1. A **probability space** is a measurable space $(S, \mathcal{M}, \mu)$ with $\mathcal{M} \subseteq \mathcal{P}(S)$ and $\mu : \mathcal{M} \to [0,1]$ with $\mu(S = 1$. In particular, we will write $\mathcal{M} = \mathcal{F}$, $\Omega = S$ and $\mu = \mathbb{P}$. The sets $E \in \mathcal{M}$ are called **events**. The σ−algebra $\mathcal{F}$ must satisfy the following axioms:
   (a) $\emptyset \in \mathcal{F}$
   (b) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$
   (c) If $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

   As well, $\mathbb{P}$ must satisfy
   (a) $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$.
   (b) $\mathbb{P}(\Omega) = 1$.
   (c) Let $A_1, A_2, \ldots$ be disjoint. Then $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

   As a result, we collect some general properties here:
   (a) $\mathbb{P}(A^c) + \mathbb{P}(A) = 1$
   (b) $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{1 \leq i_1 < \cdots < i_r \leq n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_r})$
   (c) We say $D_1, D_2, \ldots$ is a **decreasing** (resp. **increasing**) sequence of events if $D_1 \supset D_2 \cdots$ (resp. $D_1 \subset D_2 \cdots$) and define $\lim_{n\to\infty} D_n = \bigcap_{i=1}^{\infty} D_n$ (resp. $\lim_{n\to\infty} D_n = \bigcup_{i=1}^{\infty} D_n$). Then $\mathbb{P}(\lim_{n\to\infty} D_n) = \lim_{n\to\infty} \mathbb{P}(D_n)$.

2. Let $E, F \in \mathcal{F}$. Fix E with $\mathbb{P}(E) > 0$. Then define $\mathbb{P}(F|E) = \frac{\mathbb{P}(F\cap E)}{\mathbb{P}(E)}$, the **conditional probability** of F given E. In fact, the map $\mathbb{P}(\cdot|E) : \mathcal{F} \to \mathbb{R}$ is a probability measure in its own right. Conditional probability has some nice properties:
   (a) $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1|A_2 \cap \cdots A_n) \cdots \mathbb{P}(A_3|A_2 \cap A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_1)$.
   (b) Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$ be a partition of Ω. Then for any $B \in \mathcal{F}$,

   $$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

   and

   $$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}$$

3. Let $A_1, A_2, \ldots, A_n \in \mathcal{F}$. We say the events are **independent (as a collection)** if

   $$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$$

   We say the evens are **conditionally independent** with respect to B if they are independent with respect to the conditional probability measure:

   $$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}|B) = \mathbb{P}(A_{i_1}|B) \cdots \mathbb{P}(A_{i_k}|B)$$

4. A **random variable** is a measurable function $f : S \to \mathbb{R}$. We usually write random variables with capital letters, like X. We say X is **discrete** if its range is countable.

Let $\mathcal{B}$ denote the Borel measurable subset of $\mathbb{R}$. The **distribution** of a measurable function is a measure $\rho : \mathcal{B} \to \mathbb{R}$ given by $\rho(S) = \mathbb{P}(f^{-1}(S))$. This is indeed measurable: by definition, X is measurable if it pulls open sets in $\mathbb{R}$ to elements of $\mathcal{F}$.

For some reason, probabilists dislike writing sets down properly. We say $\mathbb{P}(X = i) = \mathbb{P}(X^{-1}(i))$, $\mathbb{P}(X \leq i) = \mathbb{P}(X^{-1}(-\infty, i]))$, and other similar notations.

The **expected value** of a random variable X is given by $\mathbb{E}(X) = \int_\Omega X \, d\mathbb{P}$. When X is discrete and $X(\Omega) = \{x_1, x_2, \ldots\}$, this becomes $\mathbb{E}(X) = \sum_{k=1}^{\infty} x_k \mathbb{P}(X = x_k)$. Note that the sum need not necessarily exist.

(a) If X is discrete and $g : \mathbb{R} \to \mathbb{R}$, then $\mathbb{E}(g(X)) = \sum_{k=1}^{\infty} g(x_k)\mathbb{P}(X = x_k)$.

(b) $\mathbb{E}(aX + Y) = a\,\mathbb{E}(X) + \mathbb{E}(Y)$.

We define the **variance** $\mathrm{Var}(X) = \inf_{a \in \Omega} \mathbb{E}[(X - a)^2]$. If $X \in L^2$, then this value is minimized for $a = \mathbb{E}(X)$ and $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$.

(a) $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

(b) $\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X)$

5. We say that a discrete random variable is **Poisson**, i.e. $X \sim \mathrm{Poi}(\lambda)$ if it has image $\{0, 1, 2, \ldots\}$ and density function

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Poisson random variables arise from Binomial distributions, taking the limit as $n \to \infty$ and $p \to$ so that $np \to \lambda$. They also arise as a **Poisson process** with intensity λ if there is a point process such that for any interval I with length h, the probability that there is exactly one point in I is λh (with perhaps some sublinear additional terms). If a point process satisfies the above conditions, I_t is an arbitrary interval with length t, and N_t counts the number of impacts with I_t, then $N_t \sim \mathrm{Poi}(\lambda t)$. A sum of Poisson random variables X_i with parameters $_i$ is Poisson with paramaters $\sum_{i=1}^{n} \lambda_i$.

6. Let $(\Omega, \mathcal{M}, \mathbb{P})$ be a measure space, and let $X : \Omega \to \mathbb{R}$ be a random variable. Then define P_X on $\mathbb{R}$ by $P_X(E) = \mathbb{P}(X^{-1}(E))$, which is a Borel measure on $\mathbb{R}$ called the **distribution** of X. Then the function

$$F(t) = P_X((-\infty, t])$$

is called the **(cumulative) distribution function** of X. A family of functions $\{X_\alpha\}_{\alpha \in A}$ is **identically distributed** if $P_{X_\alpha} = P_{X_\beta}$ for any $\alpha, \beta \in A$. In the special case $\Omega = \mathbb{R}$ and $\mathbb{P}$ is the standard Lebesgue measure on $\mathbb{R}$, $X : \mathbb{R} \to \mathbb{R}$ is a Lebesgue measureable function. There is a particularly nice class of random variables called **absolutely continuous** random variables, in which the measure P_X has a nice form. We say X is absolutely continuous if there exists some $f_X : \mathbb{R} \to \mathbb{R}$ so that for any $E \in \mathcal{B}$,

$$P_X(E) = \int_E f_X(x)\,\mathrm{d}x$$

In the absolutely continuous case, things are particularly nice. If g is any measurable function and X is a random variable,

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x$$

and the case $g(x) = x$ gives the standard way to compute the expected value of a random variable.

Suppose k is a strictly increasing function, X is an absolutely continuous random variable with density f_X. Let $Y = k(X)$; then we have

$$f_Y(x) = f_X(k^{-1}(x)) \frac{\mathrm{d}}{\mathrm{d}x} k^{-1}(x) = \frac{f_X(k^{-1}(x))}{k'(k^{-1}(x))}$$

Given random variables X and Y, we can compute

$$F_{X+Y}(a) = \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\,\mathrm{d}y$$

which is the **convolution** of the random variables X and Y.

7. We say that ϕ is a **standard normal distribution** on $\mathbb{R}$ if it has density

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

and we denote its CDF by $\Phi(x)$. In this standard form, we have $\mathbb{E}(X) = 0$ and $\mathrm{Var}(X) = 1$. More generally, we say $X \sim \mathcal{N}(\mu, \sigma^2)$ if $X = \sigma Z + \mu$ where $Z \sim \mathcal{N}(0,1)$. If $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, then

$$X + aY \sim \mathcal{N}(\mu_1 + a\mu_2, \sigma_1^2 + a^2\sigma_2^2)$$

8. If we have a collection of random variables $(X_1, \ldots, X_n)$ then we can consider them as a map $(X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$, and the corresponding measure is called the **joint distribution** of $X_1, \ldots, X_n$. Often, we take $\Omega = \mathbb{R}^n$ and $\mathbb{P}$ to be the standard Lebesgue measure on $\mathbb{R}^n$.

9.

10. We say that random variables X, Y are **independent** if

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

for any Borel sets A, B. Equivalently, we require

$$F_{X,Y}(a,b) = F_X(a) F_X(b)$$