a single Bernoulli trial is $0 \leq p \leq 1$, the corresponding Binomial distribution reads

   $$P(X_i = n; N) = \binom{N}{n} p^n (1-p)^{N-n} \ .$$

   Here $p^n(1-p)^{N-n}$ is the probability of any one particular sequence of trials with n successes and $N-n$ failures. The binomial factor "N choose n" counts how many such sequences can be constructed. The expectation is $\langle X_i \rangle = pN$ and the variance $\langle X_i^2 \rangle - \langle X_i \rangle^2 = p(1-p)N$.

7. **Joint probability:** Given two random variables X_i and Y_i, the joint probability $P(X_i = x, Y_i = y)$ is the probability that X_i takes up value x *and* Y_i takes up value y. If X_i and Y_i are statistically independent, the joint probability is simply the product of the probabilities for each random variable $P(X_i = x, Y_i = y) = P(X_i = x)P(Y_i = y)$.

8. **Conditional probability:** Given two random variables X_i and Y_i, the conditional probability $P(X_i = x|Y_i = y)$ is the probability that X_i take up value x *given* that Y_i takes up value y. If X_i and Y_i are statistically independent, the conditional probability is independent of y, $P(X_i = x|Y_i = y) = P(X_i = x)$. If $Y_i = y$ implies $X_i = x$, then the conditional probability is one.

9. **Bayes' theorem:** the joint and conditional probabilities are related by the following identity

$$P(X_i = x, Y_i = y) = P(X_i = x|Y_i = y)P(Y_i = y) = P(Y_i = y|X_i = x)P(X_i = x) \ .$$

   This can be also written as

$$P(Y_i = y|X_i = x) = \frac{P(X_i = x|Y_i = y)P(Y_i = y)}{P(X_i = x)} \ .$$