

Lecture 2

2.1 Some probability distributions.

Let us recall some common distributions on the real line that will be used often in this class. We will deal with two types of distributions:

1. Discrete
2. Continuous

Discrete distributions.

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

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

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

$p(x)$ is called the probability function. If X is a random variable with distribution \mathbb{P} then $p(a_i) = \mathbb{P}(a_i)$ is a probability that X takes value a_i . Given a function $\varphi : \mathcal{X} \rightarrow \mathbb{R}$, 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 $p(x) \geq 0$ and $\int_{-\infty}^{\infty} p(x)dx = 1$. If a random variable X has distribution \mathbb{P} then the chance/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 $\varphi : \mathcal{X} \rightarrow \mathbb{R}$, 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 $X \sim \mathbb{P}$.

Example 1. Normal (Gaussian) Distribution $N(\alpha, \sigma^2)$ with mean α and variance σ^2 is a continuous distribution on \mathbb{R} with probability density function:

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

Normal distribution often describes continuous random variables that can be affected by a sum of many independent factors, for example, person's height or weight, fluctuations of stock market, etc. In this case, the reason for having normal distribution lies in the Central Limit Theorem.

Example 2. Bernoulli Distribution $B(p)$ 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

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

Example 3. Exponential Distribution $E(\alpha)$ 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.

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

$$\mathbb{P}(X \geq t) = \mathbb{P}(X \in [t, \infty)) = \int_t^{\infty} \alpha e^{-\alpha x} dx = e^{-\alpha t}.$$

For $s > 0$, the probability that X will exceed level $t + s$ given that it exceeded 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.

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

In other words, if we think of X as a lifetime of some object in some random conditions, then this property means that the chance that X will live longer than $t + s$ given that it lived longer than t is the same as the chance that X lives longer than t in the first place. Or, if X is “alive” at time t then it is like new, so to speak. Therefore, some natural examples that can be described by exponential distribution are the life span of high quality products, or soldiers at war.

Example 4. Poisson Distribution $\Pi(\lambda)$ is a discrete distribution with

$$\mathcal{X} = \{0, 1, 2, 3, \dots\},$$

$$p(k) = \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ for } k = 0, 1, 2, \dots$$

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:

1. Average number of objects in any region $S \subseteq T$ is proportional to the size of S , i.e. $\mathbb{E}\text{Count}(S) = \lambda|S|$. 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. $\mathbb{P}(\text{Count}(S) \geq 2)$ becomes small when the size $|S|$ gets small.

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

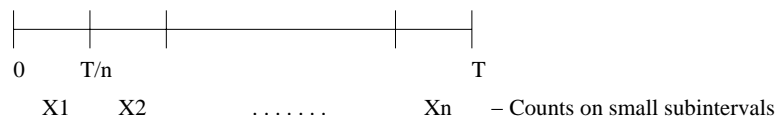


Figure 2.1: Poisson Distribution

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 i th subinterval, $i = 1, \dots, n$. 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 than 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.

$$\mathbb{P}(\text{at least one } X_i \geq 2) \leq n o\left(\frac{T}{n}\right) \rightarrow 0 \text{ as } n \rightarrow \infty,$$

we can write,

$$\begin{aligned} \mathbb{P}(\text{Count}(T) = X_1 + \dots + X_n = k) &\approx \binom{n}{k} \left(\frac{\lambda T}{n}\right)^k \left(1 - \frac{\lambda T}{n}\right)^{n-k} \\ &\rightarrow \frac{(\lambda T)^k}{k!} e^{-\lambda T} \end{aligned}$$

Example 5: Uniform Distribution $U[0, \theta]$ has probability density function

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

Finally, let us recall some properties of normal distribution. If a random variable X has normal distribution $N(\alpha, \sigma^2)$ then the r.v.

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

has standard normal distribution. To see this, we can write,

$$\begin{aligned} \mathbb{P}\left(\frac{X - \alpha}{\sigma} \in [a, b]\right) &= \mathbb{P}(X \in [a\sigma + \alpha, b\sigma + \alpha]) = \int_{a\sigma + \alpha}^{b\sigma + \alpha} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\alpha)^2}{2\sigma^2}} dx \\ &= \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy, \end{aligned}$$

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 we integrate 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,

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} y e^{-\frac{y^2}{2}} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{y}{\sqrt{2\pi}} (-y) e^{-\frac{y^2}{2}} dy \\ &= 0 + \int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \mathbb{E}Y^2. \end{aligned}$$

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

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