Lecture 4: Random Variables, Their Distribution and their Expectation

Sec 1: RANDOM VARIABLES

We are all accustomed to handling real numbers. Thus it might be a good idea to shift the arena for computing probability from the zone of abstract events to the real lines. This is done through the vehicle of random variables. But before we embark into the definition of a random variable ~~lets~~ let us get a fact straight : Random Variable is a function.

A random variable $X : \Omega \longrightarrow \mathbb{R}$, a function from the sample space Ω, to the real line $\mathbb{R}$, such that for any real number $\alpha \in \mathbb{R}$, such that the set

$$S_\alpha = \{\omega \in \Omega : X(\omega) \leq \alpha\} \subset \mathcal{F}$$

i.e. the set S_α is an event.

Consider the random experiment of tossing a coin, i.e. $\Omega = \{H, T\}$

Define random variable

$$\boxed{X(H) = 1, \quad X(T) = 0.}$$

Suppose we consider a σ-algebra made out of open intervals in $\mathbb{R}$, i.e. ~~doi~~ taking their unions, complements, etc. We call such a σ-algebra as the Borel σ-algebra. In fact for a random variable the set

$$S_I = \{\omega : X(\omega) \in I\},$$

where I is an open interval is also an event. Define for each I

$$P_X(I) = P\{\omega : X(\omega) \in I\},$$

We leave it to the reader should be able to show that $P_X(I)$ indeed satisfies, the Kolmogorov Axioms. Thus if $\mathcal{B}$ is the notation for Borel σ-algebra, then $(\mathbb{R}, \mathcal{B}, P_X)$ is a probability space induced by the random variable X.

Lecture 4: Random Variables, Their Distribution and their Expectation

Sec 1: RANDOM VARIABLES

We are all accustomed to handling real numbers. Thus it might be a good idea to shift the arena for computing probability from the zone of abstract events to the real lines. This is done through the vehicle of random variables. But before we embark into the definition of a random variable ~~lets~~ let us get a fact straight: Random Variable is a function.

> A random variable $X : \Omega \to \mathbb{R}$, a function from the sample space Ω, to the real line $\mathbb{R}$, such that for any real number $\alpha \in \mathbb{R}$, such that the set
> $$S_\alpha = \{\omega \in \Omega : X(\omega) \leq \alpha\} \subset \mathcal{F}$$
> i.e. the set S_α is an event.

Consider the random experiment of tossing a coin, i.e. $\Omega = \{H, T\}$

Define random variable

$$\boxed{X(H) = 1, \quad X(T) = 0.}$$

Suppose we consider a σ-algebra made out of open intervals in $\mathbb{R}$, i.e. ~~doi~~ taking their unions, complements, etc. We call such a σ-algebra as the Borel σ-algebra. In fact for a random variable the set

$$S_I = \{\omega : X(\omega) \in I\},$$

where I is an open interval is also an event. Define for each I

$$P_X(I) = P\{\omega : X(\omega) \in I\},$$

~~g~~ We leave it to the reader should be able to show that $P_X(I)$ indeed satisfies, the Kolmogorov Axioms. Thus if $\mathcal{B}$ is the notation for Borel σ-algebra, then $(\mathbb{R}, \mathcal{B}, P_X)$ is a probability space induced by the random variable X.

Suppose the range of X is finite, i.e. Range $X = \{x_1, \dots, x_n\}$

Then consider the sets

$$\Omega_i = \{\omega : X(\omega) = x_i\}, \quad i = 1 \dots n$$

Then $\Omega_i \cap \Omega_j = \phi, \ \forall\, i \neq j$ & $\bigcup_{i=1}^{n} \Omega_i = \Omega$, i.e. a random variable X with a finite range induce a partition of the sample space Ω. We urge the reader to extend this idea to a random variable (r.v. for short) with a countable range.

Any random variable X whose range is either finite or countable is called discrete random variable. When studying discrete random variable we often ask the question, what is the probability that X takes the value say x_k or just say x.

We denote this by $P(X = x)$, which means

$$P(X = x) = P\{\omega \in \Omega : X(\omega) = x\}$$

Now suppose the range of the r.v. is countable, i.e

$$\text{range } X = \{x_1 \dots x_n \dots\} \subseteq \mathbb{R}$$

Any ~~x whic~~ $x \in \mathbb{R}$, which is not in the range has zero probability. Thus it is meaningful to ask the question

What is $P(X = x_k) = P(\{\omega \in \Omega : X(\omega) = x_k\})$

$$= P(\Omega_k)$$

As $\Omega = \bigcup_{k=1}^{\infty} \Omega_k \Rightarrow P(\Omega) = \sum_{k=1}^{\infty} P(\Omega_k)$

$$\therefore \boxed{\sum_{k=1}^{\infty} P(X = x_k) = 1}$$

i.e. the total probability is one.

Now for any $x \in \mathbb{R}$, even for a discrete random variable X, it is meaningful to ask the question, what is the probability that X takes values less than x. Thus

$$P(X \leq x) = \sum_{x_k \leq x} P(X = x_k)$$

So more formally we write

$$\boxed{f_X(x) = P(X = x)}$$

(p.m.f for short)

which is called the probability mass function X of a discrete random variable, and

$$\boxed{F_X(x) = P(X \leq x) = \sum_{x_k \leq x} f_X(x_k)}$$

is called the cumulative distribution function (cdf), or distribution function of the random variable.

Consider the following simple example of a throwing a fair die. Define a random variable $X : \Omega \rightarrow \mathbb{R}$ as follows;

$$X(\omega) = \begin{cases} 1, & \text{if } \omega \text{ is odd} \\ 0, & \text{if } \omega \text{ is even} \end{cases}$$

The range of X is given as range $X = \{0, 1\}$. The probability mass function thus given as

$$f_X(x) = \begin{cases} P(X = 1) = 1/2 \\ P(X = 0) = 1/2 \end{cases}$$

Note: $P(X = 1) = P(\{\omega : X(\omega) = 1\}) = P(\{1, 2, 3\}) = 1/2$, Same approach for calculating $P(X = 0)$.

The distribution function is given as

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/2 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

Observe the way we do the segmenting of the variable, x. The custom is to have strict inequality on the right-hand side which allows us to graphically represented as a step function, the cdf of a discrete r.v. X. In our example of the die, the graph of ~~F(x) is given~~ F_X is given as

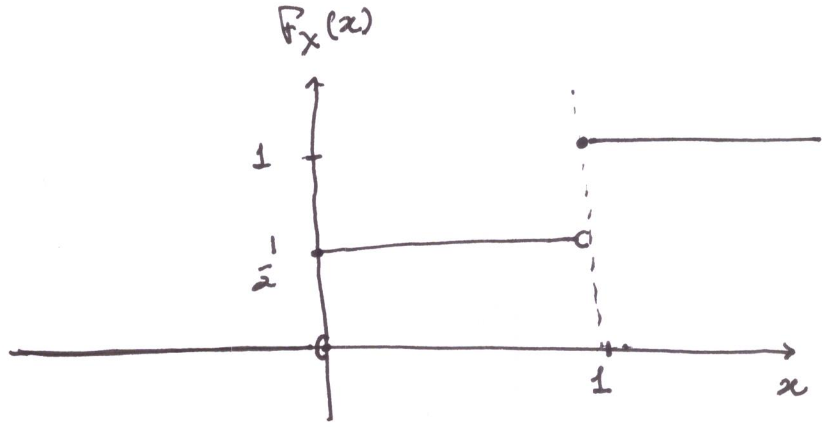


Fig 4.1 c.d.f of X in the previous example.

The c.d.f. or distribution function has the following properties.

i) F_X is a non-negative function, and naturally $F_X(x) \leq 1$

ii) If $x < y$, we have $F_X(x) \leq F_X(y)$.

This is what we gets just by observing the ~~figures~~ above figure of the cdf. It is not a difficult task to prove the above two from the very definition of a c.d.f.

Another important result shows us how to compute a p.m.f. value of the c.d.f. of a discrete r.v. is given.

Let us arrange the values of a r.v. X with finite range in ~~the~~ the ascending order

$$x_1 < x_2 < x_3 < \ldots\ldots < x_n$$

and $f(x_1) = F(x_1)$, and

$$f(x_i) = F_X(x_i) - F_X(x_{i-1}), \quad i = 2 \ldots n$$

This can be proved from the definition of F_X as

$$F_X(x_i) = F_X(x_{i-1}) + f(x_i),$$

(H)

Further observe that F_X is right continuous, i.e.

$$\lim_{x \to a+} F(x) = F(a)$$

It is however not left continuous. Note that we are writing all these properties by looking at the graph of the c.d.f in Fig 4.1. Also observe that

$$\lim_{x \to -\infty} F_X(x) = 0$$

while

$$\lim_{x \to +\infty} F_X(x) = 0$$

You might wonder why the above two properties uses limits. It is done so to consider the ~~countably infin~~ random variables which has a countably infinite range. Let us now summarize the properties of the distribution function F_X of a discrete random variable X, that we have learnt.

i) $F_X \geq 0$ & $F_X \leq 1$

ii) $\lim_{x \to a+} F_X(x) = F_X(a)$

iii) If $x < y \Rightarrow F_X(x) \leq F_X(y)$

iv) $\lim_{x \to -\infty} F_X(x) = 0$ & $\lim_{x \to +\infty} F_X(x) = 1$

Every random variable need not have a finite or countable range. A random variable may have a range which is uncountably ~~finite~~ infinite. A r.v. X is called continuous if 'its' range is uncountably infinite

Suppose we want to measure height of all students here at IIT Kanpur. Then we cannot say that there are only a finite set of ~~point~~ height measures possible. In fact we can reasonably say that hieght of students are in the interval $[4, 7]$, i.e. between 4ft and 7ft. So the hieght of a student is a random variable

$$\Omega \equiv \text{Students} \xrightarrow{\text{Hieght} = X} \mathbb{R}$$

So if X is the random variable denoting the hieght of a student, then

$$4 \leq X(\omega) \leq 7$$

and thus ~~X(ω~~ X has an ~~co~~ uncountably infinite range. Now which question is more meaningful,

$P(X(\omega) = 5\text{ft})$ or $P(4.8 \leq X(\omega) \leq 5.2)$.

So lets first consider the ~~fact~~ issue of $P(4.8 \leq X(\omega) \leq 5.2)$, Since whenever we make a measurement, random errors and observational errors creep in and it is impossible to say that an interval is of exactly ~~this l~~ a particular length. In fact if we measure the same line segment ~~for~~ five times, each time we are likely to get ~~very~~ slightly different results. Keeping our intution of the classical approach we may write

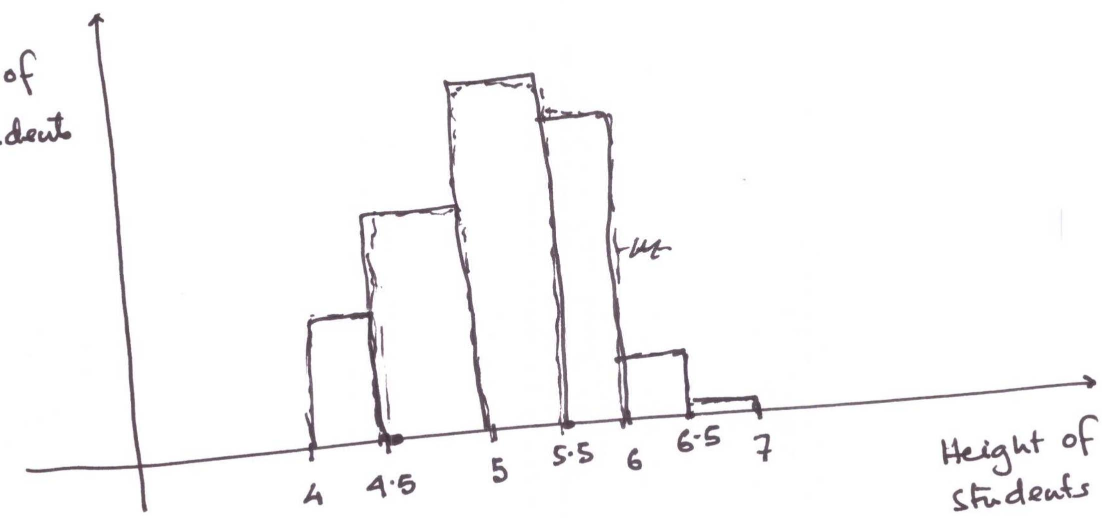
$$P(4.8 \leq x \leq 5.8) = \frac{\text{length of } [4.8, 5.8]}{\text{length of } [4, 7]} = \frac{1}{3}.$$

Hower $P(\cancel{4.8}\ x = 5) = \dfrac{\text{length of } \{5\}}{\text{length of } [4,7]} = 0$

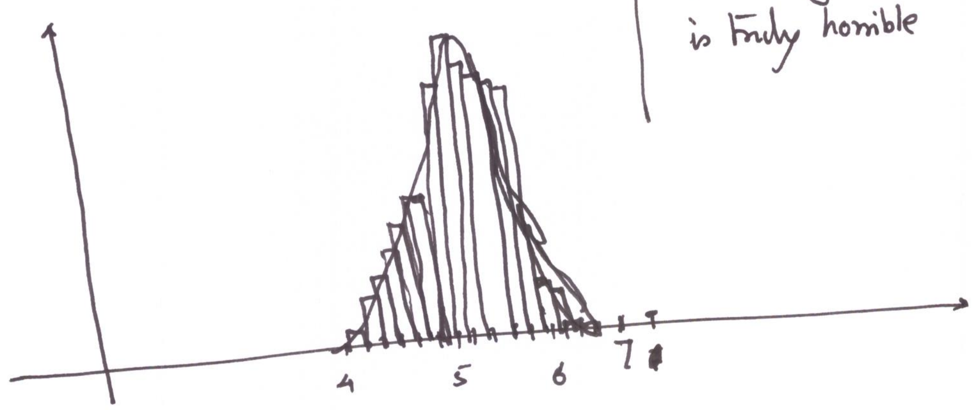
* We are aware of the perils of such an approach Recall Bertrand's paradox.

But is that the only way we deal with the case ~~cas~~ when range of X is uncountably infinite. There could be many situations where our intuition ~~coin~~ from classical probability may fail. A better path is ~~to vis~~ through pictorical representation of data.

Consider ~~a.g~~ again the hieghts of students at IIT Kanpur. Let us divide the hieght range [4,7] into intervals of length 0.5 along the horizontal axis, and in the vertical axis the number of ~~peo~~ students with hieght in ~~that interval~~ those intervals. The resulting diagram is called a histogram which we depict below

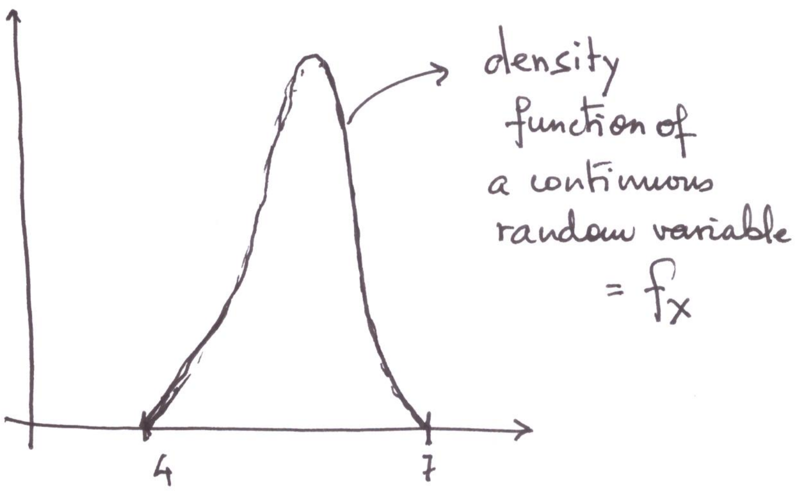


If we make more finer partitions of the interval, then we have a diagram as



This diagram is truly horrible

So as we make the partitioning intervals smaller and smaller we can represent the histogram by a smooth curve which we will call the density function, since each rectangle in some sense represents the density of the people with hieghts in that interval.



Note that the values of the density function of a continuous random variable denoted as f_X does not represent the probabilities, as at individual points the probability is zero while as we see the density function is not.

From a formal perspective, the probability density function of a continuous random variable X, is a function f_X which has the following properties

i) $f_X(x) \geq 0, \quad \forall x$

ii) $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

iii) $P(a \leq x \leq b) = \displaystyle\int_a^b f_X(x)\,dx.$

Another approach to view a continuous random variable X is to assume that it has a distribution function F_X which is continuous. Indeed we have

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(x)\,dx.$$

using (iii). Remember that all other properties of the distribution function remains same, as we have seen for the discrete variable.

Thus for a continuous random variable, X, we have

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx$$

$$= \int_{-\infty}^{b} f_X(x)\,dx - \int_{-\infty}^{a} f_X(x)\,dx$$

$$= F_X(b) - F_X(a).$$

Thus

$$\boxed{P(a \le X \le b) = F_X(b) - F_X(a)}$$

One can also show that $\boxed{\frac{dF_X}{dx} = f_X.}$

One might have a sense that somehow, all these we have discussed for the continuous case is largely intuitive. However there is a result which we will not prove, but will every thing clear.

We have already discussed that given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and X be a random variable associated with it. Then using X we can shift our working space to $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra in $\mathbb{R}$. The following Theorem shows how a unique probability measure $\hat{\mathbb{P}}$ can be defined for the space $(\mathbb{R}, \mathcal{B})$. We present the result as given in Robert Ash*

<u>Theorem 4·1</u> : Let f be a non-negative function on $\mathbb{R}$, with $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Then there exists a unique probability measure $\hat{\mathbb{P}}$ such that for any Borel subset of $B \in \mathcal{B}$ we have

$$\hat{\mathbb{P}}(B) = \int_B f(x)\,dx$$

is a probability measure on $\mathbb{R}$.

> The reader can show that $\hat{\mathbb{P}}(B)$ satisfies the Kolmogorov Axioms.

---

* Robert. B. Ash : Basic Probability Theory, Dover 2008.

As an example consider, the following c.d.f or distribution function.

$$F_X(x) = \begin{cases} 1 - e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

It is a continuous function. It is ~~increasing~~ non-decreasing also and $\lim_{x \to -\infty} F_X(x) = 0$ & $\lim_{x \to +\infty} F_X(x) = 1$. Hence F_X represents a continuous ~~func~~ random variable. distribution function. Hence X is a continuous random variable.

The probability density function associated with X is

$$f_X(x) = \begin{cases} e^{-x}, & \text{when } x \geq 0 \\ 0 & \text{when } < 0 \end{cases}$$

(We just differentiated F_X, i.e we set $f_X(x) = \frac{dF_X}{dx}$).

Let us now check that f_X is truly a density function. Of course $f_X \geq 0$, $\forall x \in \mathbb{R}$ &

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^{\infty} e^{-x}\,dx$$

$$= \left[-e^{-x}\right]_0^{\infty} = 1.$$

There are ~~severs~~ other random variables, which are of the mixed type, i.e. it has both discrete part and continuous.

For our purposes, we shall stick to discrete and continuous variables and in the next chapter we shall look at some special (means largely applicable) distribution of discrete random variables, followed by a chapter on distributions for continuous random variables. Among the continuous variable distributions the normal distribution will play a crucial role in statistics.

[* Mathematicians by the way term the normal distribution as the Gaussian distribution, after the Gauss, the originator of the idea. More on that later.]

Section 2: Expectation of a Random Variable

~~In a~~ If we take a simplistic approach, then expectation means average. We are also aware of the idea of weighted average.

Let us consider n observations (could be hieght or weight... blah blah)

$$x_1, x_2, \ldots, x_n.$$

To each of these observation assign weights $\omega_1, \omega_2 \ldots \omega_n$ respectively & $\omega_i \geq 0$, for all $i = 1 \ldots n$. Then weighted mean ~~means~~ is defined as

$$\bar{x} = \frac{\omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n}{\omega_1 + \omega_2 + \cdots + \omega_n}$$

If we set $\bar{\omega}_i = \dfrac{\omega_i}{\sum_{i=1}^{n} \omega_i}$, where $\sum_i \omega_i = \omega_1 + \cdots + \omega_n$,

then

$$\bar{x} = \bar{\omega}_1 x_1 + \bar{\omega}_2 x_2 + \cdots + \bar{\omega}_n x_n$$

Now for each i, $0 \leq \bar{\omega}_i \leq 1$ & $\sum_{i=1}^{n} \bar{\omega}_i = 1$, and thus the weights $\bar{\omega}_i$ can be viewed as probability ~~associated with occurence of the obse i-th obsow~~ of the i-th observation turning out to be x_i. Thus ~~for~~ taking a clue from the weighted mean the expectation of a discrete random variable, X is given as

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i) = \sum_{i=1}^{\infty} x_i f_X(x_i) \quad \left[\text{provided the series is convergent}\right]$$

The definition of expectation for a continuous random variable is quite analogous and is given as

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx,$$

provided the integral is finite.

Given a random variable $X : \Omega \to \mathbb{R}$, & $g : \text{Range } X \to \mathbb{R}$, the the function $Y = g \circ X$, is a function from $\Omega \to \mathbb{R}$. Is this also a random variable? (The value of Y is computed as $Y(\omega) = g(X(\omega))$

Now consider the set $\{\omega : Y(\omega) \in I\}$, for any interval I, we need to show that $\{\omega : Y(\omega) \in I\} \in \mathcal{F}$, if $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space we are working on.

Observe that $\{\omega : Y(\omega) \in I\} = Y^{-1}(I)$, i.e. to prove that Y is a random variable we have to show that $Y^{-1}(I) \in \mathcal{F}$ for any interval I in $\mathbb{R}$. Observe that

$$Y^{-1}(I) = (g \circ X)^{-1}(I)$$
$$= X^{-1} \circ g^{-1}(I)$$

Now $g^{-1}(I) \subset \text{Range } X$. ~~Thus for For any~~ Any subset of Range X is a Borel set and hence $X^{-1}(g^{-1}(I)) \in \mathcal{F}$ as X is a random variable.

So for $Y = g(X)$, we have, when X is discrete

$$E(Y) = E[g(X)] = \sum_{i=1}^{\infty} g(x_i) f_X(x_i) \text{ (provided the series converges)}$$

and when X is continuous

$$E(Y) = E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx, \text{ (provided the integral is finite).}$$

E.g. If $g(X) = X + c$, where X is constant then

$$E[X+c] = \int_{-\infty}^{\infty} (x+c) f_X(x)\,dx = \int_{-\infty}^{\infty} x f_X(x)\,dx + c\int_{-\infty}^{\infty} f_X(x)\,dx$$
$$= E(X) + c$$
$$\because \int_{-\infty}^{\infty} f_X(x)\,dx.$$

You can check out ~~som~~ some properties in the assignments. Consider ~~again~~ again the random variable X with the distribution

$$f_X(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Then $$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

$$= \int_0^{\infty} x e^{-x}\, dx$$

$$= -xe^{-x}\Big|_0^{\infty} - \int_0^{\infty} \frac{dx}{dx}(-e^{-x})\, dx$$

$$= 0 + \int_0^{\infty} e^{-x}\, dx = 1$$

i.e. $E(X) = 1$.

However for every distribution one need not have a finite expectation. Consider the discrete distribution, where the random variable has countable values, ie X has values $\{0, 1, 2, 3, 4 \ldots\ldots\}$.

Let us have

$$f_X(x) = \frac{1}{2^x}$$

Consider $g(X) = 2^X$. then

$$E(g(X)) = \sum_{x=0}^{\infty} 2^x \frac{1}{2^x} = 1 + 1 + \cdots + 1 + \cdots\cdots$$

Thus The series $1 + 1 + \cdots 1 + \cdots$ is not convergent and hence $E(g(X))$ does not exist. This is often called the St. Petersburg Paradox.

In ~~statistics~~ statistics, $E(X)$ is ~~offe~~ traditionally given the symbol μ. ~~and~~ Another important measure ~~is~~ in statistics is that of variance. Variance is a measure of dispersion. It measures the average square deviation of the random variable values from the mean. Thus for a random variable X

$$\sigma_X^2 = \text{Variance of } X = \text{Var}(X) = E(X-\mu)^2$$

Thus
$$\begin{aligned} E(X-\mu)^2 &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2 - 2\mu X) + \mu^2 \\ &= E(X^2) - 2\mu E(X) + \mu^2 \quad \text{(check this)} \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

$$\therefore \boxed{\text{Variance of } X = \text{Var } X = E(X^2) - E(X)^2}$$

Let us finish this chapter by computing the variance of the random variable X, whose pdf is
$$f_X(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

Thus
$$\begin{aligned} \text{Var}(X) &= E(X^2) - E(X)^2 \\ &= E(X^2) - 1 \\ &= \int_0^\infty x^2 e^{-x}\,dx - 1 \\ &= \left[-x^2 e^{-x}\right]_0^\infty - 2\int_0^\infty x(-e^{-x})\,dx - 1 \\ &= 0 + 2E(X) - 1 \\ &= 2 - 1 = 1. \end{aligned}$$

Chebyshev's inequality says that

$$P\left(|X-\mu| \geq r\sigma_X\right) \leq \frac{1}{r^2}, \quad \text{where } \sigma_X^2 = Var(X) \neq 0$$

~~In~~ Observe that

$$P\left(|X-\mu| \geq r\sigma_X\right) = P\left((X-\mu)^2 \geq r^2\sigma_X^2\right)$$

$$\leq \frac{Var(X)}{r^2\sigma_X^2} \quad (\text{By Theorem } 4.2)$$

$$\Rightarrow \quad P\left(|X-\mu| \geq r\sigma_X\right) \leq \frac{\sigma_X^2}{r^2\sigma_X^2}$$

$$\Rightarrow \quad \boxed{P\left(|X-\mu| \geq r\sigma_X\right) \leq \frac{1}{r^2}}$$

This is Chebyshev's inequality. What it says is the following

$$-P\left(|X-\mu| \geq r\sigma_X\right) \geq -\frac{1}{r^2}$$

$$\Rightarrow \quad 1-P\left(|X-\mu| \geq r\sigma_X\right) \geq 1-\frac{1}{r^2}$$

i.e $\quad \boxed{P\left(|X-\mu| < r\sigma_X\right) \geq 1-\frac{1}{r^2}} \longrightarrow \odot$

In fact this says that

$$\boxed{P\left(\mu - r\sigma_X < X < \mu + r\sigma_X\right) \geq 1-\frac{1}{r^2}}$$

This says that the probability, that the value of the random variable X falls in the open interval $(\mu - r\sigma_X, \mu + r\sigma_X)$ increases as r increases. We will make use of this idea later in our study.

Section 4: Moments and Moment Generating Function

The concept of moments play an important role in the practical applications. The r-th moment is simply the expectation of of the ~~random variable~~ function $g(X) = X^r$ of the random variable X, i.e.

$$\mu'_r = E(X^r)$$

If $r=1$, then $\mu'_1 = E(X) = \mu_X$, the mean or expectation.

There is also a notion of central moment, ie moment centered at the value μ_X, i.e

$$\mu_r = E\left[(X - \mu_X)^r\right]$$

In this case $\mu_1 = 0$, while $\mu_2 = \sigma_X^2$.

But how do we compute moments. This is done through the device of moment generating function, which we now discuss.

The moment generating function of a r.v. X or mgf for short is given as

$$m_X(t) = E\left[e^{tX}\right]$$

For the continuous case we have

$$m_X(t) = E\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{tX} f_X(x)\,dx$$

of course for $m_X(t)$ to be finite this integral must exist for ~~every~~ the given t. In fact it is much better that $m_X(t)$ is finite at least over an open interval, ~~for~~ i.e. $\exists\, h > 0$, s.t $\forall\, t \in (-h, h)$, the function $m_X(t)$ is finite.

For the discrete case of course we have

$$m_X(t) = E(e^{tX}) = \sum_x e^{tx} f_X(x).$$

Now if g_1 & g_2 are functions of the random variable X, then one can easily show that

$$E[g_1(x) + g_2(x)] = E[g_1(x)] + E[g_2(x)]$$

Can this be generalized to the countable case. Let

$$Y = \sum_{i=1}^{\infty} g_i(x)$$

and assume that $Y(\omega)$ is finite for all ω and $E(Y)$ exists, and if $E[g_i(x)]$ exists for each i, then

$$E[Y] = \sum_{i=1}^{\infty} E(g_i(x)) \quad \left[\text{of course under some conditions}\right]$$

provided $\sum_{n=1}^{\infty} E[g_i(x)]$ is also convergent.

$$E[e^{tX}] = E\left[1 + \frac{tx}{1!} + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \cdots + \frac{t^nx^n}{n!} + \cdots\right]$$

$$= 1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \cdots\cdots + \mu_n' \frac{t^n}{n!} + \cdots$$

In fact the coefficient of t^n, is μ_n', for $n = 1, 2, \cdots$

If the expectation $E[e^{tX}]$ is finite then one can show that the above series also converges. ~~Suppose we~~ can ~~integrate under the differential sign.~~ Suppose we can differentiate under the integral sign. Suppose we differentiate the ~~mom~~ mgf, r times, then we get

$$\frac{d^r}{dt^r} m(t) = \int_{-\infty}^{\infty} x^r e^{tx} f_X(x)\,dx$$

$$\Rightarrow \boxed{\frac{d^r}{dt^r} m(0) = \int_{-\infty}^{\infty} x^r f_X(x)\,dx = \mu_r'}$$