Institut National des Sciences
Appliquées et de Technologie
Département de Mathématiques
et Informatiques

Lecture 1: Review on Probability

0.1 Random Variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable X is just a function $X : \Omega \longrightarrow \mathbb{R}$. We distinguish two types, discrete ones, in short, d.r.v and continuous ones, in short, c.r.v.

i. A random variable X is said to be discrete if its range i.e $X(\Omega)$ is a discrete subset of $\mathbb{R}$ which is qq topological concept. Without going into details we can claim that a subset of $\mathbb{R}$ is discrete if it contains only **isolated** points.

ii. A random variable X is said to be continuous if its range i.e $X(\Omega)$ is a continuous subset of $\mathbb{R}$ i.e a union and/or an intersection of sub-intervals of $\mathbb{R}$.

0.1.1 Distribution function

To express correctly distributions associated to random variables we introduce some useful notations:

i. **The indicator function:** Let I be a subset of $\mathbb{R}$ we associate to it, its indicator function, denoted by $\mathbf{1}_I$ and that associates to each $x \in \mathbb{R}$

$$\mathbf{1}_I(x) = \begin{cases} 0 & \text{if } x \notin I \\ 1 & \text{if } x \in I \end{cases} \tag{1}$$

ii. **The Dirac function:** Let a be an element of $\mathbb{R}$, the associated Dirac function, denoted by δ_a is simply the indicator function of the single element subset $\{a\}$ i.e $\delta_a = \mathbf{1}_{\{a\}}$. For instance $\delta_0(0) = 1$ and zero elsewhere.

I. **Discrete case:** X is a discrete random variable (d.r.v) with $X(\Omega) = \{x_i\}$ iff there exists $\{p_i\} \in [0, 1]$ such that $\sum_i p_i = 1$ and for any $A \in \mathcal{F}$

$$\mathbb{P}\{X \in A\} = \mathbb{P}_X(A) = \sum_i p_i \delta_{x_i}(A). \tag{2}$$

$\mathbb{P}_X$ is called the Mass Distribution Function of the d.r.v X.

i. The Bernoulli MDF, $\mathcal{B}(p)$: $\mathbb{P}_X = (1-p)\delta_0 + p\delta_1$ for a given $p \in]0, 1[$.
It is the simplest case but very important because many complex MDF can be written in function of it. It is always considered to model random situation with two outputs "1 for Success" and "0 for Failure".

ii. The Binomial MDF, $\mathcal{B}(n,p)$: $\mathbb{P}_X = \sum_{k=0}^{n} p_k \delta_k$ where $p_k = C_n^k p^k (1-p)^{n-k}$ for a given $p \in]0,1[$ and $n \in \mathbb{N}^*$.
It is also very important, it is used in general to model the random number of "Successes" while repeating the same experiment independently n times. In this case $p \in]0,1[$ represents the probability of Success and n the number of trials.

iii. The geometric MDF, $\mathcal{G}(p)$: $\mathbb{P}_X = \sum_{k=0}^{\infty} p_k \delta_k$ where $p_k = (1-p)^k p$ for a given $p \in]0,1[$.
It is very useful, especially for counting the random number of trials before the first "failure".

iv. The negative binomial MDF, $\mathcal{NB}(r,p)$: $\mathbb{P}_X = \sum_{k=0}^{\infty} p_k \delta_k$ where $p_k = C_{k+r-1}^k p^r (1-p)^k$ for a given $p \in]0,1[$ and $r \in \mathbb{N}^*$.
It used to model the random number of trials before the first r "failures".

v. The Poisson MDF, $\mathcal{P}(\lambda)$: $\mathbb{P}_X = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \delta_k$ where $\lambda > 0$.
Used, in general, for modeling the number of arrivals during a unit time period.

II. **Continuous case:** X is a continuous random variable (c.r.v) with $X(\Omega) = I$ (subinterval of $\mathbb{R}$) iff there exists a non-negative function $f_X : I \longrightarrow \mathbb{R}_+$ such that $\int_I f_X(x)dx = 1$ and for any $A \in \mathcal{F}$

$$\mathbb{P}\{X \in A\} = \mathbb{P}_X(A) = \int_{I \cap A} f_X(x)dx. \tag{3}$$

The function f_X is called the Probability Density Function (PDF) of X.

i. The Exponential density, $\mathcal{E}(\lambda)$: $f(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x)$ for $\lambda > 0$.
It is very use to model life time. It is known for its lack of memory: If X denotes the life time of a given item (electronic devise for instance) then

$$\mathbb{P}\{X \geq t + h | X \geq t\} = \mathbb{P}\{X \geq h\}.$$

ii. The Gamma density, $\Gamma(\alpha, \lambda)$: $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x)$ for $\alpha, \lambda > 0$.
It is a generalization of exponential distribution, for instance, the sum of n independent exponential distribution $\mathcal{E}(\lambda)$ is a Gamma distribution $\Gamma(n, \lambda)$, no memoryless property.

iii. The Normal density, $\mathcal{N}(\mu, \sigma)$: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.
May be its is the most important distribution, because of the Central Theorem (CLT) (see Lecture 3).

iv. The Chi-square distribution χ_k^2: $f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{\frac{k}{2}} \mathbf{1}_{\mathbb{R}_+}$, where $k = 1, 2, \cdots$ is called degree of freedom (when $k = 1$, $\mathbb{R}_+$ should be replaced by $\mathbb{R}_+^*$) and Γ is the Gamma function defined by $\Gamma(\alpha) = \int_0^{+\infty} e^{-x} x^{\alpha-1} dx$, for $\alpha > 0$.

Actually the distribution χ^2_k can be obtained when we sum up k independent squares of standard normal distributions i.e if $\{Z_i: \ i = 1, \cdots, k\}$ are k independent standard normal distributions and $X = Z_1^2 + \cdots + Z_k^2$ then $X \sim \chi^2_k$.

0.2 Expectation

Let X be a given r.v (it could be discrete or continuous) and $g: \mathbb{R} \longrightarrow \mathbb{R}$ a regular function, we associate the expectation of $g(X)$ is defined by

$$g(X) = \int_\Omega g(X(\omega))d\mathbb{P}(\omega). \tag{4}$$

One can check that:

- if X is discrete i.e $\mathbb{P}_X$ is given by (2) then

$$\mathbb{E}\big(g(X)\big) = \sum_i g(x_i)p_i. \tag{5}$$

- if X is continuous i.e $\mathbb{P}_X$ is given by (3) then

$$\mathbb{E}\big(g(X)\big) = \int_I g(x)f_X(x)dx. \tag{6}$$

- if $g(x) = x$ we get the expectation of X.
- if $g(x) = (x - \mathbb{E}(X))^2$ we get the variance of X,

$$\mathbb{V}(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big) = \int_I \big(x - \mathbb{E}(X)\big)^2 f_X(x)dx. \tag{7}$$

- if moreover $\mathbb{E}(X^2) < \infty$ we get

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \tag{8}$$

Here are expectations and variances for some usual distributions:

Distribution	**Expectation**	**Variance**
$\mathcal{B}(p)$	p	$p(1-p)$
$\mathcal{B}(n,p)$	np	$np(1-p)$
$\mathcal{G}(p)$	$\frac{1-p}{p}$	$\frac{1-p}{p}$
$\mathcal{NB}(r,p)$	$\frac{pr}{1-p}$	$\frac{pr}{(1-p)^2}$
$\mathcal{P}(\lambda)$	λ	λ
$\mathcal{E}(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\Gamma(\alpha,\lambda)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
$\mathcal{N}(\mu,\sigma)$	μ	σ
χ^2_k	k	$2k$

0.3 Tchebychev Inequality

Theorem 1. *Let X be a given random variable with finite variance, $\mathbb{V}(X) = \sigma^2 < \infty$, denote by $\mu = \mathbb{E}(X)$. Then for any $n \in \mathbb{N}$ we have*

$$\mathbb{P}\left\{|X-\mu| \geq n\sigma\right\} \leq \frac{1}{n^2} \tag{9}$$

or equivalently

$$\mathbb{P}\left\{|X-\mu| \leq n\sigma\right\} \geq 1 - \frac{1}{n^2} \tag{10}$$

Remarks 1. *i. Theorem 1. tells us that, for any fixed $n \in \mathbb{N}$, X belongs to the interval $[\mu - n\sigma, \mu + n\sigma]$ for about more than $1 - \frac{1}{n^2}$ of chances !*

ii. Inequality (10) can also be used to give a confidence interval to μ knowing a sample of values of X, $x_1, \cdots, x_k$, of course for known σ, as follows:
μ belongs to the interval $[\overline{x} - 3\sigma, \overline{x} + 3\sigma]$ with confidence about $1 - \frac{1}{3^2} \simeq 89\%$.

iii. Inequality (9) can be formulated also as

$$\mathbb{P}\left\{|X-\mu| \geq const.\right\} \leq \frac{\sigma^2}{const.^2} \tag{11}$$