UNIT-II

2.1 - DISTRIBUTION FUNCTION OF A RANDOM VARIABLE

Definition:

Let X be a r.v on $(\Omega, \mathcal{A}, P)$ then the function $F_X(x) = P(X \le x)$

$$= P[\omega : X(\omega) \le x] ; -\infty \le x \le \infty$$

is called the distribution function of the r.v. X.

For discrete case:

$$F_X(x) = P(X \le x) = \sum_x P(x) \; ; \; x : 0, 1, 2, \cdots$$

For continuous case:

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$$

Properties of distribution functions:

(1) If $F_X(x)$ is the distribution function of the r.v X and if $a < b$, then

$$P(a < X \le b) = F(b) - F(a).$$

Proof.

The events $a < X \le b$ and $X \le a$ are disjoint and their union is the event $X \le b$.

Hence, by addition theorem,

$$P(a < X \le b) + P(X \le a) = P(X \le b).$$

$$\Rightarrow P(a < X \le b) = P(X \le b) - P(X \le a)$$

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

—x—

(2) If $F_X(x)$ is the D.F of one-dimensional r.V. X, then i) $0 \le F_X(x) \le 1$ (ii) $F_X(x) \le F_Y(y)$ if $x < y$.

i.e., all D.F's are monotonically non-decreasing between 0 & 1. i.e., $F(-\infty) = 0$ & $F(+\infty) = 1$.

Proof:

To prove (i) $0 \le F_X(x) \le 1$:

Using the axioms of certainity and non-negativity for the prob. function P, ~~it is true for~~

$0 \le F_X(x) \le 1$ is true from the definition of $F_X(x)$.

To prove (ii) $F_X(x) \le F_Y(y)$ if $x < y$

We have, for $x < y$,

$$[X \le y] = [X \le x] + [x < X \le y]$$

$$\Rightarrow P[X \le y] = P[X \le x] + P[x < X \le y]$$

$\because P[x < X \le y] \ge 0$,

$$P[X \le y] \ge P[X \le x]$$

$$\Rightarrow F_Y(y) \ge F_X(x)$$

$$\Rightarrow F_Y(y) - F_X(x) \ge 0$$

$\Rightarrow F_X(x)$ is monotone non-decreasing in x.

Consider the sequence $\{x_n'\}$ where $x_n' \longrightarrow x$.

Since, $\{x < X \le x_n'\} \to \phi$ as $x_n' \to x$,

$$F_X(x_n') - F_X(x) \to 0 \text{ as } x_n' \to x.$$

As this is true for every sequence $\{x_n'\}$, $F_X(x)$ is continuous from right.

Let us define $A_n = [X \le n]$; $n = 1, 2, \ldots$

Then $A_{n+1} \ge A_n$ and $A_n \in [X < \infty] = X^{-1}(R) = \Omega$

$\Rightarrow P[A_n] \to P[\Omega] = 1$ as $n \to \infty$.

Hence, $F_X(\infty) = 1$.

Defining $B_n = [X \le -n]$; $n = 1, 2, \ldots$

$B_n \to \phi$ (null set) as $n \to \infty$.

$\therefore P(B_n) \to 0$ as $n \to \infty$ and $F(-\infty) = 0$.

(3) If $F_X(x)$ is a D.F one dimensional r.v X,

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

and $F(+\infty) = \lim_{x \to +\infty} F_X(x) = 1$

Proof:-

Let S be the whole sample space and as a countable union of disjoint events as:

$$S = \left[\bigcup_{n=1}^{\infty} (-n < X \le -n+1)\right] \cup \left[\bigcup_{n=0}^{\infty} (n < X \le n+1)\right]$$

$$\Rightarrow P(S) = \sum_{n=1}^{\infty} P[-n < X \le -n+1] + \sum_{n=0}^{\infty} P[n < X \le n+1]$$

($\because P$ is additive

$$\Rightarrow 1 = \lim_{a\to\infty} \sum_{n=1}^{a} [F(-n+1) - F(-n)] + \lim_{b\to\infty} [F(n+1) - F(n)]$$

$$= \lim_{a\to\infty} [F(0) - F(-a)] + \lim_{b\to\infty} [F(b+1) - F(0)]$$

$$= [F(0) - F(-\infty)] + [F(\infty) - F(0)]$$

$$\Rightarrow 1 = F(\infty) - F(-\infty) \longrightarrow ①$$

Since, $-\infty < \infty$, $F(-\infty) \le F(\infty)$.

Also $F(-\infty) \ge 0$ and $F(\infty) \le 1$ [by Property ②].

$$0 \le F(-\infty) \le F(\infty) \le 1 \longrightarrow ②$$

① & ② $\Rightarrow$ $F(-\infty) = 0$ & $F(\infty) = 1$.

Probability mass/density function from D.F. :-

The pmf/pdf can be obtained from the DF by differentiating the DF with respect to x.

i.e., $f(x) = F'(x)$

$$= \frac{d}{dx} F(x) \ge 0.$$

$\Rightarrow dF(x) = f(x)\,dx$ in case of continuous r.v X,

$\Rightarrow dF(x) = p(x)\,dx$ " " discrete r.v.

If the D.F $F_X(x)$ increases by jumps, then only it is called a step function.

Suppose a r.v X induces the probability space $(R, \mathcal{B}, P)$ where P has Geometric distn, then 'X' is called Geometric r.v. Since, X takes the values $0, 1, 2, \ldots$ only, it is discrete r.v. Then $F_X(x)$ has jumps at these points and is a step function. The magnitude of the jump at x is $(1-\theta)\theta^x$; $0 < \theta < 1$.

2.2 Decomposition Theorem

Every D.F has a countable set of discontinuity points. Also $F = F_c + F_d$ where F_c is continuous and F_d is a step function. This decomposition is unique.

Proof:

Consider the interval $(l, l+1)$; $l = 0, \pm 1, \pm 2, \cdots$

Let $x_1 < x_2 < \cdots x_n$ be the 'n' points of discontinuity of $F(x)$ in $(l, l+1)$ with jumps of magnitude exceeding $1/m$.

Since, $F(l) \leq F(x_1 - 0) < F(x_1) \leq F(x_2 - 0) < F(x_2) < \cdots$

$$\cdots F(x_n - 0) < F(x_n) \leq F(l+1)$$

denoting $P(x) = F(x_n) + F(x_n - 0)$

$$\sum_{k=1}^{n} P(x_k) = \sum_{k=1}^{n} \left[F(x_k) - F(x_k - 0)\right]$$
$$\leq F(l+1) - F(l)$$

Thus, the no. of discontinuties in $(l, l+1)$ of magnitude exceeding $1/m$ cannot be larger than $m \cdot [F(l+1) - F(l)]$.

Hence, as $m \to \infty$, we may see the number of such intervals is countable.

The first part of the theorem is proved.

Let us define

$$F_d(x) = \sum_{x_k \leq x} P(x_n), \quad x \in R \qquad \text{—— (1)}$$

$$F_c(x) = F(x) - F_d(x) \qquad \text{—— (2)}$$

$F_d(x)$ is non-decreasing, non-negative and continuous on the right on account of (2).

$F_c(x)$ is also non-negative and continuous on the right as well as on the left and decreasing.

Since, for $x' < x$,

$$F_c(x') - F_c(x) = F(x') - F(x) - \sum_{x \le x_k \le x'} P(x_k)$$

$$= F(x'-0) - F(x) - \sum_{x \le x_k \le x'} p(x_k)$$

is non-negative and $\to 0$ as $x \to x'$

This composition is unique because

$$F = F_c + F_d' = F_c' + F_d$$

$$\Rightarrow F_c' - F_c = F_d' - F_d \quad \text{③}$$

The LHS of ③ is Continuous while the RHS is a step function.

This is a contradiction unless both sides vanish Simultaneously. $\Rightarrow F_c = F_c'$ and $F_d = F_d'$.

Example:

Suppose x denotes the life-time of an equipment which may fail immediately on installation with probability $(1-p)$ or live upto the age 'x' with probability $p(1-e^{-\mu x})$, $x \geq 0$.

$$F(x) = 1-p+p(1-e^{-\mu x}), \quad x \geq 0$$

$$= 0, \quad x < 0.$$

where $F(\bar{0}) = 0$, $F(0) = 1-p$, $F(\infty) = 1$.

$F(x)$ is the sum of the step function with jump of magnitude $(1-p)$ at '0' and a Continuous function and It is also a mixture of a DF with jump of magnitude unity at '0' and of an exponential DF,

$$F(x) = 1 - e^{-\mu x}, \quad x \geq 0.$$

with $(1-p)$ and p as weights

2.3 – DISTRIBUTION FUNCTION OF VECTOR RANDOM VARIABLE

Bi-variate Distribution function:

Let (X,Y) be a two-dimensional r.v. Its distribution function $F(x,y)$ is defined as

$$F(x,y) = P[X \le x, Y \le y]$$
$$= P[X \le x \cap Y \le y] \quad \text{— (1)}$$

It is a point function with domain R^2 and Range $[0,1]$.

Since, $[Y < \infty] = \Omega$,

$$F(x,\infty) = P[(X \le x) \cap \Omega]$$
$$= P(X \le x)$$
$$= F_X(x)$$

^{ly} $F(\infty,y) = F_Y(y)$ because $(X < \infty) = \Omega$

$F(x,y)$ is called the joint D.F of 'x' and 'y' while F_X & F_Y are called marginal D.F's.

Consider, $P[x_1 < X \le x_2, \ y_1 < Y \le y_2]$

$$= P[(x_1 < X \le x_2) \cap (y_1 < Y \le y_2)] \quad \text{— (2)}$$

But $(x_1 < X \le x_2) = (X \le x_2) - (X - x_1)$

and $(y_1 < Y \le y_2) = (Y \le y_2) - (Y - y_1)$

$\therefore$ (2) can be rewritten as

$$P[(x_1 < X \le x_2), (y_1 < Y \le y_2)]$$
$$= P[(X \le x_2 - X \le x_1) \cap (Y \le y_2 - Y \le y_1)]$$
$$= P[(X \le x_2, Y \le y_2) - (X \le x_1, Y \le y_2) - (X \le x_2, Y \le y_1) + (X \le x_1, Y \le y_1)]$$
$$= F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1)$$

where $x_1 < x_2$, $y_1 < y_2$.

② Marginal Distribution function:

In case of joint discrete distribution function, $F(x,y)$ the Marginal distribution function of X is given by

$$F_X(x) = \sum_y P(X \le x, Y \le y)$$

and the Marginal distribution function of Y is given by

$$F_Y(y) = \sum_x P(X \le x, Y \le y)$$

In case of joint Continuous distribution function the marginal distribution functions of X and of Y are

$$F_X(x) = \int_{-\infty}^{x} \left\{ \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy \right\} dx \quad \text{and}$$

$$F_Y(y) = \int_{-\infty}^{y} \left\{ \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx \right\} dy$$

③ Distribution function in k-variable case:

Suppose $(X_1, X_2, \dots X_k)$ is a k-dimentional vector r.v. on Ω to R^k, Then sets of the form

$$[-\infty < X_1 \le x_1, \le x_2 \le \dots \le x_k] = \bigcap_{i=1}^{k} [-\infty < X_i \le x_i] \quad \text{①}$$

are measurable.

The probability of ① is denoted by $F(x_1, x_2, \dots x_k)$ and is called d.f of $(X_1, X_2, \dots X_k)$

The LHS will be ϕ if any one of the sets is ϕ and if all the sets in the RHS are equal to Ω.

We have,

$$\begin{aligned} 0 &= F(-\infty, x_1, x_2, \dots x_k) \\ &= F(x_1, -\infty, \dots x_2) \\ &= F(x_1, x_2, \dots x_{k-1}, -\infty) \quad \text{②} \end{aligned}$$

and

$$1 = F(+\infty, \dots +\infty)$$

$$F_{X_i}(x_i) = F(\infty, \dots x_i, \dots \infty) \quad \text{------} (3a)$$

$$F_{X_i Y_j}(x_i, y_j) = F[(+\infty, \dots +\infty), (x_i+\infty), \dots x \dots (x_j+\infty), \dots +\infty] \quad \text{------} (3b)$$

$$F_{X_1, X_2, \dots X_{k-1}}(x_1, x_2, \dots x_{k-1}) = F(x_1, x_2, \dots x_{k-1}, +\infty) \quad \text{------} (3c)$$

The equations 3a, 3b and 3c represent the marginal distribution of X_i, (X_i, X_j) and $(X_1, X_2, \dots X_{k+1})$ respectively.

The necessary and sufficient condition for $F(x_1, x_2, \dots x_k)$ to be the D.F. of k dimensional vector r.V is given by

$$0 \le P\left\{\bigcap_{i=1}^{k} [a_i < x_i \le b_j]\right\}$$

$$\le P\left[\bigcap_{i=1}^{k} \{-\infty < X_i \le b_j\} - \{-\infty < X_i \le a_i\}\right]$$

$$\le \Sigma (-1)^{k-1} F(c_1, c_2, \dots c_k)$$

where c_i denotes a_i or b_i $(i = 1, 2, \dots k)$

2.4 CONDITIONAL DISTRIBUTION FUNCTION

For two dimensional r.v (X,Y), the joint D.F $F_{XY}(x,y)$, for any real numbers 'x' any 'y' is given by

$$F_{XY}(x,y) = P(X \le x, Y \le y)$$

Let A be the event $(Y \le y)$ & the event A' is said to occur when 'Y' assumes values upto y i.e. $(\le y)$.

Using conditional probabilities,

$$F_{XY}(x,y) = \int_{-\infty}^{x} P[A/X=x]\, d F_X(x)$$

The conditional distribution function of Y when X has already assumed the particular value 'x', is given by

$$F_{Y/X}(y/x) = P[Y \le y / X = x] = P[A / X = x] \quad \text{——} (1)$$

Using (1), the joint D.F. may be written as

$$F_{XY}(x,y) = \int_{-\infty}^{x} F_{Y/X}(y/x)\, d F_X(x)$$

and $$F_{XY}(x,y) = \int_{-\infty}^{y} F_{X/Y}(x/y)\, d F_Y(y).$$

The conditional pdf of Y given X for two r.v.s X & Y is defined as

$$f_{Y/X}(y/x) = \frac{\partial}{\partial y} F_{Y/X}(y/x)$$

$$= \frac{f_{XY}(x,y)}{f_X(x)}. \quad \text{if } f(x) > 0.$$

and $$f_{X/Y}(x/y) = \frac{f_{XY}(x,y)}{f_Y(y)} \quad \text{if } f(y) > 0$$

Using differential terms,

$$P[(x < X \le x+dx) / (y < Y \le y+dy)]$$

$$= \frac{P[x < X \le x+dx,\ y < Y \le y+dy]}{P[y < Y \le y+dy]}$$

$$= \frac{f_{XY}(x,y)\,dx\,dy}{f_Y(y)\,dy} = f_{X/Y}(x/y)\,dx$$ which is conditional pdf of X given $Y=y$

INDEPENDENCE / STOCHASTIC INDEPENDENCE

Definition :- (In terms of marginal pdf's).

Two r.v's X & Y with j.p.d.f $f_{xy}(x,y)$ and marginal p.d.f's $f_x(x)$ and $g_y(y)$ respectively are said to be stochastically independent iff

$$f_{xy}(x,y) = f_x(x) \cdot g_y(y)$$

Definition (In terms of D.F's)

Two jointly distributed r.v.s x & y are independent iff their joint D.F. is the product of their marginal DFs

i.e. $F_{xy}(x,y) = F_x(x) \cdot G_y(y)$.

Let us consider two r.v's X & Y with j.p.d.f $f_{xy}(x,y)$ and marginal pdf's $f_x(x)$ and $g_y(y)$ respectively.

By compound probability theorem,

$$f_{xy}(x,y) = f_x(x) \cdot g_y(y/x)$$

where $g_y(y/x)$ = conditional pdf of Y given X = x.

By the defn. of marginal pdf's and if we assume that $g(y/x)$ does not depend on x,

$$g(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{-\infty}^{\infty} f_x(x) \cdot g(y/x)\,dx$$

$$= g(y/x) \int_{-\infty}^{\infty} f(x)\,dx \quad [\because g(y/x) \text{ does not depend on } x.$$

$$\Rightarrow \quad g(y) = g(y/x) \quad [\because f(x) \text{ is p.d.f of } X.$$

& $f_{xy}(x,y) = f_x(x) \cdot g_y(y)$ provided $g(y/x)$ does not depend on X.

This motivates the above definition-1.

Theorem:- Two r.v's X & Y with pdf $f(x,y)$ are independent iff $f_{xy}(x,y)$ can be expressed as the product of a non-negative function of 'x' alone and a non-negative function of 'y' alone.

ie, if $f_{xy}(x,y) = h_x(x) \cdot k_y(y)$ where $h(\cdot)$ & $k(\cdot) \geq 0$

Proof:-

If X & Y are independent, By definition, we have,

$$f_{xy}(x,y) = f_x(x) \cdot g_y(y)$$

where $f(x)$ & $g(y)$ are Marginal pdf of X & Y respectively.

Thus, the condition $f_{xy}(x,y) = h_x(x) \cdot k_y(y)$ is satisfied. — ①

Conversely if ① is true, then we must prove that X & Y are independent.

For Continuous r.v's X & Y, the marginal pdf's are given by

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{-\infty}^{\infty} h(x) \cdot k(y)\,dy$$

$$= h(x) \cdot \int_{-\infty}^{\infty} k(y)\,dy = c_1 h(x) \quad (\text{say}) \quad — ②$$

and

$$g_y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{-\infty}^{\infty} h(x) \cdot k(y)\,dx$$

$$= k(y) \int_{-\infty}^{\infty} h(x)\,dx = c_2 \cdot k(y) \quad (\text{say}) \quad — ③$$

where c_1 & c_2 are constants, independent of x & y.

We know that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

$$\Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x)\,k(y)\,dx\,dy = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} h(x)\,dx \cdot \int_{-\infty}^{\infty} k(y)\,dy = 1$$

$$\Rightarrow c_1 \cdot c_2 = 1 \quad [\text{From ② \& ③}]$$

$$\therefore f_{xy}(x,y) = h_x(x) \cdot k_y(y) = c_1 c_2 h_x(x) \cdot k_y(y) \quad [\because c_1 c_2 = 1]$$

$$= c_1 h_x(x) \cdot c_2 k_y(y) = f_x(x)\, g_y(y) \quad [\text{using ② \& ③}]$$

$\Rightarrow$ X & Y are independent.

Theorem: If X & Y independent r.v's, then for all possible selections of the corresponding pairs of real numbers (a_1, b_1) & (a_2, b_2) where $a_1 \le b_1$ and $a_2 \le b_2$. The values $\pm\infty$ are allowed, the events $(a_1 < X \le b_1)$ and $(a_2 < Y \le b_2)$ are independent.

i.e., $P[(a_1 < X \le b_1) \cap (a_2 < Y \le b_2)]$

$$= P[a_1 < X \le b_1] \cdot P[a_2 < Y \le b_2]$$

Proof:-

Since, X & Y are independent, we have.

$$f_{XY}(x, y) = f_X(x) \cdot g_Y(y)$$

Also for continuous r.v.'s, we have

$$P[(a_1 < X \le b_1) \cap (a_2 < Y \le b_2)] = \int_{a_1}^{b_1}\int_{a_2}^{b_2} f(x,y)\,dx\,dy$$

$$= \int_{a_1}^{b_1} f_X(x)\,dx \int_{a_2}^{b_2} g_Y(y)\,dy$$

$$= P[a_1 < X \le b_1] \cdot P[a_2 < Y \le b_2]$$

Hence, the proof

2.5 0 – 1 LAWS

Basic Definition needed:

Definition Given a probability triple $(\Omega, \mathcal{F}, \mathbf{P})$ and infinitely many events $A_1, A_2, A_3, \ldots \in \mathcal{F}$, define a new event $\{A_n \text{ i.o.}\} \in \mathcal{F}$, read "$A_n$ infinitely often," as

$$\{A_n \text{ i.o.}\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k,$$

and another event $\{A_n \text{ a.a.}\} \in \mathcal{F}$, read "$A_n$ almost always", as

$$\{A_n \text{ a.a}\} = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

Regarding each A_n as a subset of the set Ω, "A_n infinitely often" is the set of all $\omega \in \Omega$ which are in infinitely many of the A_n. "A_n almost always" is more complex to interpret. If $\omega \in \Omega$ is in $\{A_n \text{ a.a.}\}$, then there exists $m_\omega \in \mathbb{N}$ such that

$$\forall\, n > m_\omega,\ \omega \in A_n, \text{ i.e. } \omega \in \bigcap_{i=m_\omega+1}^{\infty} A_i.$$

So $\omega \in \{A_n \text{ a.a.}\}$ implies that ω is in all but a finite number of the A_n. Colloquially, then, we say $\{A_n \text{ a.a.}\}$ is the event that all but finitely many of the events A_n occur.

Lastly, we should note that since $A_1, A_2, \ldots \in \mathcal{F}$, the closure of the σ-algebra under countable operations ensures that $\{A_n \text{ i.o.}\}$, $\{A_n \text{ a.a.}\} \in \mathcal{F}$, and hence are events in their own right with well defined probabilities

Borel Cantelli Lemma (Borel Cantelli 0 -1 Law)

The Borel-Cantelli Lemma *Let $A_1, A_2, \ldots \in \mathcal{F}$.*

$$\text{(i) If } \sum_n \boldsymbol{P}(A_n) < \infty, \text{ then } \boldsymbol{P}(\{A_n \ i.o.\}) = 0.$$

$$\text{(ii) If } \sum_n \boldsymbol{P}(A_n) = \infty, \{A_n\}_{n=1}^{\infty} \text{ independent, then } \boldsymbol{P}(\{A_n \ i.o.\}) = 1.$$

Proof. Let's start with i). for all $m \in \mathbb{N}$, we have:

$$\mathbf{P}(\{A_n \text{ i.o.}\}) = \mathbf{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k\right) \leq \mathbf{P}\left(\bigcup_{k=m}^{\infty} A_k\right)$$

by monotonicity. Then it follows from countable subadditivity of the probability measure that

$$\mathbf{P}\left(\bigcup_{k=m}^{\infty} A_k\right) \leq \sum_{k=m}^{\infty} \mathbf{P}(A_k).$$

Convergence of this last series implies that the terms in the summation go to zero. Hence for all $\epsilon > 0$, there exists $m \in \mathbb{N}$ such that $\sum_{k=m}^{\infty} \mathbf{P}(A_k) < \epsilon$. It follows that $\mathbf{P}(A_n \text{ i.o.}) < \epsilon$ for any ϵ, hence it is 0.

Now for ii). Because $\mathbf{P}(\{A_n \text{ i.o.}\}) = 1 - \mathbf{P}(\{A_n \text{ i.o.}\}^C)$, it suffices to show that $\mathbf{P}(\{A_n \text{ i.o.}\}^C) = 0$. By De Morgan's Laws in the preliminaries we have:

$$\mathbf{P}(\{A_n \text{ i.o.}\}^C) = \mathbf{P}\left(\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right)^C\right) = \mathbf{P}\left(\bigcup_{n=1}^{\infty} \left(\bigcup_{k=n}^{\infty} A_k\right)^C\right) = \mathbf{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^C\right).$$

Then we can write by countable subadditivity that

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^C\right) \leq \sum_{n=1}^{\infty} \mathbf{P}\left(\bigcap_{k=n}^{\infty} A_k^C\right).$$

So we need only show that for all $n \in \mathbb{N}$, $\mathbf{P}(\bigcap_{k=n}^{\infty} A_k^C) = 0$. By monotonicity we have that for all $m \in \mathbb{N}$,

$$\mathbf{P}\left(\bigcap_{k=n}^{\infty} A_k^C\right) \leq \mathbf{P}\left(\bigcap_{k=n}^{n+m} A_k^C\right) = \prod_{i=1}^{n+m} \mathbf{P}(A_k^C),$$

where the second step follows by definition of independent events, and the fact that the complements of independent events are also independent. Now using the fact that for all $x \in \mathbb{R}$, $1 - x \leq e^{-x}$, and letting $\mathbf{P}(A_k^C) = 1 - \mathbf{P}(A_k)$, we have:

$$\prod_{i=1}^{n+m} 1 - \mathbf{P}(A_k) \leq \prod_{i=1}^{n+m} e^{-\mathbf{P}(A_k)} = e^{-\sum_{i=1}^{n+m} \mathbf{P}(A_k)}.$$

Because the sum diverges to infinity, the last term goes to zero as $m \to \infty$. It follows that $\mathbf{P}(\bigcap_{k=n}^{\infty} A_k^C)) < \epsilon$ for any ϵ, so it is 0.

The lemma tells us that if the events $\{A_n\}$ are independent, then $\mathbf{P}(\{A_n \text{ i.o.}\})$ is either 0 or 1 and nothing else. The reader should consider this lemma a very weak version of a zero-one law

Examples:

Example . Consider an infinite heavily weighted coin tossing. Let our independent events be $H_1, H_2, H_3, \ldots$, where H_i is the event that the i^{th} coin is heads. Suppose also that our coins are heavily weighted *against* flipping heads, with $\mathbf{P}(H_n) = \frac{1}{n}$ (e.g. the millionth coin has only a one in a million chance at being heads). The Borel-Cantelli Lemma tells us nevertheless that we will still flip infinitely many heads, i.e. $\mathbf{P}(\{H_n \text{ i.o.}\}) = 1$. This follows from the divergence of the harmonic series, since $\sum_n \mathbf{P}(H_n) = \sum_n \frac{1}{n}$. Such a result is not at all obvious without the lemma.

For an even less obvious result, consider an infinite coin tossing heavily weighted *in favor* of heads. Again let our events be $H_1, H_2, H_3, \ldots$, and suppose that for all $n \in \mathbb{N}$, $\mathbf{P}(H_n) = (\frac{99}{100})^n$. In other words, there is a 99% chance the first coin is heads, a 98.01% chance that the second one is heads, etc. In this scenario, $\mathbf{P}(\{H_n \text{ i.o}\}) = 0$; we cannot have infinitely many heads. This follows from the fact that

$$\sum_n \mathbf{P}(H_n) = \sum_n \left(\frac{99}{100}\right)^n = \frac{\frac{99}{100}}{1 - \frac{99}{100}} = 99 < \infty$$

because this forms a geometric series with common ratio $r = \frac{99}{100}$. This result is entirely unintuitive, but revealing of the great power of the lemma.

Example As a rather amusing example, consider the event that a monkey typing at random would produce Shakespeare's *Hamlet* in an infinite amount of time. Let's ignore case sensitivity, but otherwise we still expect our monkey to type not only all letters, but spaces, quotes, commas, periods, and other punctuation correctly. To be safe, let's assume there are 45 possible characters. Moreover, let's give the monkey an old fashioned typewriter with no delete key so we need not worry about backspaces. *Hamlet* has some finite number of characters N, with N large. Now consider the infinite string produced by our monkey typing at random. We assume for simplicity that each character has the same probability of being hit and that hits are independent. We seek a substring that is the text of *Hamlet*. If we pick an arbitrary starting point in the infinite string the probability that this is the beginning of a full text of *Hamlet* is:

$$\mathbf{P}(H) = \left(\frac{1}{45}\right)^N = \epsilon > 0.$$

Now consider a sequence of events $S_1, S_{N+1}, S_{2N+1}, \ldots$, where S_i is the event that the i^{th} character is the start of a full text of *Hamlet*. These events are independent because they specify the start of a *Hamlet*-length substring of our infinite string with no overlap. Clearly then $\mathbf{P}(S_i) = \epsilon$ always. It follows that

$$\sum_{i=1}^{\infty} \mathbf{P}(S_{2i+1}) = \sum_{i=1}^{\infty} \epsilon = \infty.$$

So by Borel-Cantelli, $\mathbf{P}(\{S_i \text{ i.o.}\}) = 1$. In other words, our monkey will not only type *Hamlet*, but will do so infinitely many times.

Borel 0–1 Law:

If $A_1, A_2, \ldots$ are independent and if E belongs to the σ-field generated by the class $(A_n, A_{n+1}, \ldots)$ for every 'n', then $P(E)$ is 0 or 1.

KOLMOGOROV'S ZERO-ONE LAW

First a few lemmas on independence that we'll need to make use of.

Lemma 1. *Let $B_1, B_2, B_3 \ldots$ be independent. Then $\sigma(B_1, \ldots, B_{i-1}, B_{i+1}, \ldots)$ and $\sigma(B_i)$ are independent classes, i.e. for all $X \in \sigma(B_1, \ldots, B_{i-1}, B_{i+1}, \ldots)$, $\boldsymbol{P}(B_i \cap X) = \boldsymbol{P}(B_i)\boldsymbol{P}(X)$.*

Lemma 2. *Let $A_1, A_2, \ldots, B_1, B_2, \ldots$ be a collection of independent events.*

(i) If $X \in \sigma(A_1, A_2, \ldots)$, then $X, B_1, B_2, \ldots$ are independent.

(ii) The σ-algebras $\sigma(A_1, A_2, \ldots)$ and $\sigma(B_1, B_2, \ldots)$ are independent classes, i.e. if $X \in \sigma(A_1, A_2, \ldots), Y \in \sigma(B_1, B_2, \ldots)$, then $\boldsymbol{P}(X \cap Y) = \boldsymbol{P}(X)\boldsymbol{P}(Y)$.

Definition. Given a sequence of events $A_1, A_2, \ldots \in \mathcal{F}$, we define their *tail field* as

$$\tau = \bigcap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, A_{n+2}, \ldots).$$

The tail field is a σ-algebra whose members we call *tail events*.

The tail field has some interesting attributes in the case that the A_i are independent. Then any $T \in \tau$ cannot depend on any particular event A_i, or on any finite number of events $A_{n_1}, A_{n_2}, \ldots, A_{n_m}$, $n_i \in \mathbb{N}$. If n_{max} is the highest index, then none of these are in $\sigma(A_{n_{max}+1}, A_{n_{max}+2}, \ldots)$ and hence are not in the tail field by Lemma 1 (see below), so bear no relation whatsoever to T. All tail events clearly depend strongly on the tail of our sequence of events; events of this nature depend on infinitely many A_i. As an easy example, $\{A_n \text{ i.o.}\}, \{A_n \text{ a.a.}\} \in \tau$. While the Borel-Cantelli Lemma can only be applied to the events stated, Kolmogorov's Zero-One Law is more powerful in that it applies to any tail event

KOLMOGOROV'S ZERO-ONE LAW

Kolmogorov's Zero-One Law *Given a probability triple $(\Omega, \mathcal{F}, \boldsymbol{P})$ and a sequence of independent events $A_1, A_2, \ldots \in \mathcal{F}$ with tail field τ, if $T \in \tau$, then $\boldsymbol{P}(T) \in \{0, 1\}$.*

Proof. We have an independent collection $A_1, A_2, A_3, \ldots$, and our tail event $T \in \tau$. Then for any $n \in \mathbb{N}$, as $T \in \sigma(A_{n+1}, A_{n+2}, \ldots)$, we have $T, A_1, A_2, \ldots, A_n$ independent by Lemma 2 (i). It follows that $T, A_1, A_2, \ldots$ is an independent collection. If we pick any finite subcollection with indices $m_1, m_2, \ldots m_k$, with m_{max} the largest of these, we need only let $n > m_{max}$ to automatically have T independent from $A_{m_1}, \ldots, A_{m_k}$ by above.

So, with $T, A_1, A_2, \ldots$ independent, by Lemma 5.1 we then have T and S independent for any $S \in \sigma(A_1, A_2, \ldots)$. But we also know by definition that $T \in \tau \subseteq \sigma(A_1, A_2, \ldots)$, i.e. T is independent of itself! It follows that

$$\mathbf{P}(T) = \mathbf{P}(T \cap T) = \mathbf{P}(T)\mathbf{P}(T) = \mathbf{P}(T)^2,$$

so $\mathbf{P}(T) = 0$ or $\mathbf{P}(T) = 1$.

Example Let a_n be a fixed sequence of real numbers. Consider the space $\Omega = \{\pm 1\}^{\infty}$. Consider a *sign sequence* $\omega_n \in \Omega$, such that $\mathbf{P}(\{\omega_i = 1\}) = \frac{1}{2}$ and $\mathbf{P}(\{\omega_i = -1\}) = \frac{1}{2}$, where ω_i is the i^{th} component of the infinite sequence ω_n. Then the event that $\sum_n \omega_n a_n$ converges is a tail event. If the sum converges, then changing the signs of finitely many terms will yield a sum that must also converge; and similarly if it diverges, changing the signs of finitely many terms will still yield a divergent series. As no finite number of sign switches can change the convergence, then the event that the series converges resides in the tail field. Hence, by Kolmogorov's Zero-One Law,

$$\mathbf{P}\left(\sum_n \omega_n a_n \text{ converges}\right) \in \{0, 1\}.$$

End of Unit-II