Last Updated: January 10, 2024

Chapter 2

The Random Variable

2.1 Introduction

The term "Random variable" is a bit of a misnomer. The random variable is actually a function.

$$X : \Omega \rightarrow \mathcal{R} \ s.t. \ X^{-1}((a, b]) \in F$$

Here, every interval inside the brackets corresponds to a valid event. So why do we need such a function? Why do we need the Random variable? We have dealt with many questions till now, none of which required any external assistance from such a function. We need random variables because, in real life, several patterns exist in the probabilities of certain events. Certain trends are very common. Take student performance in an exam, for example. If you ever see the scores, you will notice- very few students have extremely low scores, and very few have extremely high ones. Many are near the middle (or average) score. If you plotted the number of students getting certain scores v/s the scores themselves, you would be able to observe a distinctive pattern. This isn't the only place where patterns emerge. They pop up everywhere- even the number of people in the dining halls v/s the time- patterns exist! These patterns are so common that it is essential to model them mathematically so that useful information can be extracted from them and predictions can be made- after all, it's all about probability!

The random variable is a function. We are given an experiment specified by the space Ω, and from the axiomatic definitions, we know that the field of subsets of Ω are called events and probability is assigned to these events. Now, for every possible outcome of this experiment, we assign a number. Consider an example: Throwing a dice. Say we assign the number 10 if 1 comes up, 25 if 2 comes up, 35 if 3 comes up, 49 if 4 comes up, 50 if 5 comes up and 51 if 6 comes up. Thus, we have a random variable for our experiment. Let's call this random variable **X**. Thus, we have: $\{1,2,3,4,5,6\} \rightarrow \{10,25,35,49,50,51\}$. This can be anything we please. We could've chosen $\{1,2,3,4,5,6\}$ as the outputs of the function too. Or even $\{50,100,2345678, 2345678, 100000000, 10000000000\}$- this is our wish for now. Take the first case. We can ask the question, what is the probability that $\mathbf{X} < 20$. The answer would be $\frac{1}{6}$. Why? **X** can be less than 20 only if it is 10 (no other value in the range of the function is less than 20). Similarly, one can ask, what is the probability that **X** is more than 29? The answer would be $\frac{4}{6} = \frac{2}{3}$
When defining a random variable, the following 2 conditions are necessary:

1. The set $\{\mathbf{X} \leq x\}$ is an event for every value of x
2. The probabilities of events $x = \infty$ and $x = -\infty$ are both 0

2.2 Distribution and Density Functions

2.2.1 Distribution Functions

As defined above, the set $\{\mathbf{X} \leq x\}$ is an event for every value of x. Thus, a probability measure can be associated with it. As x varies, the probability associated with the events generated by various x will also vary (in most cases). But irrespective of exceptions, we can always say that $P\{\mathbf{X} \leq x\}$ is a function of x. This function is called the cumulative distribution function (CDF) of the random variable **X**.

The notation we shall use for this function is $F_{\mathbf{X}}(x)$. In many places, the subscript is dropped. It is assumed that if the constant is a lowercase 'x', the random variable we are dealing with is **X**.

Every CDF follows the following properties:

1. $F(\infty) = 1$
2. $F(-\infty) = 1$
3. If $x_1 < x_2$, then $F(x_1) \leq F(x_2)$
4. If $F(x_0) = 0$, then $F(x) \forall x \leq x_0 = 0$
5. $P\{\mathbf{X} > x\} = 1 - P\{\mathbf{X} \leq x\} = 1 - F(x)$
6. $F(x)$ is right continuous
7. $P\{x_1 < \mathbf{X} \leq x_2\} = F(x_2) - F(x_1)$

2.2.2 Density Functions

The derivative of the CDF is known as the probability density function(pdf). It is denoted by $f_{\mathbf{X}}(x)$ for the random variable X. Like $F(x)$, the subscript is dropped, and $f(x)$ is understood to be the distribution function corresponding to the random variable **X**. Thus, we have:

$$f(x) = \frac{d}{dx}F(x)$$

or,

$$F_{\mathbf{X}}(x) = \int_{-\infty}^{x} f_{\mathbf{X}}(u)du$$

Since $F(\infty) = 1$, we have $\int_{-\infty}^{\infty} f_{\mathbf{X}}(x)dx = 1$
The above discussion on density functions was for continuous random variables, i.e., those random variables which are defined on a domain with only a finite number of discontinuities. However, we can also have discrete random variables. For any discrete random variable, the density function is called the probability mass function (pmf). Furthermore:

$$f_{\mathbf{X}}(x) = \sum_i p_i \delta(x - x_i)$$

It should satisfy $\sum_i p_i = 1$, $p_i \leq 1 \ \forall i$

2.2.3 Some common probability density functions (discrete)

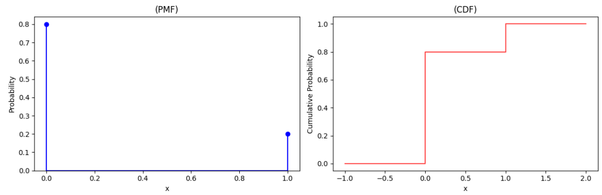
*Note: For all plots in this section, 'x' implies 'k'

The Bernoulli Distribution

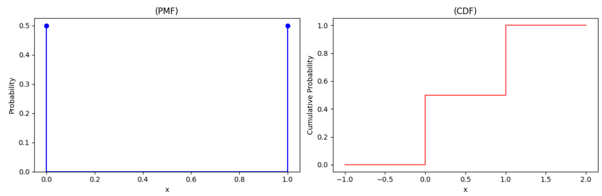
The Bernoulli random variable can only take two distinct values. They are taken to be 0 and 1, though other values may be used at times. An example would be the tossing of a coin. There are only two outcomes possible, heads or tails. One can assign 1 to heads and 0 to tails or vice-versa. **X** is a Bernoulli random variable if :

$$P\{\mathbf{X} = 1\} = p \ and \ P\{\mathbf{X} = 0\} = 1 - p$$

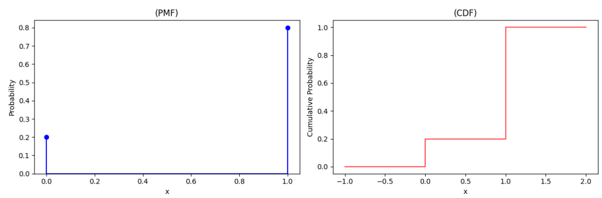
It is often modelled by $Ber(p)$ or $Ber(\theta)$ or $Ber \sim (p)$, where p is the probability of success (generally, this is taken to be 1. However, one is free to choose it as 0 as well). Any event which has only two possible outcomes can be modelled by the Bernoulli Distribution.



(a) p=0.2



(b) p=0.5



(c) p=0.8

Figure 2.1: The Bernoulli Distribution

The Binomial Distribution

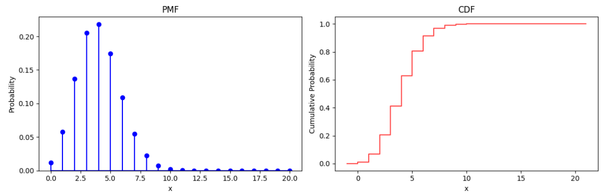
Say we repeat a Bernoulli trial n times, each trial independent from all others (but identical to each of them). If p represents the probability of success in each trial and **Y** represents the total number of favourable outcomes, then **Y** represents a binomial random variable. For example, if we toss a coin 20 times, and the probability of getting a head (which is "success" here/ this is the favorable outcome) is 0.2, then what is the probability of getting 1 head in 20 tosses? How about 2 heads in 20 tosses? What about 10? 15? The probability of getting a certain number of heads (say k heads) is the probability of a favorable outcome/ the probability of success. For every k, we have a probability. When we plot this probability for every value of k from 1 to n, we have the Binomial Distribution. It is often modelled by $Bin(n, p)$ or $Bin \sim (n, p)$, where p is the probability of success. Any experiment which has only two possible outcomes when repeated n times can be modelled as a Binomial Distribution. The special case when n=1 is nothing but the Bernoulli Distribution, since k can only take the values 0 and 1 in this case.

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

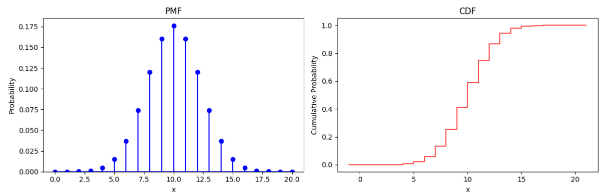
The Geometric Distribution

The geometric distribution can be looked at in two ways. Firstly, it can be modelled as the number of independent and identical Bernoulli trials needed to get the first "success". Secondly, it can be modelled as the number of failures before the first success. For example, say a coin is tossed a certain number of times. If we define getting a "tail" as "success" (and has a probability p), a geometric distribution would indicate the following: If n tosses are needed to get the first tail, then the first n-1 tosses resulted in a "head." Another example would be as follows: Say a set of clinical trials take place to find out whether a set of drugs work on a patient. Assume that the probability of a drug working effectively to cure a patient is p. How many drugs must be tested on the patient until he is finally cured? This can be modelled by a Geometric distribution. It is represented as $Geo(p)$ or $Geo \sim (p)$. Thus, we effectively have below the probability of k-1 failures in the first k-1 experiments, followed by a success on the k^{th} attempt.

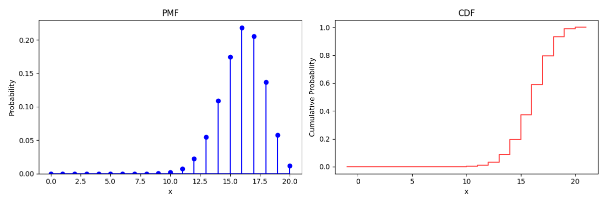
$$P\{\mathbf{X} = k\} = (1-p)^{k-1} p$$



(a) n=20, p=0.2

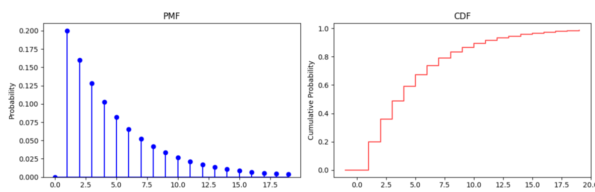


(b) n=20, p=0.5

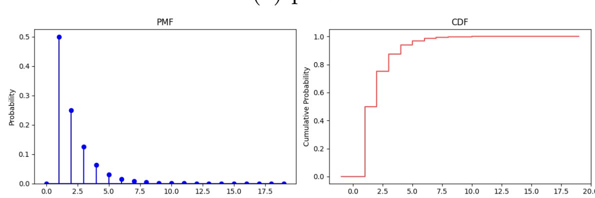


(c) n=20, p=0.8

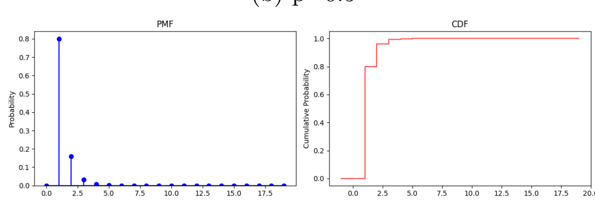
Figure 2.2: The Binomial Distribution



(a) p=0.2



(b) p=0.5



(c) p=0.8

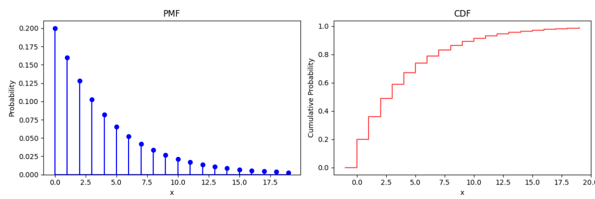
Figure 2.3: The Geometric Distribution

The Negative Binomial Distribution

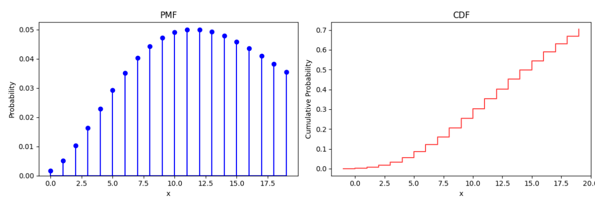
The Negative Binomial Distribution models the number of failures before a certain number of successes in a sequence of independent and identical Bernoulli trials. Say a coin is flipped again and again. The probability of getting a "heads" is p, which we define as success. What is the probability that on the sixth flip, "heads" has occurred for the fourth time? This can be modelled using the Negative Binomial Distribution. Note that by defining the number of successes, we inherently define the number of failures too. What we are effectively doing is as follows: from k+r-1 experiments, we choose r-1 successes and k failures, and on the final experiment ($(k+r)^{th}$), we must get a success, so multiply by p again.

$$P\{\mathbf{X}=k\} = \binom{k+r-1}{r-1} p^{r-1}(1-p)^k p = \binom{k+r-1}{r-1} p^r(1-p)^k = \binom{k+r-1}{k} p^r(1-p)^k$$

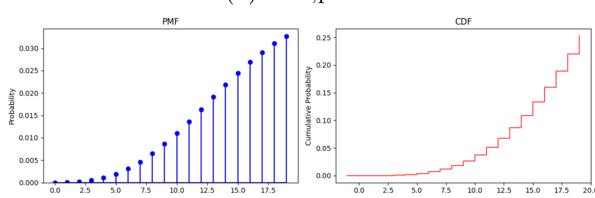
It is represented as $NB(k,r,p)$ or $NB \sim (k,r,p)$



(a) n=1,p=0.2



(b) n=4,p=0.2



(c) n=7,p=0.2

Figure 2.4: The Negative Binomial Distribution

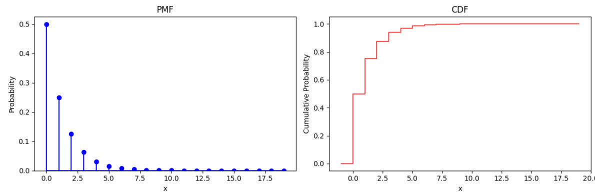
The Hypergeometric Distribution

The Hypergeometric Distribution is a subtle variant of the Binomial Distribution. In the binomial case, the possible set of outcomes is the same for every trial. Here, it isn't. It can be modelled as k successes in n trials without replacement, whereas the binomial distribution can be thought as of k successes in n trials with replacement. Consider the following example: Say there are 6 men and 19 women attending a party. A lucky draw is held. All their names are put in a hat and drawn. In total, 5 names are drawn. What is the probability that 3 men and 2 women are picked? Notice how the trials are not independent (unlike the binomial case). Once a name is picked, it can't be picked again. The equivalent binomial problem would be: the first name is picked and put back into the hat, then the second one is picked and put back into the hat and so on.

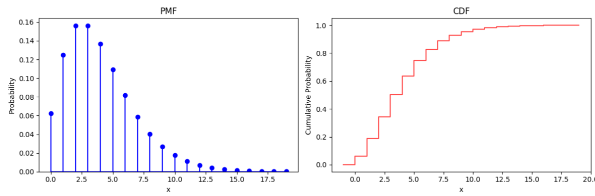
$$P\{\mathbf{X}=k\} = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

The Poisson Distribution

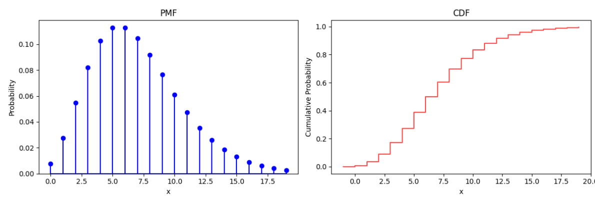
The Poisson Distribution measures the probability of a given number of events happening in a specified time period. To model the number of students entering the cafeteria between noon and, say, an hour after that, we would use the Poisson Distribution. Something like traffic flow can also be modelled using this distribution. The Poisson DIstribution is actually a limiting case of the



(a) n=1,p=0.5

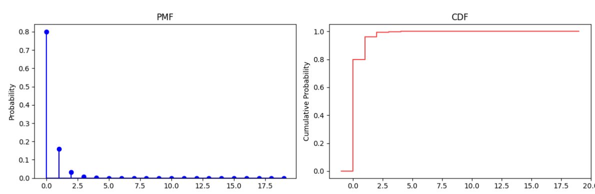


(b) n=4,p=0.5

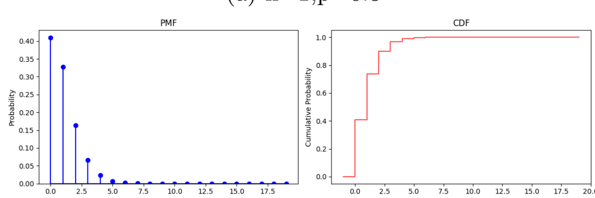


(c) n=7,p=0.5

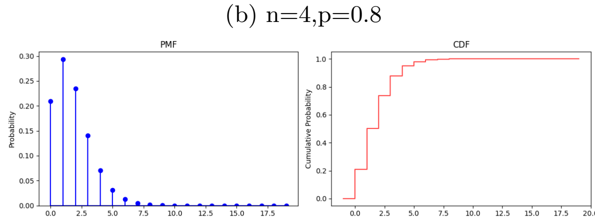
Figure 2.5: The Negative Binomial Distribution



(a) n=1,p=0.8



(b) n=4,p=0.8



(c) n=7,p=0.8

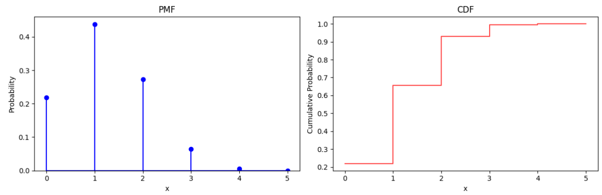
Figure 2.6: The Negative Binomial Distribution

Binomial Distribution when n becomes very large, and p becomes very small. For the Binomial Distribution, we have:

$$P\{\mathbf{X}=k\} = \binom{n}{k} p^k (1-p)^{n-k};\ set\ \lambda = np$$

$$\implies P\{\mathbf{X}=k\} = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$\implies P\{\mathbf{X}=k\} = \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$



(a) N=25,k=6,n=5

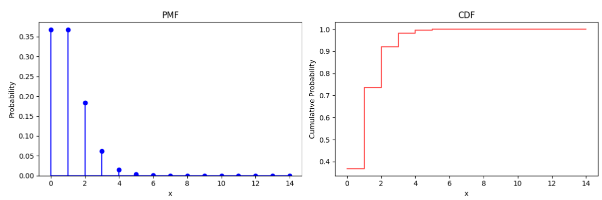
Figure 2.7: The Hypergeometric Distribution

$$\implies P\{\mathbf{X}=k\} = \frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n}\left(\frac{\lambda^k}{k!}\right)\left(1-\frac{\lambda}{n}\right)^{n-k}$$

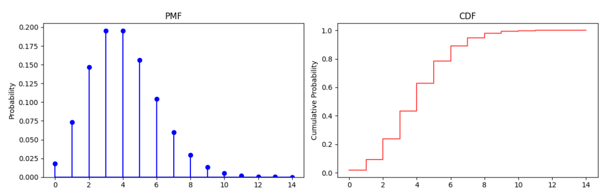
$$\implies P\{\mathbf{X}=k\} = \frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n}\left(\frac{\lambda^k}{k!}\right)\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-k}$$

Take the limit as n tends to ∞. This gives:

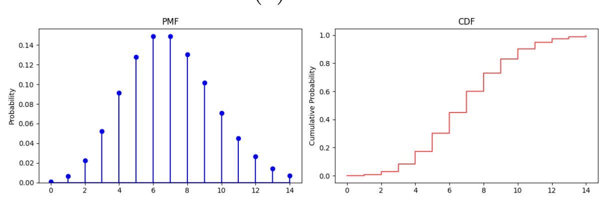
$$P\{\mathbf{X}=k\} = \frac{e^{-\lambda}\lambda^k}{k!}$$



(a) $\lambda = 1$



(b) $\lambda = 4$



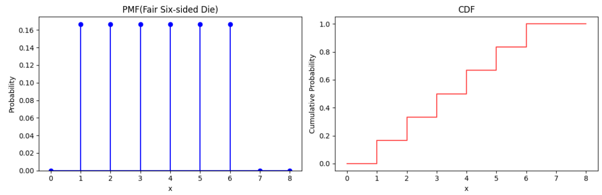
(c) $\lambda = 7$

Figure 2.8: The Poisson Distribution

The Uniform Distribution

Though generally strictly defined in the continuous realm, we can have a discrete formulation of the same. In the Uniform Distribution, every value of x in the domain is equally likely to occur. For example, tossing a fair coin or rolling a fair die can be modelled as a Uniform Distribution.

$$P\{\mathbf{X}=k\} = \frac{1}{n}$$



(a)

Figure 2.9: The Uniform Distribution

2.2.4 Some common probability density functions (continuous)

The Gaussian Distribution

Possibly the most common distribution one encounters is the Gaussian or the 'Normal' (since it is encountered so often) Distribution. It is a bell-shaped curve. It is modelled by two parameters, the mean and variance, both of which we will analyze in much greater detail in the upcoming chapter. The distribution is symmteric about this point. Furthermore, it has infinite support, i.e., x can take any value from 0 to ∞. You might see the term "alphabet" also being used for the set of values x can take, though this is slightly more common in the discrete distributions. μ and σ^2 denote the mean and the variance, respectively. The scores students obtain in an exam often follow a bell-shaped curve and are very close to being approximated as Gaussian. If $\mu = 0$ and $\sigma = 1$, we call it the Standard Normal Distribution. The pdf of the Normal Distribution is as follows:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The cdf is:

$$F_X(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right]$$

As you might have noticed, the function seems non-integrable, and in fact, there are three special functions associated with the Gaussian Distribution. The first of them is the Q function, the integral of the standard normal from x to ∞.

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}}\, dt$$

$Q(\infty) = 0$ and $Q(-\infty) = 1$. Why? On substituting $-\infty$ for x, we get the pdf for the Standard Normal Distribution. Since it is a valid pdf, its integral should be 1 over its support.

$$\implies \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\, dt = 1$$
$$\implies \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} = \sqrt{2\pi}$$

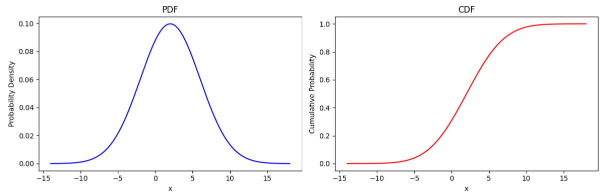
Q is a decreasing function of x. As is obvious by a simple manipulation of signs, $Q(x) = 1 - Q(-x)$. Here, $Q(-x)$ is defined as $\Phi(x)$, the CDF of the Standard Normal Distribution. $Q(x)$ can also be defined in terms of two other functions: The error function and the complementary error function: $erf(x)$ and $erfc(x)$.

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\frac{t^2}{2}}\, dt$$

$$Q(x) = \frac{1}{2} - \frac{1}{2} erf(\frac{x}{\sqrt{2}}) = \frac{1}{2} erfc(\frac{x}{\sqrt{2}})$$

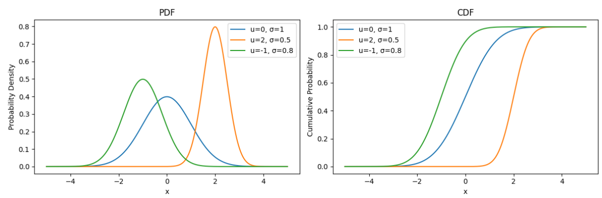
The Exponential Distribution

It is used to model the time elapsed between events. If occurrences of events over nonoverlapping intervals are independent, then the waiting time distribution of these events can be modelled by Exponential Distribution. Examples include bus arrival times at a



(a) $\mu = 2, \ \sigma = 4$

Figure 2.10: The Gaussian Distribution



(a)

Figure 2.11: Comparing Gaussian Distributions

bus stop. All such distributions exhibit a memorylessness property, which we will analyze in the conditional distributions section. It is defined for only $x > 0$. λ is known as the rate or the inverse scale parameter.

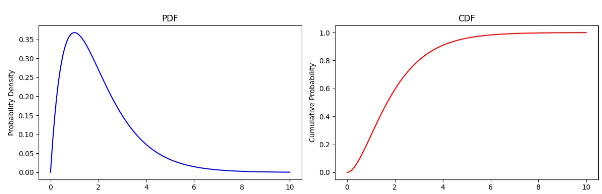
$$f_X(x) = \lambda e^{-\lambda x}$$

or

$$f_X(x) = \frac{1}{b} e^{-\frac{x}{b}}$$

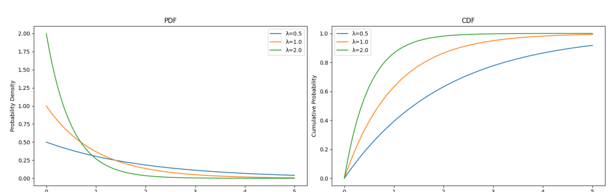
The cdf is:

$$F_X(x) = 1 - e^{-\lambda x}$$



(a) $\lambda = 0.5$

Figure 2.12: The Exponential Distribution



(a)

Figure 2.13: Comparing Exponential Distributions

The Gamma Distribution

The Gamma Distribution is often used to model the time between independent events that occur at a constant average rate. It is also used to model the distribution of rainfall and other natural phenomena. λ is known as the rate or inverse scale parameter. k is known as the shape parameter. If λ is set as $\frac{1}{\theta}$, then θ is known as the scale parameter.

$$f_X(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$$

The cdf is:

$$F_X(x) = \frac{1}{\Gamma(k)}\gamma(k, \lambda x)$$

The Gamma Function is defined as follows ($x > 0$):

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$$

It is used to define factorials and has the property: $\Gamma(n+1) = (n)\Gamma(n)$.

$$\Gamma(n) = \int_0^\infty t^{n-1} e^{-t}\, dt$$
$$\implies \Gamma(n+1) = \int_0^\infty t^{n} e^{-t}\, dt$$

Using integration by parts, we get:

$$\Gamma(n+1) = -t^n e^{-t}\Big|_0^\infty + \int_0^\infty n t^{n-1} e^{-t}\, dt$$
$$\implies \Gamma(n+1) = (0-0) + n\Gamma(n)$$
$$\implies \Gamma(n+1) = n\Gamma(n)$$

Furthermore, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ as:

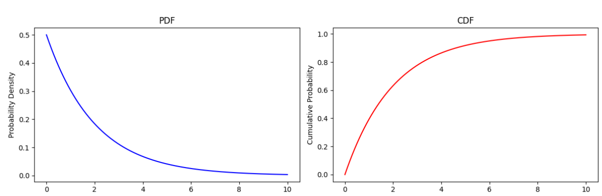
$$\Gamma(n) = \int_0^\infty t^{n-1} e^{-t}\, dt$$
$$\implies \Gamma\left(\frac{1}{2}\right) = \int_0^\infty t^{\frac{-1}{2}} e^{-t}\, dt$$
$$= \int_0^\infty \frac{e^{-t}}{\sqrt{t}}\, dt$$

Substitute $x = \sqrt{t}$. Thus, $dx = \frac{dt}{2\sqrt{t}}$.

$$\implies \Gamma\left(\frac{1}{2}\right) = 2\int_0^\infty e^{-x^2}\, dx$$
$$= \int_{-\infty}^\infty e^{-x^2}\, dx = \sqrt{\pi}$$

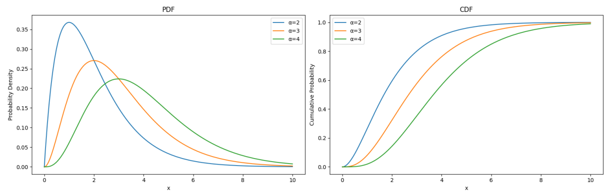
Finally, $\Gamma(1) = 1$ and $\Gamma(n)\cdot\Gamma(1-n) = \frac{\pi}{sin(n\pi)}$

Several other distributions are derived from the Gamma Distribution. On setting $\lambda = 1$, we get the Exponential Distribution,



(a) $\lambda = 2,\ k = 1$

Figure 2.14: The Gamma Distribution



(a)

Figure 2.15: Comparing Gamma Distributions

which has been discussed above. On setting $k = \frac{n}{2}$ and $\lambda = \frac{1}{2}$, we obtain the Chi-square distribution (with n degrees of freedom). We will later provide a relation between the Chi-square Distribution and the Gaussian Distribution. It has several applications in communication theory and hypothesis testing, among many others. Another distribution that arises is when we set $k = 1$. This gives the Erlang Distribution. It was developed to model the time between incoming telephone calls at a call center along with the expected number of calls.

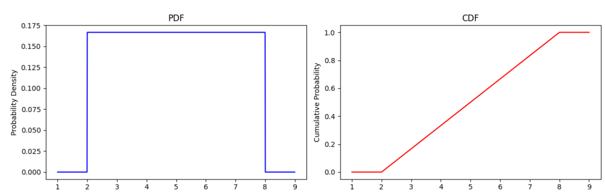
The Uniform Distribution

It is used to model scenarios where all outcomes are equally likely. The window under consideration is $[a, b]$. It need not be a continuous window. We could have a union, such as $\cup_{i=1}^{n}[a_i, b_i]$ such that each window of operation is disjoint from the rest, i.e., $[a_i, b_i] \cap [a_j, b_j] = \phi \ \forall i \neq j$

$$f_X(x) = \frac{1}{b-a}$$

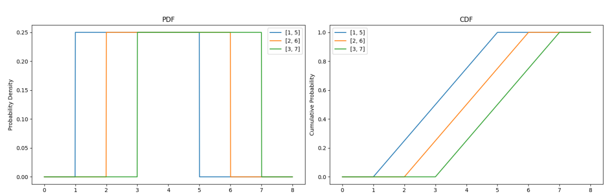
The cdf is:

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x > b \end{cases}$$



(a) $a = 2, \ b = 6$

Figure 2.16: The Uniform Distribution



(a)

Figure 2.17: Comparing Uniform Distributions

The Beta Distribution

It is defined in the range (0,1) (maybe including or excluding the boundary points). α and β are both shape parameters. Interestingly, as α and β tend to 0, the Beta Distribution approaches the Bernoulli Distribution. It is often used to model user interactions such as advertisement click-through rates and so on.

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$$

The cdf is:

$$F_X(x) = \int_0^x \frac{t^{\alpha-1}(1-t)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}\,dt$$

The Beta Function is defined as follows:

$$B(m,n) = \int_0^1 t^{m-1}(1-t)^{n-1}\,dt$$

We have the relation, $B(m,n) = \frac{\Gamma(m)\cdot\Gamma(n)}{\Gamma(m+n)}$. This can be proved as follows: The gamma function is

$$\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}\,dx$$

Set $x = ty$

$$\begin{aligned}
\implies \Gamma(n) &= \int_0^\infty (ty)^{n-1}e^{-ty}\,tdy \\
&= t^n \int_0^\infty (y)^{n-1}e^{-ty}\,tdy \\
&= t^n \int_0^\infty (x)^{n-1}e^{-tx}\,dx
\end{aligned}$$

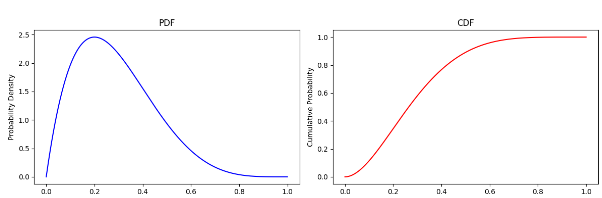
Multiply both sides by $e^{-t}t^{m-1}$.

$$\implies \Gamma(n)e^{-t}t^{m-1} = e^{-t}t^{m-1}t^n \int_0^\infty (x)^{n-1}e^{-tx}\,dx$$

Integrating both sides from 0 to ∞:

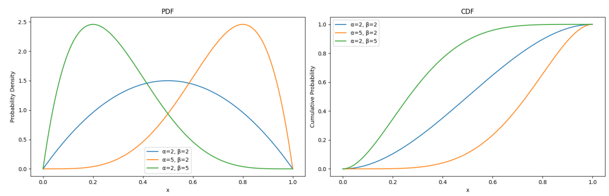
$$\begin{aligned}
\int_0^\infty \Gamma(n)e^{-t}t^{m-1} &= \int_0^\infty e^{-t}t^{m-1}t^n \int_0^\infty (x)^{n-1}e^{-tx}\,dx \\
\implies \Gamma(n)\Gamma(m) &= \int_0^\infty \int_0^\infty e^{-t(1+x)}t^{(m+n-1)}(x)^{n-1}\,dxdt \\
\implies \Gamma(n)\Gamma(m) &= \int_0^\infty (x)^{n-1} \int_0^\infty e^{-t(1+x)}t^{(m+n-1)}\,dxdt \\
\implies \Gamma(n)\Gamma(m) &= \int_0^\infty (x)^{n-1}\frac{\Gamma(n+m)}{(1+x)^{(m+n)}}\,dx \\
\implies \frac{\Gamma(n)\Gamma(m)}{\Gamma(n+m)} &= B(m,n)
\end{aligned}$$

The last step is the standard property $\int_0^1 x^{m-1}(1-x)^{n-1} = \int_0^\infty \frac{(x)^{n-1}}{(1+x)^{m+n}}$. This can be proved by using a simple substitution $x = \frac{1}{1+u}$



(a) $\alpha = 2,\ \beta = 5$

Figure 2.18: The Beta Distribution



(a)

Figure 2.19: Comparing Beta Distributions

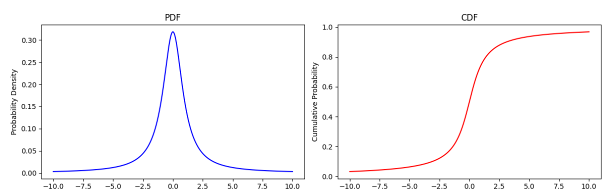
The Cauchy Distribution

It is also known as the Lorentzian Distribution. We will analyze its relation with the Normal Distribution in later chapters. It can model resonance behavior, and is also used in robustness analysis. Its pdf is as follows:

$$f_X(x) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

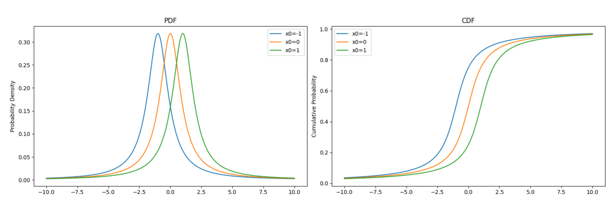
The cdf is:

$$F_X(x) = \frac{1}{\pi}\arctan\left(\frac{x-x_0}{\gamma}\right) + \frac{1}{2}$$



(a) $x_0 = 0,\ \gamma = 1$

Figure 2.20: The Cauchy Distribution



(a)

Figure 2.21: Comparing Cauchy Distributions

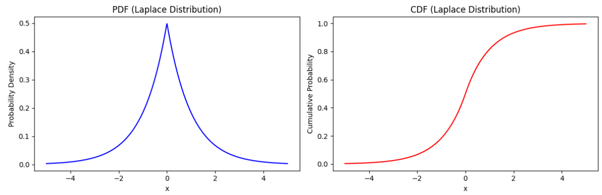
The Laplace Distribution

The Laplace Distribution is closely related to the Exponential Distribution. It is known as the Double Exponential Distribution. It is used in image and speech recognition, hydrology, and finance, Its pdf is given below:

$$f_X(x) = \frac{1}{2b}e^{-\frac{|x-\mu|}{b}}$$

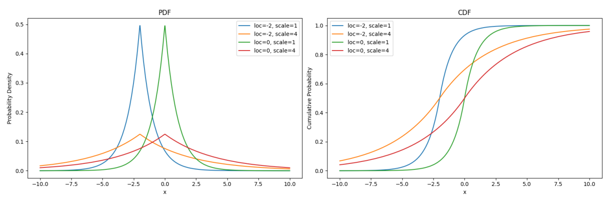
The cdf is:

$$F_X(x) = \begin{cases} \frac{1}{2}e^{\frac{x-\mu}{b}} & \text{if } x < \mu \\ 1 - \frac{1}{2}e^{-\frac{x-\mu}{b}} & \text{if } x \geq \mu \end{cases}$$



(a) $\mu = 0,\ b = 1$

Figure 2.22: The Laplace Distribution



(a)

Figure 2.23: Comparing Laplace Distributions

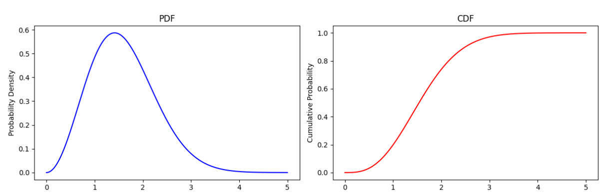
The Maxwell Distribution

A direct consequence of the KTG (Kinetic Theory of Gases), the Maxwell/ Maxwell-Boltzmann Distribution has several use cases in ensemble analysis. Its pdf is:

$$f_X(x) = \sqrt{\frac{2}{\pi}} \frac{x^2 e^{-\frac{x^2}{2a^2}}}{a^3}$$

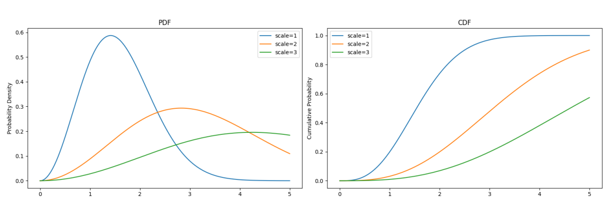
The cdf is:

$$F_X(x) = 1 - e^{-x^2/2}$$



(a) $a = 1$

Figure 2.24: The Maxwell Distribution



(a)

Figure 2.25: Comparing Maxwell Distributions

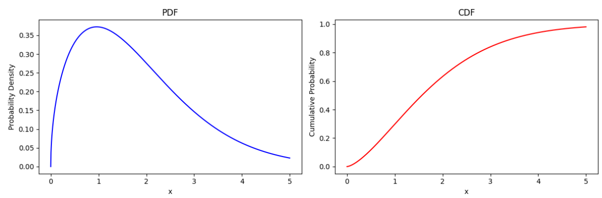
The Weibull Distribution

The Weibull Distribution is used to model the time between events and failure rates. k is a shape parameter, and λ is the scale parameter, both of which must be greater than 0. It is defined as:

$$f_X(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$$

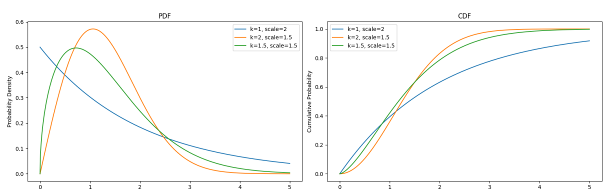
The cdf is:

$$F_X(x) = 1 - e^{-(x/\lambda)^k}$$



(a) $\lambda = 2, k = 1.5$

Figure 2.26: The Weibull Distribution



(a)

Figure 2.27: Comparing Weibull Distributions

A special case of the Weibull Distribution is the Rayleigh Distribution. It is used in Communication Theory, especially in non-coherent detection techniques. Its pdf is:

$$f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$$

The Rician Distribution

The Rician Distribution is extremely useful in Communication Theory, especially in non-coherent detection techniques. It is defined as follows:

$$f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2+\nu^2}{2\sigma^2}} I_0\left(\frac{x\nu}{\sigma^2}\right)$$

The cdf is:

$$F_X(x) = 1 - e^{-\frac{x^2+\nu^2}{2\sigma^2}} I_0\left(\frac{x\nu}{\sigma^2}\right)$$

*Note: Most of the pdfs and cdfs are not only functions of x but of other variables too. However, these haven't been indicated explicitly.

2.2.5 Other Distributions

There are several other probability distributions (discrete and continuous), each one useful for a set of specific applications. Some of them have been enumerated below:

1. The Log-Normal Distribution
2. The Truncated Normal Distribution
3. The Student's T Distribution

4. The Snedecor-Fisher's F Distribution (or simply, The F Distribution)
5. The Nakagami-M Distribution
6. Johnson's SU Distribution
7. The Pareto Distribution

And many more...

2.3 Conditional Distributions

$F_{\mathbf{X}}(x) = P(\mathbf{X} \leq x)$
$F_{\mathbf{X}}(x|M) = P(\mathbf{X} \leq x|M) = \frac{P(\mathbf{X} \leq x, M)}{P(M)}$. Here, the numerator is the probability of the event that M takes place and $\mathbf{X} \leq x$. The result has been derived as a direct consequence of the definition of conditional probability introduced in Chapter 1. Since it is a valid cdf, it satisfies all properties discussed before. Thus, a conditional distribution is a distribution of values for one variable that exists when you specify the values of other variables.
Let's now discuss the memorylessness property for certain distributions:

1. The Exponential Distribution:
$$\begin{aligned} P(\mathbf{X} > t + s|\mathbf{X} > s) &= \frac{P(\mathbf{X} > t + s}{\mathbf{X} > s} \\ &= \frac{\int_{t+s}^{\infty} \lambda e^{-\lambda(t+s)}}{\int_{s}^{\infty} \lambda e^{-\lambda(s)}} \\ &= e^{-\lambda(t)} = P(\mathbf{X} > t) \end{aligned}$$
which is independent of s.
2. The Geometric Distribution:
$$\begin{aligned} P(\mathbf{X} > t + s|\mathbf{X} > s) &= \frac{P(\mathbf{X} > t + s}{\mathbf{X} > s} \\ &= \frac{\sum_{m=t+s}^{\infty}(1-p)^m p}{\sum_{m=s}^{\infty}(1-p)^m p} \\ &= (1-p)^t = P(\mathbf{X} > t) \end{aligned}$$
which is independent of s.

Practice Questions

1. Derive the cumulative distribution functions from the probability density functions for the following:
   (a) Gaussian Distribution
   (b) Exponential Distribution
   (c) Laplace Distribution
   (d) Rayleigh Distribution
   (e) Weibull Distribution
   (f) Maxwell Distribution
   (g) Cauchy Distribution
2. The time to complete a task is an exponentially distributed random variable with parameter 1/2. If the time exceeds 2.5 hours, what is the probability that it exceeds 8 hours?
3. Two players, A and B, play a set of games. The first player to three wins wins the whole set. The outcome of each game is independent. Given that A wins the first game, what is the probability that wins the whole set?
4. What will the total probability theorem discussed in Chapter 1 look like for distributions? Similarly, extend Bayes' Theorem for distributions.
5. When we do not have prior information about an event, we consider its apriori pdf to be a Uniform Distribution. Let the probability of obtaining a head in a coin toss be p. Now, we toss the coin n times, and we observe k heads. Based on this information, what distribution does the probability of getting a head follow? What is the probability of getting a head in the (n+1)th toss?
6. If a random variable $\mathbf{X}$ is normally distributed with mean 2 and variance 1, find $f_{\mathbf{X}}(x|(\mathbf{X}-2)^2 < 1)$
7. n balls are placed in m boxes, such that the number of boxes is at least as much as the number of balls. Find the probability that on selecting n of the m boxes, we will find one ball in each box. Consider all the balls to be distinct.
8. Repeat the previous question for the following cases:
   (a) The balls are not distinct, and there can be at most one ball in a box.
   (b) The balls are not distinct, but multiple balls can be in the same box.
9. An electrical circuit has 20 capacitors, 4 resistors and 10 inductors. The probability that any of them will fail in the time interval (T_1, T_2) is $e^{-T_1/T} - e^{-T_2/T}$. Find the probability that at least 15 components fail in the interval $(T/2, T)$
10. Simpify the following: $P(A|\mathbf{X} \leq x)F_{\mathbf{X}}(x) - P(A) + P(A|\mathbf{X} > x)(1 - F_{\mathbf{X}}(x))$