Engineering Probability Lecture 5
Random Variables

In many chance experiments the outcomes are only *indirectly* known through some measurement or *observable*. It is a bit like getting a read-out from an instrument. The read-out function does not produce the same value every time you do the experiment. This is the essence of the random or chance nature of the experiment. We call such observable functions *random variables*.

Definition: Given an experiment $\mathcal{E}$ with sample space Ω, a random variable associated to the experiment is a function $X : \Omega \to \mathbb{R}$. [Initially, we confine attention to real-valued observables.]

The following picture captures the main idea.

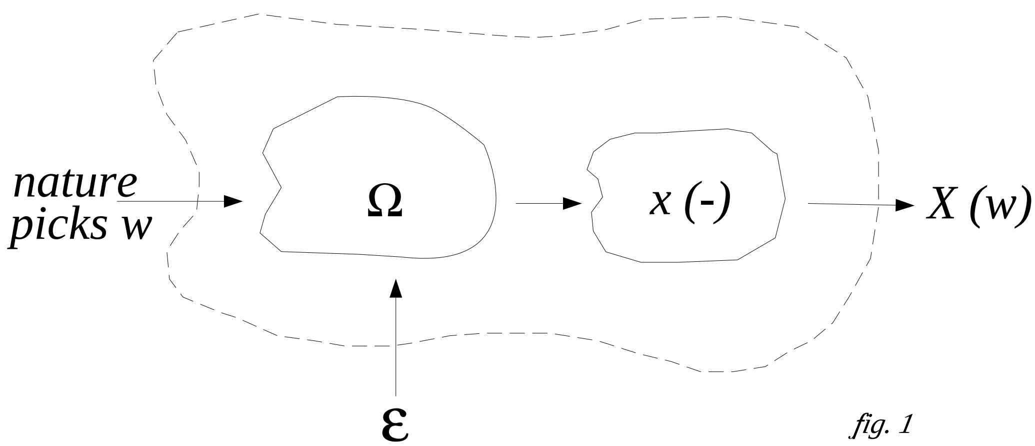

fig. 1

If the experimenter can access only the value of X, it is as if there is a (dotted) box as in *figure 1* and nothing inside the box is directly accessible.

Example 1:
$\mathcal{E}$:= insert a light bulb in a socket, switch on the light, wait till it burns out. Record when this happens

$$\Omega := \text{possible dates and times of burn out.}$$

$$X : \Omega \to \mathbb{R}_+ = \text{non-negative real numbers}$$

$$\omega \longmapsto X(\omega) = \text{lifetime of bulb}$$

$$= \omega - \text{ (time when the bulb was switched on)} \qquad \square$$

Example 2:
$\mathcal{E}$:= Pair of coin tosses

$$\Omega := \{HH, HT, TH, TT\}$$

Suppose we make up X and Y as follows

$$X(\omega) = \begin{cases} \frac{1}{2} & \text{if} \quad \omega \epsilon \{HH, TT\} \\ -\frac{1}{2} & \text{if} \quad \omega \epsilon \{TH, HT\} \end{cases}$$

$$Y(\omega) = \begin{cases} 1 & \text{if} \quad \omega \epsilon \{HH\} \\ 0 & \text{if} \quad \omega \epsilon \{HH, HT\} \\ \frac{1}{2} & \text{if} \quad \omega \epsilon \{TH, TT\} \end{cases}$$

Only X is a random variable, and Y is not. Why? Y is **not** a function. So it cannot be a random variable. $\square$

To avoid mixing up a function and its value, we reserve uppercase letters for functions that are random variables. A value $X(\omega)$ is denoted as x.

Example 3: Suppose we have a sequence of n tosses of a given coin. For each toss we have $\Omega = \{H, T\}$. For the i^{th} toss, let $X_i : \Omega \to \mathbb{R}$ be defined by

$$X_i(\omega) = \begin{cases} 1 & \omega = H \\ 0 & \omega = T \end{cases}$$

$$i = 1, 2, \ldots, n.$$

Thus we have n random variables associated to the entire sequence of coin tosses.

We can take the entire sequence of experiments as *one giant experiment* $\tilde{\mathcal{E}}$ with sample space

$$\begin{aligned} \tilde{\Omega} &= \{HH\cdots H, \quad HTH\cdots H, \cdots\} \\ &= \text{set of sequences in H and T, of length n.} \end{aligned}$$

Then $X : \tilde{\Omega} \rightarrow \mathbb{R}$ may be defined as,

$$\begin{aligned} X(\omega) &= \text{total number of times H came up} \\ &= \sum_{i=1}^{n} X_i(\omega_i) \end{aligned}$$

where ω_i = outcome of just the i^{th} coin toss and X_i as before. Since X aggregates the X_i, it is "less informative" than the collection of X_i. □

Returning to *fig. 1*, one can stay entirely on the outside of the dotted line by defining the events in the range of X.

Definition: Let $\mathcal{E}$ be an experiment, Ω the sample space and $X : \Omega \rightarrow \mathbb{R}$ a random variable. Let R_X = range of X = set of values taken by $X(\omega)$ as ω varies in Ω. One can work with an algebra of events, denoted as $\mathcal{R}_X$, analogous to the Boolean algebra of interesting events (subsets of Ω) introduced before. Whenever $B \epsilon \mathcal{R}_X$, *i.e.* $B \subset R_X$ is such that,

$$A = \{\omega \epsilon \Omega \mid \quad X(\omega) \epsilon B\} \triangleq X^{-1}(B) \epsilon \mathcal{A},$$

one can define the probability of occurrence of B to be simply

$$P(A) = P\left(X^{-1}(B)\right).$$

See *fig. 2*

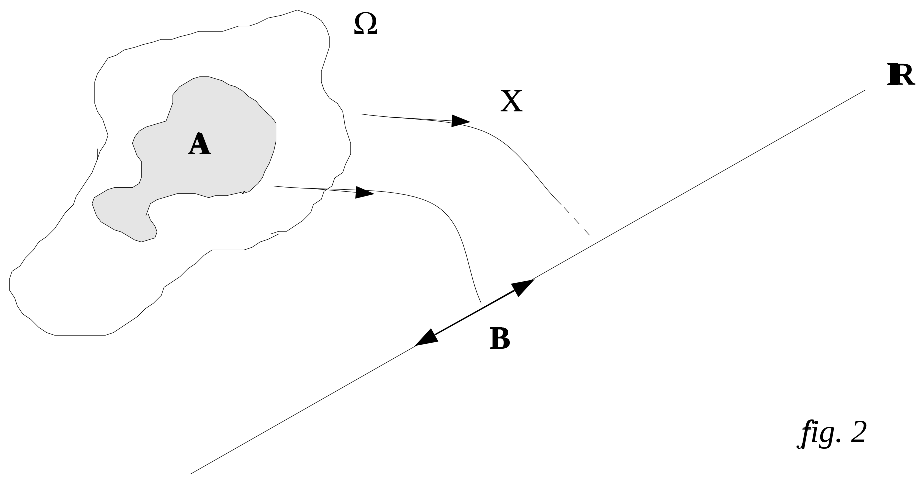

fig. 2

To keep track of how A is related to B, we write:

$$\boxed{P_X(B) \stackrel{\triangle}{=} P(A) = P\left(X^{-1}(B)\right)}$$

Definition: A random variable X is said to be discrete if R_X is a finite or countable set.

Example 4: Let N_t = number of α particles emitted by a gram of Radium in a time interval $[0, t]$. Then N_t is a discrete random variable.

For a discrete random variable X with $R_X = \{x_1, x_2, x_3 \cdots\}$, one associates a *probability mass function* given by,

$$p(x_i) = \text{Prob}\,\{X = x_i\}.$$

It is standard to use the short-hand p_i for $p(x_i)$. The probability mass function is simply the sequence $\{p_1, p_2, p_3, \cdots\}$.

Note that $p_i \geq 0$ and $\displaystyle\sum_{i=1}^{\infty} p_i = 1$

Example 5: (**Geometric** Distribution) Consider an experiment $\mathcal{E}$ in which one tosses a coin repeatedly until it turns up 'Head.' Assume that the successive tosses are independent and the probability of a 'Head' in a single toss

$= p$ (coin may be biased, so p need not be $= 1/2$). Then

$$\Omega = \{H,\ TH,\ TTH,\ TTTH, \cdots\}$$

is a space of sequences. Consider the random variable,

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega &\longmapsto \text{length of } (\omega) \\ \text{Then } R_X &= \{1, 2, 3, \ldots\} \end{aligned}$$

The probability mass function in this case is given by,

$$\begin{aligned} p_k &= \text{Prob}\{X = k\} \\ &= \text{Prob}\{\text{first}(k-1)\text{tosses come up TAIL and } k^{th}\text{comes up HEAD}\} \\ &= (1-p)^{k-1} \cdot p \qquad\qquad k = 1, 2, 3, \cdots \end{aligned}$$

The probability mass function $\{p_1, p_2, \ldots\}$ is a geometric sequence and hence
X is called a *geometric random* variable. $\square$

NOTE: Since $\left(1 + r + r^2 + \cdots + r^{N-1}\right) = \dfrac{1 - r^N}{1 - r}$, it follows that,

$$\begin{aligned} \sum_{k=1}^{\infty} p_k &= lim_{N\to\infty} \sum_{k=1}^{N} p_k \\ &= lim_{N\to\infty} \sum_{k=1}^{N} (1-p)^{k-1} p \\ &= p \ \ lim_{N\to\infty} \frac{\left(1 - (1-p)^N\right)}{(1 - (1-p))} \\ &= p \cdot \frac{1}{1 - 1 + p} \cdot \qquad (\text{since}(1-p) < 1) \\ &= 1 \qquad \text{as it should be.} \end{aligned}$$

Example 6: (**Binomial** Random Variable) Consider an experiment involving n successive, independent coin tosses, with a coin as in Example 5. The sample space $\Omega = \{HH\ldots H, HTHT\ldots H, \ldots\}$ is a space of 2^n sequences. The random variable X is defined by $X(\omega)$ = number of head in ω. Clearly,

$$\begin{aligned} R_X &= \{0, 1, 2, \ldots, n\}, \text{and} \\ p_k &= \text{Prob}\{X = k\} \\ &= \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, 2, \ldots, n. \quad \square \end{aligned}$$

What happens when $n \to \infty,\ p \to 0,$ but $np \to a$ in Example 6?

$$\begin{aligned} \frac{p_k}{p_{k-1}} &= \frac{n!}{k!(n-k)!} \frac{(k-1)!(n-k+1)!}{n!} \frac{p^k(1-p)^{n-k}}{p^{k-1}(1-p)^{n-k+1}} \\ &= \frac{n-k+1}{k} \frac{p}{1-p} \\ &= \frac{np-(k-1)p}{k(1-p)} \to \frac{a}{k} \\ & \qquad \text{as } \ n \to \infty,\ p \to \infty,\ np \to a \end{aligned}$$

Thus

$$\begin{aligned} p_k \to \left(\frac{a}{k} \cdot \frac{a}{k-1} \cdots \frac{a}{1}\right) p_0 &= \frac{a^k}{k!} p_0 \\ p_0 &= (1-p)^n \sim \left(1 - \frac{a}{n}\right)^n \to e^{-a} \text{as } \ n \to \infty. \end{aligned}$$

Thus the sequence $p_k \to e^{-a}\dfrac{a^k}{k!}$ $\qquad \square$.

Definition: The random variable X with $R_X = \{0, 1, 2, \cdots\}$ and probability mass function given by,

$$\begin{aligned} p_k &= \text{Prob } (x = k) \\ &= e^{-a}\frac{a^k}{k!} \qquad k = 0, 1, 2, \cdots \end{aligned}$$

is called a **Poisson** random variable. We have shown that the binomial $\rightarrow$ Poisson.

Example 7: (Lottery) How many lottery tickets should I buy to make the probability of winning at least ϵ?

Solution: In a lottery, out of a total of N tickets, there are M winning tickets. A purchase is a Bernoulli trial with probability $p = \frac{M}{N}$. A set of n purchases is a sequence of n Bernoulli trials, with Prob (holding k winning tickets) $= \frac{a^k}{k!}e^{-a}$ (approximately) where $a = np = \frac{nM}{N}$. We are asking to choose n

$$\begin{array}{rrcl} \text{so that,} & \epsilon & \leq & 1 - P(0) = 1 - e^{a}, \\ \text{equivalently,} & e^{-a} & \leq & 1 - \epsilon, \text{or } \; e^{-\frac{nM}{N}} \leq 1 - \epsilon, \\ \text{equivalently,} & \frac{-nM}{N} & \leq & ln(1 - \epsilon), \\ \text{equivalently,} & \frac{nM}{N} & \geq & -\ln(1 - \epsilon), \\ \text{equivalently,} & n & \geq & -\frac{N}{M} ln(1 - \epsilon) \qquad \square \end{array}$$

Cumulative Distribution Function

As already discussed, it is possible to work with R_X instead of Ω. Similarly, instead of probabilitites defined on an algebra $\mathcal{A}$ of interesting events, one can work with an equivalent concept of the cumulative distribution function.

Definition: Let X be a random variable. The cumulative distribution function associated to X denoted as the $\mathrm{c \cdot d \cdot f \cdot}$ $F_X(\cdot)$ is defined by

$$F_X(x) = \text{Prob}\,\{\omega \epsilon \Omega : X(\omega) \leq x\} \quad \square$$

Knowing the $\mathrm{c \cdot d \cdot f \cdot}$ we can determine interesting probabilities.
Suppose $x_1 < x_2$
Prob $\{\omega \epsilon \Omega : x_1 < X(\omega) \leq x_2\} =$ Prob $\{\omega \epsilon \Omega : X(\omega) \leq x_2\}-$Prob $\{\omega \epsilon \Omega : X(\omega) \leq x_1\}$

This follows from the disjoint union,

$$\{\omega : X(\omega) \leq x_2\} = \{\omega : X(\omega) \leq x_1\} \cup \{\omega : x_1 < X(\omega) \leq x_2\}$$

For a discrete random variable X with sample space Ω and range

$$R_X = \{x_1, x_2, x_3, \cdots\}$$

where the x_i's are ordered $x_k < x_{k+1} \qquad k = 1, 2, \ldots,$
and probability mass function given by,

$$p_k = P(X = x_k) = \text{Prob}\{\omega : X(\omega) = x_k\},$$

the cumulative distribution function is given by,

$$F_X(x) = \sum_{k=1}^{\infty} p_k U(x - x_k)$$

Here $U(x - x_k)$ is the unit step function

$$U(x - x_k) \quad = \quad \begin{cases} 0 & x < x_k \\ 1 & x \geqslant x_k \end{cases}$$

This follows directly from the definition. The resulting picture of the c · d · f· is that of a staircase function.

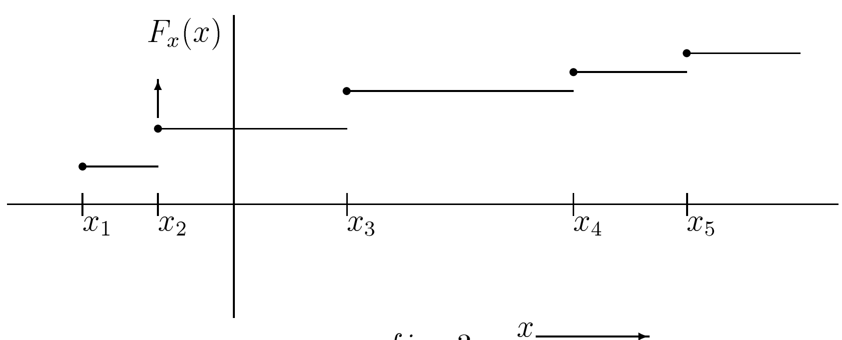

fig. 3

The jump at $x = x_k$ is p_k.

Irrespective of whether a random variable is of the discrete variety or something else, the very definition of the c · d · f· leads to some basic properties.

$$
\begin{array}{llcl}
(i) & F_X(x) & \geqslant & 0 \qquad \forall x \epsilon \mathbb{R} \\
(ii) & \lim_{x \to -\infty} F_X(x) & = & 0 \\
 & \lim_{x \to +\infty} F_X(x) & = & 1 \\
(iii) & x, y \text{ such that } x & \leq & y \Rightarrow F_X(x) \leq F_X(y) \\
 & & & \text{(monotone increasing property)}
\end{array}
$$

(iv) F_X is right continuous, i.e.

$$F_X(x) = \lim_{h \downarrow 0_+} F_X(x+h)$$

Only the last in the list above requires some extra concepts which we consider beyond our scope.

The random variable in Example 1, the lifetime T of a bulb, it not a discrete random variable. So, we do not speak of a probability mass function in this case. Instead, one can start directly from a cumulative distribution function as a given (from physics or from experimental data). The following c · d · f· seems natural,

$$
\begin{aligned}
F_T(t) &= \begin{cases} 1 - e^{-\lambda t}, & t \geqslant 0 \\ 0 & t < 0 \end{cases} \\
\text{Thus } P(T > t) &= 1 - P(T \leq t) \qquad t > 0 \\
&= 1 - (1 - e^{-\lambda t}) \\
&= e^{-\lambda t}
\end{aligned}
$$

Here $\lambda > 0$ is a parameter. The formula above implies that *long lifetimes are highly unlikely*. It implies more. What is the probability that the bulb will last an extra time δ, given that it has lasted until time a ? This is,

$$
\begin{aligned}
P\{T > a + \delta | T > a\} & \\
&= \frac{P(\{T > a + \delta\} \cap \{T > a\})}{P\{T > a\}} \\
&= \frac{P\{T > a + \delta\}}{P\{T > a\}} \\
&= \frac{e^{-\lambda(a+\delta)}}{e^{-\lambda a}} \\
&= e^{-a\delta} \qquad \Box
\end{aligned}
$$

The above formula suggests that your (or bulb's) present age is immaterial to how much longer you (or the bulb) will live. This *memory-less property* may be unrealistic and one may want a different one. Notice that in the present instance, one can differentiate F_T to obtain (fixing-up things at 0), the *density*

$$
\begin{aligned}
p_T(t) &= \frac{dF_T(t)}{dt} \\
&= \begin{cases} \lambda e^{-\lambda t} & t \geqslant 0 \\ 0 & t < 0 \end{cases}
\end{aligned}
$$

The graph of p_T is as in *figure 4.a.* We call T an **exponential** randon vaiable.

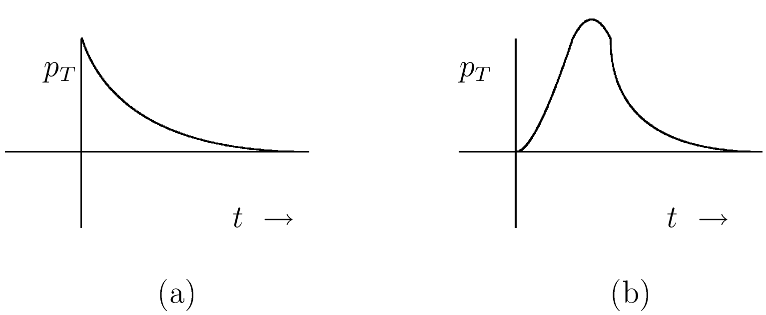

fig. 4

A better choice of density might be as in *fig. 4(b)*. We now are led to the following:

Definition: A random variable X is said to be *continuous* if the associated cumulative distribution can be expressed as

$$F_X(x) = \int_{-\infty}^{x} p_X(y)dy$$

for a suitable piecewise continuous function p_X which we will call the *probability density function* (p · d · f·) of X. It follows that

$$p_X(x) = \frac{dF_X}{dx}$$

and by the properties of the c · d · f·, we conclude

$$\begin{aligned} p_X(x) &\geqslant 0 \qquad\qquad x \epsilon \mathbb{R} \\ \int_{-\infty}^{\infty} p_X(y)dy &= 1 \\ \text{Prob}\{a < X \leq b\} &= \int_a^b p_X(y)dy \end{aligned}$$

NOTE: Even for discrete random variables, if one is willing to work with Dirac delta functions, we can use $\delta(x - x_k) = \frac{d}{dx}U(x - x_k)$ to write

$$p_X(x) = \sum_{k=1}^{\infty} p_k \delta(x - x_k)$$

This is a convenient mnemonic but not essential.

Example 8: (**Uniform** Distribution) Suppose a point ξ is "marked at random" in an interval $[a, b]$. This means that the probability of the mark falling in the sub-interval $[\xi',\ \xi''] \subset [a, b]$ *does not depend* on the location of $[\xi', \xi'']$, just the length of the subinterval $= \xi'' - \xi'$.

Let $P(s)$ denote the probability of falling into a subinterval of length x. Then,

$$P(x + t) = P(s) + P(t)$$

by hypothesis (and axion of addition). This is true for all $s,\ t$ such that the subintervals are in $[a,\ b]$.

Essentially, one function satifies the above functional equation: $P(s) = k \cdot s$ and $k = \frac{1}{b-a}$ because $P(b-a) = 1$. Thus

$$\begin{aligned} P(\xi' < \xi \leq \xi'') & \\ &= \frac{\xi'' - \xi'}{b-c} \\ &= \int_{\xi'}^{\xi''} \frac{1}{b-a} dx \end{aligned}$$

Thus the random variable ξ is continuous with density

$$p_\xi(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & x \notin [a, b] \end{cases} \qquad \square$$

Example 9: (Service System) Suppose customers arrive into a service system according to the law:

$$\begin{aligned} N_t &= \text{number of arrivals in the time} \\ & \quad\ \text{interval } [0, t] \qquad \text{(counter)} \\ \text{Prob}\{N_t = n\} &= e^{-\lambda t} \frac{(\lambda t)^n}{n!} \qquad n = 0, 1, 1, \ldots \end{aligned}$$

The parameter λ is called the *rate* of the arrival process. Let $t_n =$ time instant of n^{th} arrival. Let $T =$ time to the *next* arrival.

$$\text{Prob}\{T \leq \delta\} = 1 - P\{T > \delta\}$$

Proceeding on the assumption that the time t_n is as good an origin of time as 0 for the Poisson counter,

$$\begin{aligned} P\{T > \delta\} &= P\{N_\delta = 0\} \\ &= e^{-\lambda\delta} \\ \text{So } P\{T \leq \delta\} &= 1 - e^{-\lambda\delta} \end{aligned}$$

The interarrival time random variable T is what we call an **exponential** random variable, short for *exponentially distributed*. One can prove (later) that the *assumption* above is correct.

Example 1 0: (**Gaussian** random variable) X is a Gaussian random variable if it has a density

$$p_X(x) \;=\; \frac{1}{c}\exp\left(\frac{-(x-\mu)^2}{2\delta^2}\right) \qquad -\infty < x < \infty$$

where $\mu\epsilon\mathbb{R}$ and $\delta > 0$ are parameters.

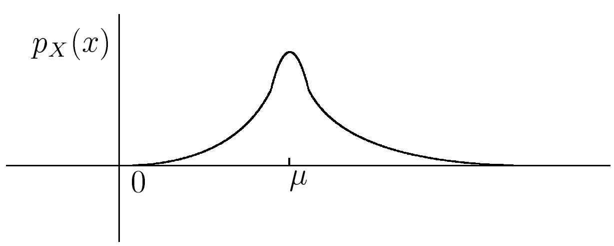

fig. 5

What do these parameters signify? μ is the point about which p_X is symmetric. σ measures the spread of the density function. Greater σ is greater is the spread. What is c? c has to ensure that

$$\int_{-\infty}^{\infty} p_X dx = 1.$$

$$\text{Thus} \qquad c = \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

$$= \int_{-\infty}^{\infty} \exp\left(\frac{-y^2}{2\sigma^2}\right) dy \qquad (\text{change variable} \quad y = x - \sigma)$$

$$= \sqrt{2}\sigma \int_{-\infty}^{\infty} \exp(-z^2) dz \qquad (\text{change} \quad z = \tfrac{y}{\sqrt{2}\sigma})$$

What is $\int_{-\infty}^{\infty} e^{-z^2} dz$? First denote it as $\mathbf{I}$. Then

$$\begin{aligned}
\mathbf{I}^2 &= \int_{-\infty}^{\infty} e^{-z^2} dz \int_{-\infty}^{\infty} e^{-w^2} dw \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-(z^2+w^2)} dz dw,
\end{aligned}$$

a double integral on the (z, w) plane.

Do a change of variable, $(z, w) \rightarrow (r, \theta)$ where $z = r\cos(\theta)$, $\quad w = r\sin(\theta)$. Then $z^2 + w^2 = r^2$, $\quad dzdw = rdrd\theta$ and,

$$\begin{aligned}
\mathbf{I}^2 &= \int_0^{\infty}\int_0^{2\pi} e^{-r^2} r dr d\theta \\
&= 2\pi \int_0^{\infty} e^{-r^2} r dr \\
&= \frac{2\pi}{2}\int_0^{\infty} e^{-y} dy \qquad\qquad (y = r^2) \\
&= \pi \\
\text{Hence } \mathbf{I} &= \sqrt{\pi}
\end{aligned}$$

Thus, $c = \sqrt{2}\ \sigma\ \sqrt{\pi} = \sqrt{2\pi}\sigma$, so the Gaussian density is

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

We also refer to X as a *normal* random variable denoted as,

$$X \sim N(\mu, \sigma^2) \qquad \square$$

Example 11: (**Cauchy** random variable, also sometimes referred to as **Lorentzian**) A random variable X is said to be a Cauchy random variable if

$$p_X(x) = \frac{1}{\pi}\frac{1}{1+(x-x_0)^2} \qquad -\infty < x < \infty$$

with graph symmetrical about x_0 (see *figure 6*).

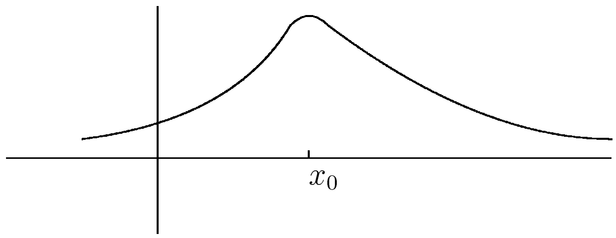

fig. 6

The graph decays slower than that of the Gaussian as $x \to \infty$. $\square$

We now introduce a new concept of quantifying an uncertain or random function. This involves the process of *averaging.*

Suppose a chance experiment, repeated n times, produces the observations $x_1, x_2, x_3, \cdots, x_n$, of a random variable X. Consider the average

$$Avg(X) = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

It has the properties:

$$\begin{array}{llcl}
\text{(i)} \quad \text{Suppose} & X & \geqslant & 0, \text{i.e. } R_X \subset [0, \infty), \text{then} \\
 & Avg(X) & \geqslant & 0. \\
\text{(ii)} & Avg(cX) & = & cAvg(X) \\
\text{(iii)} & Avg(X+Y) & = & Avg(X) + Avg(Y) \\
\text{(iv)} & Avg(1) & = & 1
\end{array}$$

the left hand side, 1 denotes the "random variable" which always takes the value 1.

If X is a discrete random variable with $R_X = \{\alpha_1, \alpha_2, \alpha_3, \cdots\}$, then

$$Avg(X) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$
\begin{aligned}
&= \sum_{i=1}^{\infty} \alpha_i \; r_i \\
\text{where } r_i &= (\text{number of times } \alpha_i \text{ occurs})/n \\
&= \text{relative frequency of } \alpha_i
\end{aligned}
$$

Recall that for large number of trials $r_i \to p_i$ by the frequentist intepretation of p_i. We are led to substituting p_i for r_i in the above expression for average, and hence,

Definition: For a discrete random variable X with $R_X = \{\alpha_1, \alpha_2, \alpha_3, \cdots\}$ and probability mass function given by,

$$
p_i = \text{Prob } \{X = \alpha_i\} \qquad\qquad i = 1, 2, 3, \cdots
$$

the **expectation** of X is

$$
E(X) \;=\; \sum_{i=1}^{\infty} \alpha_i \; p_i \qquad\qquad \Box
$$

All of the properties of the average Avg are satisfied by the expectation. For a continuous random variable X with density p_X,

$$
E(X) \;=\; \int_{-\infty}^{\infty} x p_X(x) dx.
$$

Example 11 (Some expectations)
(a)

$$
\begin{aligned}
X &\sim \text{Binomial } (n, p) \\
R_X &= \{0, 1, 2, \ldots, n\} \\
p_k &= \text{Prob } (X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, 2, \ldots, n
\end{aligned}
$$

$$
E(X) \;=\; \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}
$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} \frac{k\,n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-1-k+1)!} p^k (1-p)^{n-k} \\
&= np \sum_{r=0}^{n-1} \frac{(n-1)!}{r!(n-1-r)!} p^r (1-p)^{n-1-r} \\
&= np(p + (1-p))^{n-1} \\
&= np
\end{aligned}
$$

$$
\begin{aligned}
\text{Thus } p &= \frac{E(X)}{n} \\
&= E\left(\frac{X}{n}\right) \\
&= E\text{ (relative frequency)}
\end{aligned}
$$

(b)

$$
\begin{aligned}
X &\sim \text{Poisson}(a) \\
R_X &= \{0, 1, 2, \ldots\} \\
p_k &= \text{Prob } (X = k) \\
&= e^{-a} \frac{(a)^k}{k!} \qquad k = 0, 1, 2, \ldots \\
E(X) &= \sum_{k=0}^{\infty} k \cdot e^{-a} \frac{a^k}{k!} \\
&= e^{-a} \cdot a \sum_{k=1}^{\infty} \frac{(a)^{k-1}}{(k-1)!} \\
&= e^{-a} \cdot a \cdot e^{a} \\
&= a.
\end{aligned}
$$

(c)

For a Poisson arrival process with $N_t = \#$ arrivals in $[0,t]$ satisfying

$$\begin{aligned}
\text{Prob}(N_t = n) &= e^{-\lambda t}\frac{(\lambda t)^n}{n!} \qquad n = 0, 1, 2, 3, \ldots \\
E(N_t) &= \lambda t \\
\text{Hence } \lambda &= \frac{E(N_t)}{t} \\
&= E\left(\frac{N_t}{t}\right)
\end{aligned}$$

Thus we see a justification for calling λ the arrival rate.

(d)

$$\begin{aligned}
X &\sim \text{Uniform } ([a,b]) \\
R_X &= [a,b] \\
p_X(x) &= \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & x \notin [a,b] \end{cases} \\
\text{Then,} \quad E(X) &= \int_a^b \frac{x}{b-a} dx \\
&= \frac{1}{b-a}\frac{x^2}{2} \Big|_a^b \\
&= \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \\
&= \text{center of range}
\end{aligned}$$

(e)

$$\begin{aligned}
X &\sim N(\mu, \sigma^2) \\
E(X) &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= \int_{-\infty}^{\infty} (x - \mu + \mu) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \\
&\quad +\mu \\
&= 0+\mu \quad \text{because the integrand in the first integral is odd} \\
E(X) &= \mu
\end{aligned}
$$

(f)

$$X \sim \text{Cauchy/Lorentzian}$$

Then $E(X)$ does not exist! Why?