COMMON DISTRIBUTIONS AND CENTRAL LIMIT THEOREM

1. Common Probability Distributions

We have looked at discrete and continuous random variables. These random variables are described using Probability Density Functions (PDFs) or PMFs (mass functions) for continuous, and discrete random variables. Let's now look at some common distributions that we'll encounter.

1.1. **Uniform distributions.** Perhaps the simplest of PDFs are uniform distributions. In the discrete form, they specify that each of values in the sample space are equally likely. For instance, the rolling of a fair die takes one of 6 possible values and each value is equally likely. If there are n possible values in the sample space, then each value is equally likely with a probability of $\frac{1}{n}$. The mean of the uniform distribution can be proven to be $\mu = \frac{n+1}{2}$ and the variance is $\sigma^2 = \frac{n^2-1}{12}$. For a fair die, $\mu = 3.5$ and $\sigma^2 = \frac{35}{12}$.

For a continuous random variable taking values between a and b, the PDF is given by $f(x) = \frac{1}{b-a}$. We calculate $\mu = \frac{a+b}{2}$ and $\sigma^2 = \frac{(b-a)^2}{2}$.

1.2. **Bernoulli distributions.** A Bernoulli trial is something that has two outcomes. One is called often called success and the other one is denoted as failure. Success occurs with a probability p whereas the failure occurs with a probability of $q = 1 - p$. The random variable X takes on two values: 1 for success and 0 for failure. $f(x) = p, x = 1$ and $f(x) = q, x = 0$. The mean $\mu = p$ and the variance is $\sigma^2 = pq$.

An example of a Bernoulli trial is that of fair coin toss. Here the mean is $\mu = 0.5$ and $\sigma^1 = 0.25$.

1.3. **Exponential Distribution.** Another common distribution we encounter often is the Exponential distribution. This usually occurs when we want to describe the phenomenon of events occurring uniformly at random but at a fixed rate per unit time. For instance, let's say that you operate a web server and it sees roughly 100 requests per second. λ, the rate of events per unit time is 100 in this case. We can describe a random variable X which gives the time to the next event. Then, the PDF for X is $f(x) = \lambda e^{-\lambda x} \forall x \in [0, \infty)$. This essentially says that the random variable X takes values between 0 and ∞ and there is a exponential decay in the probability of having gap between events. The mean gap between events is $\mu = \frac{1}{\lambda}$ and the variance $\sigma^2 = \frac{1}{\lambda^2}$.

1.4. **Normal Distribution.** Perhaps the distribution that is encountered most commonly in practice is the Normal Distribution aka Gaussian Distribution. Normal distribution is ubiquitous in probability and statistics. A Normal distribution is defined by 2 parameters, mean μ and standard deviation σ and is denoted by $N(\mu, \sigma)$. It is defined as

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

We'll see why Normal distributions are so common in statistics when we study the Central Limit Theorem.

2. Central Limit Theorem

Central Limit Theorem is perhaps the most important concept in probability theory. It states that given certain assumptions about probability distributions, the arithmetic mean or average of sets of samples of independently distributed random variables will follow the normal distribution. Given a probability distribution with a well-defined mean and variance, let's assume that we draw several independent samples from it, each of some size n. As n grows large, the mean value of each such sample set will be normally distributed around the mean of the original distribution.

It doesn't matter what distribution we start with, the sample means are normally distributed. Let's see an example. Let's start with a discrete uniform distribution between 0 and 10. From this distribution, let's draw 5 samples several times. In one iteration, perhaps we get $(0, 4, 5, 9, 7)$. The mean of this set is 5. In another iteration, perhaps we get $(7, 3, 9, 6, 8)$. The mean of this set is 6.6. If we take many such sets of samples, we'll get several means, one for each sample. The Central Limit Theorem states that these means will obey the normal distribution and will be centered at the mean of the distribution. In this case, it will be normally distributed around 5, which is the mean of the uniform distribution we started with.

Proof of the Central Limit Theorem requires a bit of familiarity with Calculus and Taylor Series approximations, which are coming later in the course. For now, we'll assume that the theorem is true and prove it later when we encounter Calculus later in the course. It is however, extremely easy to empirically verify by simulation.

Central Limit Theorem is not to be confused with the Law of Large Numbers. Law of Large Numbers states that if you take a sufficiently large sample, then the sample mean approaches the Expected Value of the distribution. For instance, let's start a uniform distribution between $(0, 1)$. If we take a large sample from this distribution, the sample mean of that distribution will approach the Expected Value of the distribution, which is 0.5. The Law of Large Numbers is very intuitive and is frequently used when we estimate sample means and variances from data. The more data points we have, the closer we get to the Expected Value of the distribution. Central Limit Theorem, on the other hand, is used when you are faced with taking multiple samples from a distribution. It tells you how those means will be distributed. When we cover statistics and Hypothesis testing, you'll see how Central Limit Theorem plays a role in computing p or significance values of these tests.