* A probability space is a triple (Ω, F, P) where Ω is an arbitrary set, F is a σ-field of subsets of Ω, and P is a measure on F such that:

$$P(\Omega) = 1,$$

* Sample space: any set Ω

- Any subset $E \subseteq \Omega$ of the sample space is known as an event.
- set of events (F)
  σ-algebra : $\quad \Omega \in F \;,\; E \in F \Rightarrow E^c \in F \;,\; E_i \in F \Rightarrow \bigcup_i E_i \in F$

- For each event E of the sample space Ω, we assume that a number $P(E)$ is defined and satisfies the following three axioms:

1) $0 \leq P(E) \leq 1 \qquad \forall E \in F$

2) $P(\Omega) = 1$

3) $E \cap F = \emptyset \Rightarrow P(E \cup F) = P(E) + P(F)$

$\Rightarrow P(E^c) = 1 - P(E)$ & $P(\emptyset) = 0$ & if $E \subset F \Rightarrow P(E) \leq P(F)$

3′) generalization: sequence of mutually exclusive events $E_1, E_2, \dots$

$E_i \cap E_j = \emptyset$ when $i \neq j$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

* Inclusion - Exclusion principle:

$P(E \cup F) = P(E) + P(F) - P(E \cap F)$ & E, F two arbitrary events.

* Random variable : a function from Ω (sample space) to $\mathbb{R}$ discrete / Continuous

~~exampl~~ $X: \Omega \longrightarrow \mathbb{R}$

$X: \omega \longrightarrow X(\omega)$

- example : total number of heads in tossing a coin 3 times.

$$\Omega = \{ \underset{\downarrow\, 3}{HHH}, \underset{\downarrow\, 2}{HHT}, \underset{\downarrow\, 2}{HTH}, \cdots, \underset{\downarrow\, 0}{TTT} \}$$

- example: roll a dice 2 times $\Rightarrow \Omega = \{ (i,j) : 1 \leq i,j \leq 6 \}$

$$\omega = (i,j) \Rightarrow \begin{cases} X(\omega) = i \in 1, \cdots, 6 \\ Y(\omega) = \max(i,j) \in 1, \cdots, 6 \\ Z(\omega) = i+j \in 2, \cdots, 12 \end{cases}$$

* Distribution function :

$$F_X(x) \triangleq \mathbb{P}(X \leq x) \quad , \quad x \in \mathbb{R}$$

Properties:

+ non decreasing
+ right-continuous
+ $\mathbb{P}(X > x) = 1 - F_X(x)$
+ $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$
+ $\mathbb{P}(X = x) = F_X(x) - F_X(x-)$
+ $\lim_{x \to -\infty} F_X(x) = 0$
+ $\lim_{x \to +\infty} F_X(x) = 1$

* Probability mass function (Pmf, for discrete r.v.)

$$P_X(x) = \mathbb{P}(X = x) \quad , \quad x \in \mathbb{Z}$$

- examples:

Bernoulli(p) : $P(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \\ 0 & \text{o.w.} \end{cases}$

Binomial (n, P): $P(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, \cdots, n \\ 0 & \text{o.w.} \end{cases}$

Poisson (λ) $P(x) = \begin{cases} e^{-\lambda} \cdot \frac{\lambda^x}{x!} & x = 0, 1, \cdots \\ 0 & \text{o.w.} \end{cases}$

- exercise: Coin with $(p, 1-p)$ is tossed until the first time K Heads obtained. X = number of tosses

$\Omega = ?$, $P_X(x) = ?$

$$P(X=n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

* The probability density function (pdf for continuous r.v.)

$$f_X(x) \triangleq \frac{dF_X(x)}{dx}$$

+ $P(a < x \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\,dx$

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

- examples:

Uniform: $U[a,b]$ $\quad f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{o.w.} \end{cases}$

Gaussian/Normal: $\mathcal{N}(\mu, \sigma^2)$ $\quad f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

Exponential: $E(\lambda)$ $\quad f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{o.w.} \end{cases}$

* Transformations of random variables:

$X \sim F_X$ is continuous r.v.

$Y = h(X) \sim F_Y \Rightarrow F_Y = ?$

+ h is strictly monotonic:

$$F_Y(y) = \mathbb{P}(h(X) \leq y) = \begin{cases} P(X \leq h^{-1}(y)) = F_X(h^{-1}(y)) & : h \text{ increasing} \\ \mathbb{P}(X \geq h^{-1}(y)) = 1 - F_X(h^{-1}(y)) & : h \text{ decreasing} \end{cases}$$

+ h is strictly monotonic and differentiable:

$$f_Y(y) = \left.\frac{f_X(x)}{|h'(x)|}\right|_{x = h^{-1}(y)}$$

* Expectation (mean):

$$\mu_x = E(X) = \sum_x x P_X(x) \quad : \text{discrete r.v.}$$

$$\mu_x = E(X) = \int_{-\infty}^{+\infty} x f_X(x)\, dx \quad : \text{Continuous r.v.}$$

+ Properties : linear

$$E(aX+b) = a\,E(X) + b$$

* Variance :

$$\mathrm{Var}(X) = E[(X - E(X))^2] = E(X^2) - [E(X)]^2 = E[(X-\mu_x)^2] \geq 0$$

it is not linear:

$$\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$$

* Covariance : (is a measure of how much two r.v.s change together)

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

* Cauchy-Schwarz ineq.

$$(E[XY])^2 \leq E[X^2]\,E[Y^2]$$

---

* Conditioning :

$$P(A|B) \triangleq \frac{\overbrace{P(A\cap B)}^{P(A,B)}}{P(B)} \quad \text{if } P(B) \neq 0$$

* is again a probability with all properties ...

+ chain rule: $P(A,B,C) = P(A)\,P(B|A)\,P(C|A,B)$

+ $B_1, \dots, B_n$ disjoint & $\cup B_i = \Omega$ & $P(B_i) > 0$

+ ~~law of total probability~~ Bayes rule:

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A|B_j)\,P(B_j)}{\sum_{i=1}^{n} P(A|B_i)\,P(B_i)}$$

+ law of total probability:

$$P(A) = \sum_{i=1}^{n} P(A|B_i)\,P(B_i)$$

- example: for the example we had for dice

* $P_{X|Y}(x|4) = ?$

$$\text{event } (Y=4) = \{(1,4),(2,4),(3,4),\overset{(4,4)}{,}(4,3),(4,2),(4,1)\} \Rightarrow P(Y=4) = \tfrac{7}{36}$$

$$\Rightarrow P_{X|Y}(x|4) = \begin{cases} \tfrac{1}{7} & x = 1,2,3 \\ \tfrac{4}{7} & x = 4 \\ 0 & \text{o.w.} \end{cases}$$

* $P_{Y|X}(y|5) = ?$

$$\text{event } (X=5) = \{(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)\} \Rightarrow P(X=5) = \tfrac{6}{36} = \tfrac{1}{6}$$

$$\Rightarrow P_{Y|X}(y|5) = \begin{cases} 0 & y<5 \\ \tfrac{5}{6} & y=5 \\ \tfrac{1}{6} & y=6 \end{cases}$$

---

* Independence of events:

+ $A \perp\!\!\!\perp B \iff P(A\cap B) = P(A)\cdot P(B) \Rightarrow P(A|B) = P(A)$ if $P(B) \neq 0$

+ $A_1, \ldots, A_n$ are independent if for any sub-collection $A_{i_1}, \ldots, A_{i_k}$

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

+ $A \perp\!\!\!\perp B \Rightarrow A^c \perp\!\!\!\perp B^c$

* Independence of R.V.s:

$$X \perp\!\!\!\perp Y \iff P[X\leq x, Y\leq y] = P[X\leq x]\cdot P[Y\leq y] \qquad \forall x,y \in \mathbb{R}$$

$$F_{XY}(x,y) = F_X(x)\cdot F_Y(y)$$

* indep $\Rightarrow E[X_1 \cdots X_n] = E[X_1]\cdots E[X_n]$ & $Var(X_1+X_2) = Var(X_1) + Var(X_2)$

$$\text{indep} \underset{\not\Leftarrow}{\Longrightarrow} \text{uncorrelated}$$

- example: $X \sim U[-1,1]$, $Y = X^2 \Rightarrow E[XY] = E[X]E[Y] = 0$ but $X \not\perp\!\!\!\perp Y$

* Conditional expectation:

$E[X|Y]$ is a random variable

$= \psi(Y)$

+ in fact $\forall y$ $E[X|Y=y] = \sum_x x\,\mathbb{P}[X=x|Y=y] = \psi(y)$: for discrete r.v.
+ $E[\psi(Y)] = E_Y[E_X[X|Y]] = E_X[X]$
+ $E[X|X] = X$

---

* Stochastic process (Random process):

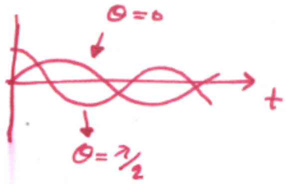
a family of random variables (indexed sequence of r.v.)

$$\{X_t(\omega) : \underbrace{t \in T}_{\text{discrete/continuous}}\}$$

+ fixed $\omega \in \Omega$ $\Rightarrow$ $\{X_t(\omega) : t \in T\}$ is a realization of r.p. (a function of time)
+ fixed $t \in T$ $\Rightarrow$ $\{X_t(\omega) : \omega \in \Omega\}$ random variable

- example:

$$\Theta \sim U[0, 2\pi] \quad \Rightarrow \quad X_t(\Theta) = \sin(\underset{\text{cte}}{a}t + \Theta)$$



* Stationarity of a r.p.

$$\mathbb{P}(X_{t_1} = x_1, \dots, X_{t_n} = x_n) = \mathbb{P}(X_{t_1+h} = x_1, \dots, X_{t_n+h} = x_n) \qquad \begin{array}{l} \forall n \\ \forall t_1, \dots, t_n \\ \forall h > 0 \end{array}$$

* Markov chain (first order)

a stochastic process $X_1, X_2, \cdots$ is a Markov chain or Markov process if

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \cdots, X_1 = x_1) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

+ "time homogeneous" Markov chain or "time invariant" or "stationary"

$$\mathbb{P}(X_{n+1} \cancel{X_n} = b \mid X_n = a) = \mathbb{P}(X_2 = b \mid X_1 = a)$$

+ a time invariant Markov chain is characterized by its initial state and a probability transition matrix $P = [P_{ij}]$ $\quad 1 \leq i,j \leq S \longrightarrow$ # states

$$P_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

+ $\nu_{n+1} = \nu_n \cdot P$: ν_n is the vector of ~~pr~~ state probability

$\nu_{n+k} = \nu_n \cdot P^k$ $\qquad \nu_n = [\nu_n(1) \; \cdots \; \nu_n(S)]$ $\quad S \rightarrow$ # states.

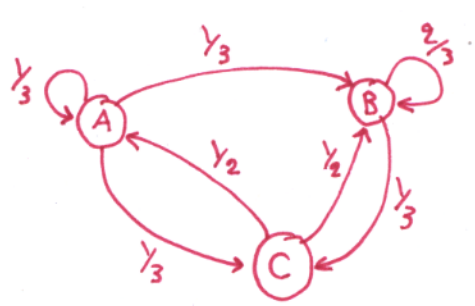
$\mathbb{P}(X_n = i) = \nu_n(i)$

+ stationary distribution :

$$\nu_{n+1} = \nu_n \;\Rightarrow\; \pi = \pi P \quad \text{such that} \quad \pi(i) \geq 0 \,, \; \sum_{j=1}^{S} \pi(j) = 1$$

- example:

$$P = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 2/3 & 1/3 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$



$$[a \;\; b \;\; c] \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 2/3 & 1/3 \\ 1/2 & 1/2 & 0 \end{bmatrix} = [a \;\; b \;\; c]$$

$$a = \frac{3}{16} \,, \quad b = \frac{9}{16} \,, \quad c = \frac{4}{16}$$

* Weak law of large numbers (WLLN)

$X_1, X_2, \dots$ iid r.v. $\mu < \infty$, σ^2, $E(X_k^2) < \infty$ $\forall k \geq 1$

$$A_n = \frac{X_1 + \dots + X_n}{n} \quad : \text{ sample average}$$

$$\Rightarrow \forall \varepsilon > 0, \lim_{n\to\infty} \mathbb{P}(|A_n - \mu| < \varepsilon) = 1 \quad \text{or} \quad A_n \xrightarrow{P} \mu \text{ for } n \to \infty$$

* Strong law of large numbers (SLLN) $\mathbb{E}(|X_k|) < \infty$ $\forall k \geq 1$

$$\mathbb{P}(\lim_{n\to\infty} A_n = \mu) = 1 \quad \text{or} \quad A_n \xrightarrow{a.s.} \mu \text{ for } n \to \infty$$

* Convergence in probability:

$$X_n \xrightarrow{P} X \quad \text{if} \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X| < \varepsilon) = 1 \quad \forall \varepsilon > 0$$

* Almost sure convergence:

$$X_n \xrightarrow{a.s.} X \quad \text{if} \quad \mathbb{P}\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = 1$$

* Convergence in distribution:

$$X_n \xrightarrow{d} X \quad \text{if} \quad \mathbb{P}[X_n \leq x] \xrightarrow{n\to\infty} \mathbb{P}[X \leq x]$$ (at every cont. point of $F_X(x)$)

* Convergence in square mean:

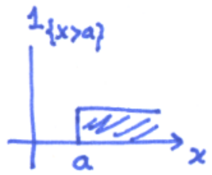
$$X_n \xrightarrow{L^2} X \quad \text{if} \quad \lim_{n\to\infty} E(|X_n - X|^2) = 0$$

* Relationship between different type of convergence:

+ $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$
+ $X_n \xrightarrow{L^2} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$
+ no relation between $\xrightarrow{L^2}$ and $\xrightarrow{a.s.}$

*

$\mathbb{P}\{X \geq a\} = \mathbb{E}[\mathbb{1}_{\{X \geq a\}}]$ where $\mathbb{1}_{\{X \geq a\}}$ is indicator function

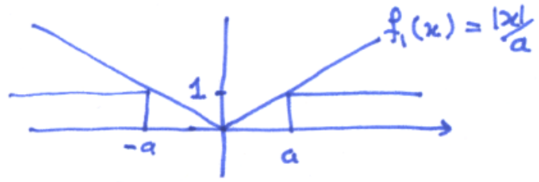


because $\mathbb{E}[\mathbb{1}_{\{X \geq a\}}] = \sum_{x} P_X(x) \cdot \mathbb{1}_{\{x \geq a\}} = \sum_{x \geq a} P_X(x)$

$= \mathbb{P}\{X \geq a\}$

* Markov's inequality : $(\cancel{X \geq 0})$

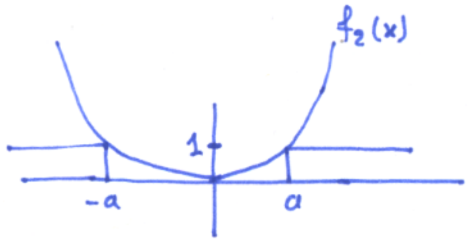
$\mathbb{1}_{\{|x| \geq a\}} \leq f_1(x) \quad \forall a > 0$



$\Rightarrow \mathbb{E}[\mathbb{1}_{\{|X| \geq a\}}] \leq \mathbb{E}[f_1(X)] \Rightarrow \mathbb{P}\{|X| \geq a\} \leq \frac{\mathbb{E}[|X|]}{a} \quad \forall a > 0$

* chebyshev's inequality:

$\mathbb{1}_{\{|x| \geq a\}} \leq f_2(x) = \frac{x^2}{a^2}$



$\Rightarrow \mathbb{E}[\mathbb{1}_{\{|X| \geq a\}}] \leq \mathbb{E}[f_2(X)] \Rightarrow \mathbb{P}\{|X| \geq a\} \leq \frac{\mathbb{E}[X^2]}{a^2}$

it works for any r.v. $X \Rightarrow \mathbb{P}\{|X - \mu| \geq a\} \leq \frac{\mathbb{E}[(X-\mu)^2]}{a^2} = \frac{\mathrm{Var}(X)}{a^2}$