2.1 Probability mass functions.

A probability mass function is a function *pX* : → [0, 1] such that

$$p_X(u) = P(X = u) \text{ for all } u \in \mathbb{R}. \tag{2.16}$$

Note: If the random variable is continuous, then *pX*(*u*) = 0 for all values of *u*.

Thus,

$$\sum_{u \in \mathbb{R}} p_X(u) = \begin{cases} 0 & \text{if } X \text{ is continuous} \\ 1 & \text{if } X \text{ is discrete} \\ \text{neither 0 nor 1} & \text{if } X \text{ is mixed} \end{cases} \tag{2.17}$$

where the sum is done over the entire set of real numbers. What follows are some important examples of discrete random variables and their probability mass functions.

Discrete Uniform Random Variable:

A random variable *X* is discrete uniform is there is a finite set of real numbers $\{x_1, \ldots, x_n\}$ such that

$$p_X(u) = \begin{cases} 1/n & \text{if } u \in \{x_1, \ldots, x_n\} \\ 0 & \text{else} \end{cases} \tag{2.18}$$

For example a uniform random variable that assigns probability 1/6 to the numbers {1, 2, 3, 4, 5, 6} and zero to all the other numbers could be used to model the behavior of fair dies.

Bernoulli Random Variable:

Perhaps the simplest random variable is the so called **Bernoulli** random variable, with parameter μ ! [0, 1]. The Bernoulli random variable has the following probability mass function

$$p_X(y) = \begin{cases} \mu & \text{if } y = 1 \\ 1-\mu & \text{if } y = 0 \\ 0 & \text{if } y \neq 1 \text{ and } y \neq 0 \end{cases} \tag{2.19}$$

For example, a Bernoulli random variable with parameter μ = 0.5 could be used to model the behavior of a random die. Note such variable would also be discrete uniform.

Binomial Random Variable: A random variable *X* is binomial with parameters μ ϵ [0, 1] and *n* ϵ if its probability mass function is as follows

$$p_X(y) = \begin{cases} \binom{n}{y}\mu^y(1-\mu)^{n-y} & \text{if } u \in \{0,1,2,3,\ldots\} \\ 0 & \text{else} \end{cases} \tag{2.20}$$

Binomial probability mass functions are used to model the probability of obtaining *y* "heads" out of tossing a coin *n* times. The parameter μ represents the probability of getting heads in a single toss. For example if we want to get the probability of getting 9 heads out of 10 tosses of a fair coin, we set *n* = 10, μ = 0.5 (since the coin is fair).

$$p_X(9) = \binom{10}{9}(0.5)^9(0.5)^{10-9} = \frac{10!}{9!(10-9)!}(0.5)^{10} = 0.00976 \tag{2.21}$$

Poisson Random Variable

A random variable *X* is Poisson with parameter λ > 0 if its probability mass function is as follows

$$p_X(u) = \begin{cases} \frac{\lambda^u e^{-\lambda}}{u!} & \text{if } u \geq 0 \\ 0 & \text{else} \end{cases} \tag{2.22}$$

Poisson random variables model the behavior of random phenomena that occur with uniform likelihood in space or in time. For example, suppose on average a neuron spikes 6 times per 100 millisecond. If the neuron is Poisson then the probability of observing 0 spikes in a 100 millisecond interval is as follows

$$p_X(0) = \frac{6^0 e^{-6}}{0!} = 0.00247875217 \qquad (2.23)$$

2.2 Probability density functions.

The probability density of a random variable *X* is a function f_X : $\longrightarrow$ such that for all real numbers *a* > *b* the probability that *X* takes a value between *a* and *b* equals the area of the function under the interval [*a*, *b*]. In other words

$$P(X \in [a, b]) = \int_a^b f_X(u)du \qquad (2.24)$$

Note if a random variable has a probability density function (pdf) then

$$P(X = u) = \int_u^u f_X(x)dx = 0 \text{ for all values of } u \qquad (2.25)$$

and thus the random variable is continuous.

Interpreting probability densities:

If we take an interval very small the area under the interval can be approximated as a rectangle, thus, for small Δx

$$P(X \in (x, x + \Delta x]) \approx f_X(x)\Delta x \qquad (2.26)$$

$$f_X(u) \approx \frac{P(X \in (x, x + \Delta x])}{\Delta x} \qquad (2.27)$$

Thus the probability density at a point can be seen as the amount of probability per unit length of a small interval about that point. It is a ratio between two different ways of measuring a small interval: The probability measure of the interval and the length (also called Lebesgue measure) of the interval. What follows are examples of important continuous random variables and their probability density functions.

Continuous Uniform Variables:

A random variable X is continuous uniform in the interval $[a, b]$, where a and b are real numbers such that $b > a$, if its pdf is as follows;

$$f_X(u) = \begin{cases} 1/(b-a) & \text{if } u \in [a, b] \\ 0 & \text{else} \end{cases} \qquad (2.28)$$

Note how a probability density function can take values larger than 1. For example,a uniform random variable in the interval [0, 0.1] takes value 10 inside that interval and 0 everywhere else.

Continuous Exponential Variables:

A random variable X is called exponential if it has the following pdf

$$f_X(u) = \begin{cases} 0 \text{ if } u < 0 \\ \lambda \exp(-\lambda x) & \text{if } u \geq 0 \end{cases} \qquad (2.29)$$

we can calculate the probability of the interval [1, 2] by integration

$$P(\{X \in [1, 2]\}) = \int_1^2 \lambda \exp(-\lambda x) dx = \Big[-\exp(-\lambda x) \Big]_1^2 \qquad (2.30)$$

if $\lambda = 1$ this probability equals $\exp(-1) - \exp(-2) = 0.2325$.

Gaussian Random Variables:

A random variable X is Gaussian, also known as **normal**, with parameters $\mu \in$ and $\sigma^2 > 0$ if its pdf is as follows

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \qquad (2.31)$$

where $\pi = 3.1415 \cdots$, μ is a parameter that controls the location of the center of the function and σ is a parameter than controls the spread of the function.. Hereafter whenever we want to say that a random variable X is normal with parameters μ and σ^2 we shall write it as $X \quad N(\mu, \sigma^2)$. If a Gaussian random variable X has zero mean and

standard deviation equal to one, we say that it is a **standard Gaussian random variable**, and represent it $X \quad N(0, 1)$.

The Gaussian pdf is very important because of its ubiquitousness in nature thus the name "Normal". The underlying reason why this distribution is so widespread in nature is explained by an important theorem known as **the central limit theorem**. We will not prove this theorem here but it basically says that observations which are the result of a sum of a large number of random and independent influences have a cumulative distribution function closely approximated by that of a Gaussian random variable. Note this theorem applies to many natural observations: Height, weight, voltage fluctuations, IQ... All these variables are the result of a multitude of effects which when added up make the observations distribute approximately Gaussian.

One important property of Gaussian random variables is that linear combinations of Gaussian random variables produce Gaussian random variables. For example, if X_1 and X_2 are random variables, then $Y = 2+4X_1+6X_2$ would also be a Gaussian random variable.