Lecture 1

Overview of some probability distributions.

In this lecture we will review several common distributions that will be used often throughtout the class. Each distribution is usually described by its probability function (p.f.) in the case of discrete distributions or probability density function (p.d.f.) in the case of continuous distributions. Let us recall basic definitions associated with these two cases.

Discrete distributions.

Suppose that a set $\mathcal{X}$ consists of a countable or finite number of points,

$$\mathcal{X} = \{a_1, a_2, a_3, \cdots\}.$$

Then a probability distribution $\mathbb{P}$ on $\mathcal{X}$ can be defined via a function $p(x)$ on $\mathcal{X}$ with the following properties:

1. $0 \leq p(a_i) \leq 1$,

2. $\sum_{i=1}^{\infty} p(a_i) = 1$.

A function $p(x)$ is called the probability function. If X is a random variable with distribution $\mathbb{P}$ then $p(a_i) = \mathbb{P}(X = a_i)$ - a probability that X takes value a_i. Given a function $\varphi : \mathcal{X} \to \mathbb{R}$, the expectation of $\varphi(X)$ is defined by

$$\mathbb{E}\varphi(X) = \sum_{i=1}^{\infty} \varphi(a_i)p(a_i).$$

(Absolutely) continuous distributions.

Continuous distribution $\mathbb{P}$ on $\mathbb{R}$ is defined via a probability density function (p.d.f.) $p(x)$ on $\mathbb{R}$ such that

$$p(X) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} p(X)dx = 1.$$

If a random variable X has distribution $\mathbb{P}$ then the probability that X takes a value in the interval $[a, b]$ is given by

$$\mathbb{P}(X \in [a, b]) = \int_a^b p(x)dx.$$

Clearly, in this case for any $a \in \mathbb{R}$ we have $\mathbb{P}(X = a) = 0$. Given a function $\varphi : \mathcal{X} \to \mathbb{R}$, the expectation of $\varphi(X)$ is defined by

$$\mathbb{E}\varphi(X) = \int_{-\infty}^{\infty} \varphi(x)p(x)dx.$$

Notation. The fact that a random variable X has distribution $\mathbb{P}$ will be denoted by $X \sim \mathbb{P}$.

Normal (Gaussian) Distribution $N(\alpha, \sigma^2)$. Normal distribution is a continuous distribution on $\mathbb{R}$ with p.d.f.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\alpha)^2}{2\sigma^2}} \text{ for } x \in (-\infty, \infty).$$

Here $-\infty < \alpha < \infty, \sigma > 0$ are the parameters of the distribution. Let us recall some properties of a normal distribution. If a random variable X has a normal distribution $N(\alpha, \sigma^2)$ then the r.v.

$$Y = \frac{X - \alpha}{\sigma} \sim N(0, 1)$$

has a *standard normal distribution* $N(0, 1)$. To see this, we can write,

$$\mathbb{P}\Big(\frac{X - \alpha}{\sigma} \in [a, b]\Big) = \mathbb{P}(X \in [a\sigma + \alpha, b\sigma + \alpha]) = \int_{a\sigma+\alpha}^{b\sigma+\alpha} \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\alpha)^2}{2\sigma^2}}dx$$

$$= \int_a^b \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy,$$

where in the last integral we made a change of variables $y = (x - \alpha)/\sigma$. This, of course, means that $Y \sim N(0, 1)$. The expectation of Y is

$$\mathbb{E}Y = \int_{-\infty}^{\infty} y\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy = 0$$

since the integrand is an odd function. To compute the second moment $\mathbb{E}Y^2$, let us first note that since $\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$ is a probability density function, it integrates to 1, i.e.

$$1 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy.$$

If we integrate this by parts, we get,

$$1 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy = \frac{1}{\sqrt{2\pi}}ye^{-\frac{y^2}{2}}\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{y}{\sqrt{2\pi}}(-y)e^{-\frac{y^2}{2}}dy$$

$$= 0 + \int_{-\infty}^{\infty} y^2\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy = \mathbb{E}Y^2.$$

Thus, the second moment $\mathbb{E}Y^2 = 1$. The variance of Y is

$$\text{Var}(Y) = \mathbb{E}Y^2 - (\mathbb{E}Y)^2 = 1 - 0 = 1.$$

2

It is now easy to compute the mean and the variance of $X = \alpha + \sigma Y \sim N(\alpha, \sigma^2)$,

$$\mathbb{E}X = \alpha + \sigma\mathbb{E}Y = \alpha, \ \ \mathbb{E}X^2 = \mathbb{E}(\alpha^2 + 2\alpha\sigma Y + \sigma^2 Y^2) = \alpha^2 + \sigma^2,$$

$$\mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \alpha^2 + \sigma^2 - \alpha^2 = \sigma^2.$$

Thus, parameter α is a mean and parameter σ^2 is a variance of a normal distribution. Let us recall (without giving a proof) that if we have several, say n, independent random variables $X_i, 1 \leq i \leq n$, such that $X_i \sim N(\alpha_i, \sigma_i^2)$ then their sum will also have a normal distribution

$$X_1 + \ldots + X_n \sim N(\alpha_1 + \ldots + \alpha_n, \sigma_1^2 + \ldots + \sigma_n^2).$$

Normal distribution appears in one of the most important results that one learns in probability class, namely, a *Central Limit Theorem (CLT)*, which states the following. If $X_1, \ldots, X_n$ is an i.i.d. sample such that $\sigma^2 = \mathrm{Var}(X) < \infty$, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mathbb{E}X_i) = \sqrt{n}(\bar{X} - \mathbb{E}X_1) \to^d N(0, \sigma^2)$$

converges in distribution to a normal distribution with zero mean and variance σ^2, where convergence in distribution means that for any interval $[a, b]$,

$$\mathbb{P}\left(\sqrt{n}(\bar{X}_n - \mathbb{E}X_1) \in [a, b]\right) \to \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx.$$

This result can be generalized for a sequence of random variables with different distributions and it basically says that the sum of many independent random variables/factors approximately looks like a normal distribution as long as each factor has a small impact on the total sum. A consequence of this phenomenon is that a normal distribution gives a good approximation for many random objects that by nature are affected by a sum of many independent factors, for example, person's height or weight, fluctuations of a stock's price, etc.

Bernoulli Distribution $B(p)$**.** This distribution describes a random variable that can take only two possible values, i.e. $\mathcal{X} = \{0, 1\}$. The distribution is described by a probability function

$$p(1) = \mathbb{P}(X = 1) = p, \ \ p(0) = \mathbb{P}(X = 0) = 1 - p \text{ for some } p \in [0, 1].$$

It is easy to check that

$$\mathbb{E}X = p, \ \ \mathrm{Var}(X) = p(1 - p).$$

Binomial Distribution $B(n, p)$**.** This distribution describes a random variable X that is a number of successes in n trials with probability of success p. In other words, X is a sum of n independent Bernoulli r.v. Therefore, X takes values in $\mathcal{X} = \{0, 1, \ldots, n\}$ and the distribution is given by a probability function

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

It is easy to check that
$$\mathbb{E}X = np, \ \ \text{Var}(X) = np(1-p).$$

Exponential Distribution $E(\alpha)$. This is a continuous distribution with p.d.f.
$$p(x) = \begin{cases} \alpha e^{-\alpha x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Here, $\alpha > 0$ is the parameter of the distribution. Again, it is a simple calculus exersice to check that
$$\mathbb{E}X = \frac{1}{\alpha}, \ \ \text{Var}(X) = \frac{1}{\alpha^2}.$$

This distribution has the following nice property. If a random variable $X \sim E(\alpha)$ then probability that X exceeds level t for some $t > 0$ is
$$\mathbb{P}(X \geq t) = \mathbb{P}(X \in [t, \infty)) = \int_t^\infty \alpha e^{-\alpha x} dx = e^{-\alpha t}.$$

Given another $s > 0$, the conditional probability that X will exceed level $t + s$ given that it will exceed level t can be computed as follows:
$$\begin{aligned} \mathbb{P}(X \geq t + s | X \geq t) &= \frac{\mathbb{P}(X \geq t + s, X \geq t)}{\mathbb{P}(X \geq t)} = \frac{\mathbb{P}(X \geq t + s)}{\mathbb{P}(X \geq t)} \\ &= e^{-\alpha(t+s)}/e^{-\alpha t} = e^{-\alpha s} = \mathbb{P}(X \geq s), \end{aligned}$$

i.e.
$$\mathbb{P}(X \geq t + s | X \geq t) = \mathbb{P}(X \geq s).$$

If X represent a lifetime of some object in some random conditions, then the above property means that the chance that X will "live" longer then $t + s$ given that it will "live" longer than t is the same as the chance that X will live longer than t in the first place. Or, in other words, if X is "alive" at time t then it is "like new". Therefore, some natural examples that can be described by exponential distribution are the lifetime of high quality products (or, possibly, soldiers in combat).

Poisson Distribution $\Pi(\lambda)$. This is a discrete distribution with
$$\mathcal{X} = \{0, 1, 2, 3, \ldots\},$$
$$p(k) = \mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} \text{ for } k = 0, 1, 2, ,\ldots$$

It is an exercise to show that
$$\mathbb{E}X = \lambda, \text{Var}(X) = \lambda.$$

Poisson distribution could be used to describe the following random objects: the number of stars in a random area of the space; number of misprints in a typed page; number of wrong connections to your phone number; distribution of bacteria on some surface or weed in the field. All these examples share some common properties that give rise to a Poisson distribution. Suppose that we count a number of random objects in a certain region T and this counting process has the following properties:

4

1. Average number of objects in any region $S \subseteq T$ is proportional to the size of S, i.e. $\mathbb{E}\mathrm{Count}(S) = \lambda|S|$. Here $|S|$ denotes the size of S, i.e. length, area, volume, etc. Parameter $\lambda > 0$ represents the intensity of the process.

2. Counts on disjoint regions are independent.

3. Chance to observe more than one object in a small region is very small, i.e. $\mathbb{P}(\mathrm{Count}(S) \geq 2)$ becomes small when the size $|S|$ gets small.

We will show that under these assumptions will imply that the number $\mathrm{Count}(T)$ of objects in the region T has Poisson distribution $\Pi(\lambda|T|)$ with parameter $\lambda|T|$.
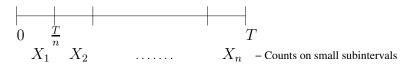

Figure 1.1: Poisson Distribution

For simplicity, let us assume that the region T is an interval $[0, T]$ of length T. Let us split this interval into a large number n of small equal subintervals of length T/n and denote by X_i the number of random objects in the ith subinterval, $i = 1, \ldots, n$. By the first property above,

$$\mathbb{E}X_i = \frac{\lambda T}{n}.$$

On the other hand, by definition of expectation

$$\mathbb{E}X_i = \sum_{k \geq 0} k\mathbb{P}(X_i = k) = 0 + \mathbb{P}(X_i = 1) + \varepsilon_n,$$

where $\varepsilon_n = \sum_{k \geq 2} k\mathbb{P}(X_i = k)$, and by the last property above we assume that ε_n becomes small with n, since the probability to observe more that two objects on the interval of size T/n becomes small as n becomes large. Combining two equations above gives, $\mathbb{P}(X_i = 1) \approx \lambda\frac{T}{n}$. Also, since by the last property the probability that any count X_i is ≥ 2 is small, i.e.

$$\mathbb{P}(\text{at least one } X_i \geq 2) \leq no\left(\frac{T}{n}\right) \to 0 \text{ as } n \to \infty,$$

$\mathrm{Count}(T) = X_1 + \cdots + X_n$ has approximately binomial distribution $B(n, \lambda|T|/n)$ and we can write

$$\mathbb{P}(\mathrm{Count}(T) = X_1 + \cdots + X_n = k) \approx \binom{n}{k}\left(\frac{\lambda T}{n}\right)^k\left(1 - \frac{\lambda T}{n}\right)^{n-k}$$
$$\to \frac{(\lambda T)^k}{k!}e^{-\lambda T}.$$

The last limit is a simple calculus exercise and this is also a famous Poisson approximation of binomial distribution taught in every probability class.

Uniform Distribution $U[0, \theta]$**.** This distribution has probability density function

$$p(x) = \begin{cases} \frac{1}{\theta}, & x \in [0, \theta], \\ 0, & otherwise. \end{cases}$$

Matlab review of probability distributions.

Matlab Help/Statistics Toolbox/Probability Distributions.

Each distribution in Matlab has a name, for example, normal distribution has a name 'norm'. Adding a suffix defines a function associated with this distribution. For example, 'normrnd' generates random numbers from distribution 'norm', 'normpdf' gives p.d.f., 'normcdf' gives c.d.f., 'normfit' fits the normal distribution for a given dataset (we will look at this last type of functions when we discuss Maximum Likelihood Estimators). Please, look at each function for its syntax, input, output, etc. Type 'help normrnd' to quickly see how the normal random number generator works. Also, there is a graphic user interface tools like 'disttool' (to run it just type disttool in the main Matlab window) that allows you to play with different distributions, or 'randtool' that generates and visualizes random samples from different distributions.