Probability Theory: Review

Anubha Gupta, PhD.
Professor
SBILab, Dept. of ECE,
IIIT-Delhi, India
Contact: anubha@iiitd.ac.in; Lab: http://sbilab.iiitd.edu.in

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY **DELHI**

Probability Theory: Review

(Random Variables)

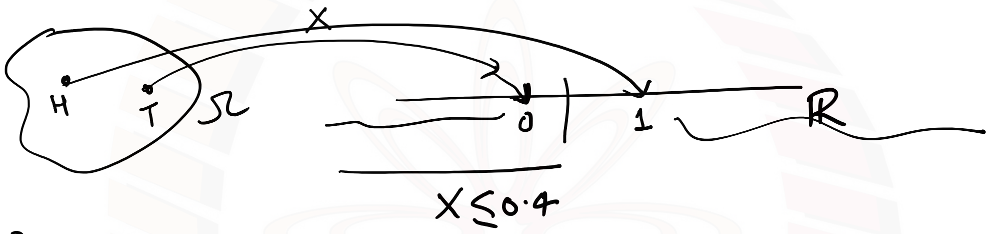
Learning Objectives

- Definition of a random variable
- Types- discrete and continuous
- Probability Distribution Function (PDF) OR Cumulative Distribution Function (CDF)
    - Examples
- Probability Density Function (pdf)
- Examples of famous random variables
- Expectation & Variance
- Joint distribution and density functions
- Central and non-central moments
- Independence, uncorrelatedness, orthogonality of random variables
- Central limit theorem
- Random vector: mean vector, covariance matrix

Reference Text.: Stark, Henry, and John William Woods. "Probability and random processes with applications to signal processing." (2002).

Random Variable

$X: \ \Omega \rightarrow \mathcal{R}$



What is random here?

It is actually not a variable but a function or a mapping whose domain is Ω and range is $\mathcal{R}$.

Consider the experiment: Tossing of a coin; X: {H} → 1; X: {T} → 0

If the coin is fair (unbiased or not biased to generate any specific outcome)

P[X=1]=P[H]=0.5; P[X=0]=P[T]= 0.5

$\underbrace{\{H\}}_{X=1}, \ \underbrace{\{T\}}_{X=0}, \ \underbrace{\{H,T\}}_{\Omega}, \ \phi$

Compute:

(a) $P[X < 0] = 0$ event is: $\{X < 0\} = \phi$

(b) $P[X \leq 0] = \frac{1}{2}$ event is: $\{X \leq 0\}$ ✓ includes $\{T\}$

(c) $P[X \leq 0.4] = \frac{1}{2}$ event is: $\{X \leq 0.4\}$

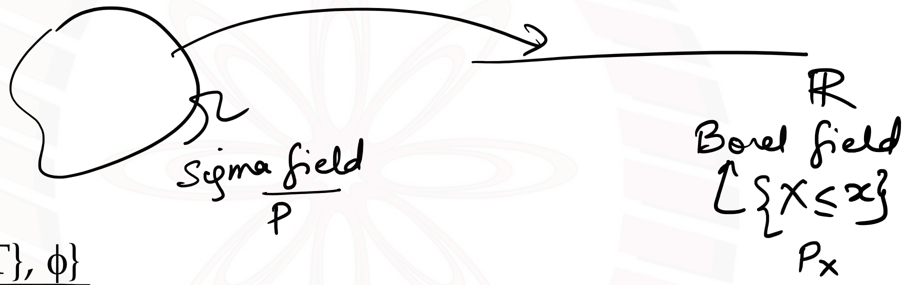
(d) $P[X \leq 1] = 1$ event is: $\{X \leq 1\} = \Omega$

$P[X > 1] = 0$

Or, an event is specified as $\{X \leq \underline{x}\}$

Random Variable

X: $(\Omega, \mathcal{F}, P) \rightarrow (\mathcal{R}, \mathcal{B}, \underline{Px})$



Sigma Field: $\mathcal{F} = \underline{\{\Omega, \{H\}, \{T\}, \phi\}}$
Borel Field: $\mathcal{B}$ = subsets on the real line
P= Probability measure associated with the events in $\mathcal{F}$.
Px= Probability measure associated with the events in $\mathcal{B}$.

$$X^{-1}\underline{\{x \leq 0\}} = \{T\}$$

X is a random variable (r.v.) only if the inverse image under X of all Borel subsets in $\mathcal{R}$ making up the Borel field $\mathcal{B}$ are valid events in Ω.
Or, $X^{-1}(B) = E_B$ = valid event in $\mathcal{F}$
where $B \in \mathcal{B}$

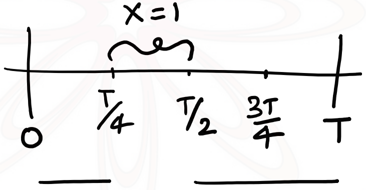
$$X^{-1}\{x < 0\} = \phi$$
$$X^{-1}\{x \leq 1\} = \Omega$$

Let us Try!

A bus arrives at random in [0,T]. Let t be the time of arrival of bus. The bus is equally likely to arrive at any time t between the interval [0,T]. The arrival time is uniformly distributed over [0, T].

Example 1: Define a discrete (binary) random variable X ∈ {0, 1} as below:

$$\text{Define } X(t) = \begin{cases} 1 & \text{if } t \in [T/4, T/2] \\ 0 & \text{otherwise} \end{cases}$$



$$P\{X=1\} = \frac{\frac{T}{2}-\frac{T}{4}}{T} = \frac{1}{4}$$

$$P\{X=0\} = 1 - P\{X=1\} = \frac{3}{4}$$

(a) P[X(t) = 1] = 1/4

(b) P[X(t) = 0] = 3/4

(c) P[X(t) ≤ 2] = 1

Example 2: Define a continuous random variable X ∈ [0, T] as below:

Define X = time of arrival t

$$P[X(t) \le t] = \frac{t}{T}$$

(a) P[X(t) ≤ 0] = 0

(b) P[X(t) ≤ T] = 1

(c) P[X(t) > T] = 0 $= 1 - P[X(t) \le T] = 1 - 1 = 0$

Probability Distribution Function (PDF) or Cumulative Distribution Function (CDF)

This is a point-wise function and is defined as

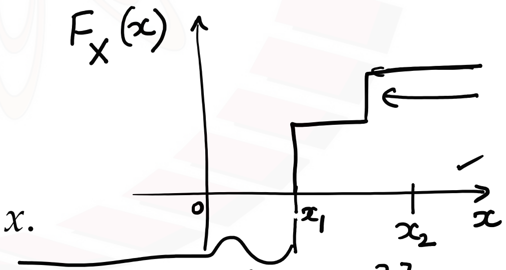
$$F_X(x) = P[X \leq x] = P\,[\{-\infty, x\}]$$

X= notation of the random variable

x = the value taken by the random variable

Properties of a PDF function:

1) $F_X(\infty) = 1$; ✓ $\qquad F_X(-\infty) = 0$



2) If $x_1 \leq x_2$, then $F_X(x_1) \leq F_X(x_2)$
i.e., $F_X(x)$ is a non-decreasing function of x.

$$F_X(\infty) = P[\{-\infty, \infty\}] = 1$$

$$\{-\infty, x_1\} \subseteq \{-\infty, x_2\}$$

$$P[\{-\infty, x_1\}] \leq P[\{-\infty, x_2\}]$$

3) $F_X(x)$ is continuous from the right, i.e.,

$$F_X(x) = \lim_{\varepsilon \to 0} F_X(x + \varepsilon), \text{ where } \varepsilon > 0$$

Examples of CDF

A bus arrives at random in [0,T]. Let t be the time of arrival of bus. The bus is equally likely to arrive at any time t between the interval [0,T]. The arrival time is uniformly distributed over [0, T].

Example 1: Define a discrete (binary) random variable X ∈ {0, 1} as below:

$$\text{Define } X(t) = \begin{cases} 1 & \text{if } t \in [T/4, T/2] \\ 0 & \text{otherwise} \end{cases}$$

Compute and plot the CDF of the above random variable.

$P[X=0] = \frac{3}{4}$

$P[X=1] = \frac{1}{4}$

① $x < 0$

$F_X(x) = P[X \leq x]$

$x = 0.01$

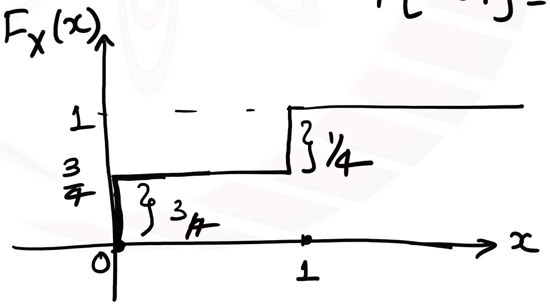
② $F_X(x) = P[X \leq 0.01]$

$= \frac{3}{4}$

③ $x > 1$

$F_X(x) = P[X \leq x]$

let $x=2$ $F_X(2) = P[X \leq 2] = 1$



Examples of CDF

A bus arrives at random in [0,T]. Let t be the time of arrival of bus. The bus is equally likely to arrive at any time t between the interval [0,T]. The arrival time is uniformly distributed over [0, T].

Example 1: Define a discrete (binary) random variable $X \in \{0, 1\}$ as below:

$$\text{Define } X(t) = \begin{cases} 1 & \text{if } t \in [T/4, T/2] \\ 0 & \text{otherwise} \end{cases}$$

Compute and plot the CDF of the above random variable.

Solution: Since the train is equally likely to arrive within 0 to time T,

$P[X = 1] = (T/2-T/4)/T = 1/4$ and $P[X = 0] = 3/4$.

1) Case 1: $-\infty < x < 0$

$F_X(x) = P[X \leq x] = 0$ because no event of X has occurred yet (it can take only two values 0 and 1).

2) Case 2:

$F_X(x) = P[X \leq 1] = 1$ because the ~~train~~ bus would have arrived and hence, $P[X \leq \infty] = P[X \leq 1] = \ldots = 1$;

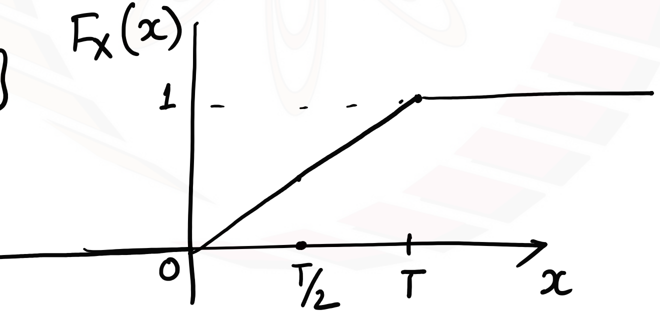
Examples of CDF

A bus arrives at random in [0,T]. Let t be the time of arrival of bus. The bus is equally likely to arrive at any time t between the interval [0,T]. The arrival time is uniformly distributed over [0, T].

Example 2: Define a continuous random variable X as the time of arrival t. Compute and plot the CDF of the above random variable.

$X = t \qquad t \in [0, T]$

$F_X(0) = Pr[X \leq 0]$
$= 0$



$P[\{X \leq t\}] = \frac{t}{T}$

when $t \in [0, T]$

Examples of CDF

A bus arrives at random in [0,T]. Let t be the time of arrival of bus. The bus is equally likely to arrive at any time t between the interval [0,T]. The arrival time is uniformly distributed over [0, T].

Example 2: Define a continuous random variable X as the time of arrival t. Compute and plot the CDF of the above random variable.

Solution:

1) For T $\leq t \leq \infty$

$F_X(t) = P[X \leq t] = 1$ because the ~~train~~ bus would have arrived upto time T and hence,

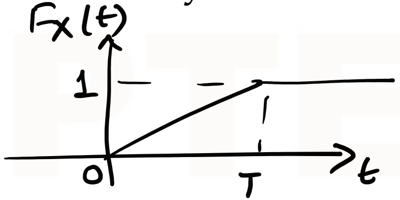
$P[X \leq \infty] = P[X \leq T] = \ldots. = 1;$

2) For $t \leq 0$

$F_X(t) = P[X \leq t] = 0$ because the ~~train~~ bus arrives only after time t ≥ 0

$F_X(- \infty) = 0$

$F_X(t)$



Probability Density Function (pdf)

If $F_X(x)$ is continuous and differentiable, then the probability density function is defined as

$$f_X(x) = \frac{d\, F_X(x)}{dx}$$

The pdf, if it exists, satisfies the following properties:

1. $f_X(x) \geq 0$
2. $F_X(x) = \int_{-\infty}^{x} f_X(\xi) d\xi$

$$F_X(\infty) = 1 \implies \int_{-\infty}^{\infty} f_X(\xi) d\xi = 1$$

Area under pdf = 1

3. $P[x_1 < X \leq x_2] = \int_{x_1}^{x_2} f_X(\xi) d\xi$

Mean and Variance of a Random Variables

Expectation operator ⇒ linear operator

- Mean (μ): E [X] = $\mu_X = \int_{-\infty}^{\infty} x f_X(x) dx$
- Variance (σ^2) :

$$\sigma^2 = E[(X - \mu_X)^2]$$
$$= \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$$

- Standard Deviation (σ):

$$\sigma = \sqrt{E[(X - \mu_X)^2]}$$

$E[X+Y] = E[X] + E[Y]$ → additivity

$E[aX + bY] = aE[X] + bE[Y]$ → linearity

$E[aX] = aE[X]$ → homogeneity

Examples of Continuous Random Variable

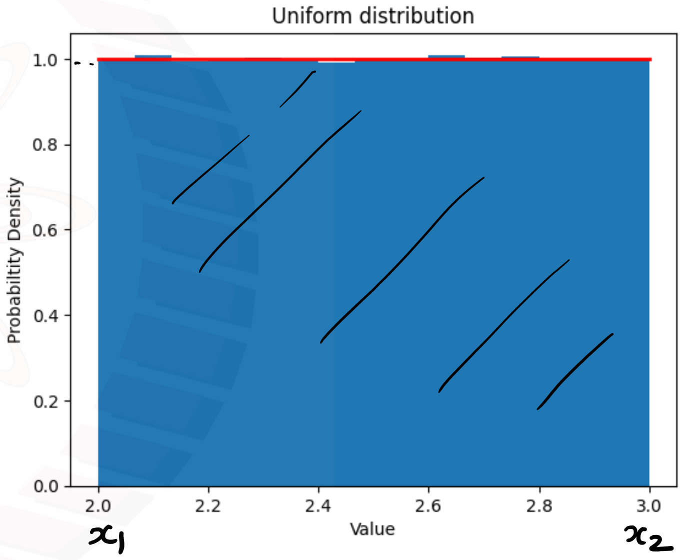
1. Uniformly distributed random variable

The PDF is given by

$$f_X(x) = \begin{cases} \dfrac{1}{x_2 - x_1}, & x_1 < x \le x_2 \\ 0, & otherwise \end{cases}$$

- Mean $= \frac{(x_2+x_1)}{2} = \mu_x$
- Variance $= \frac{(x_2-x_1)^2}{12}$

$$\int_{x_1}^{x_2} \frac{(x-\mu_x)^2}{(x_2-x_1)} dx$$



$x_1 = 2, x_2 = 3$, number of samples= 10 Lakh

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{x_1}^{x_2} \frac{1}{x_2 - x_1} dx = \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} dx$$

$$= \frac{1}{x_2 - x_1} [x]_{x_1}^{x_2} = \frac{1}{x_2 - x_1} (x_2 - x_1) = 1$$

Examples of Continuous Random Variable

2. Gaussian [Normal] distributed random variable

The PDF is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

where

μ : mean of the Gaussian random variable

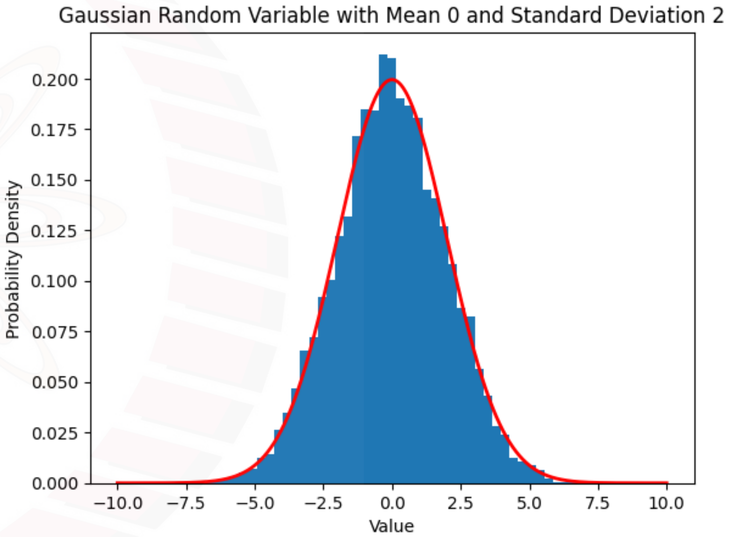
σ : standard deviation of Gaussian random variable

σ^2 : variance of Gaussian random variable

$$\boxed{X : \mathcal{N}(0, \sigma^2)}$$

$$X : \mathcal{N}(\mu, \sigma^2)$$

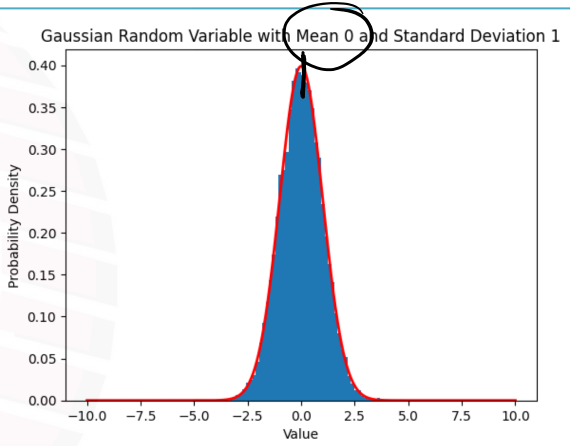
$$\mu = 0 \;, \quad \sigma^2 = 1 \Rightarrow \underline{\text{normal}}$$



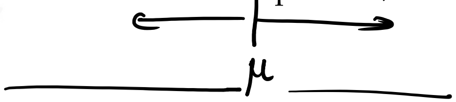
number of samples= 1,000

Examples of Continuous Random Variable

- Since a Gaussian random variable is symmetric about mean, the area under $f_X(x)$:
  - $-\infty < x \leq \mu : \int_{-\infty}^{\mu} f_X(x)dx = \frac{1}{2}$ ✓
  - $\mu < x \leq \infty : \int_{\mu}^{\infty} f_X(x)dx = \frac{1}{2}$ ✓



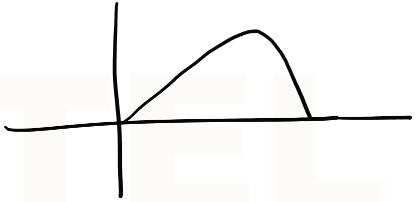
number of samples= 1,000



- Mean (μ): E [X] $= \mu = \int_{-\infty}^{\infty} x f_X(x)dx$
- Mode: Mo[X] $= \frac{d\, f_X(x)}{dx} = 0$ $\Longrightarrow$ $Mo[X] = 0 = \mu$
- Median (x_m) : $\int_{-\infty}^{x_m} f_X(x)dx = \int_{x_m}^{\infty} f_X(x)dx$ $\quad x_m = \mu$

Let us define $Y = \frac{X-\mu}{\sigma} \equiv \mathcal{N}(0,1)$

$$\boxed{f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}}$$



Examples of Continuous Random Variable

68-95-99 Rule

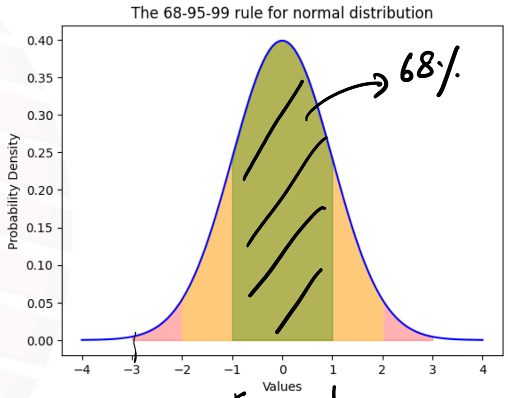
$\mu \pm \sigma \equiv 68\%$

- 68% of the population is within 1 standard deviation of the mean.

$\mu \pm 2\sigma \equiv 95\%$

- 95% of the population is within 2 standard deviation of the mean.
- 99.7% of the population is within 3 standard deviation of the mean.

$\mu \pm 3\sigma \equiv 99.7\%$



$\mu = 0, \sigma^2 = 1$

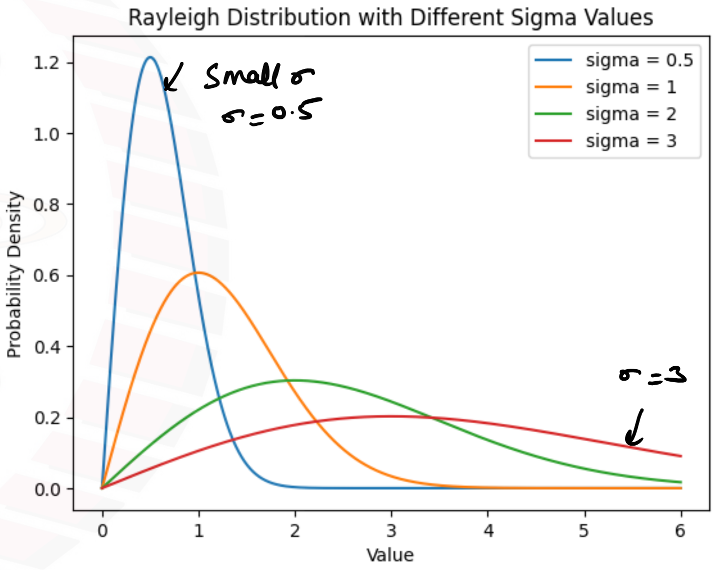
$\mu - \sigma$ μ $\mu + \sigma$

Examples of Continuous Random Variable

3. Rayleigh distributed random variable

The PDF is given by

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, & x > 0 \\ 0, & otherwise \end{cases}$$



where,

scale parameter~~, $\sigma > 0$~~

- Mean $= \sigma\sqrt{\frac{\pi}{2}}$ ✓
- Variance $= \sigma^2\left(2 - \frac{\pi}{2}\right)$ ✓

Examples of Continuous Random Variable

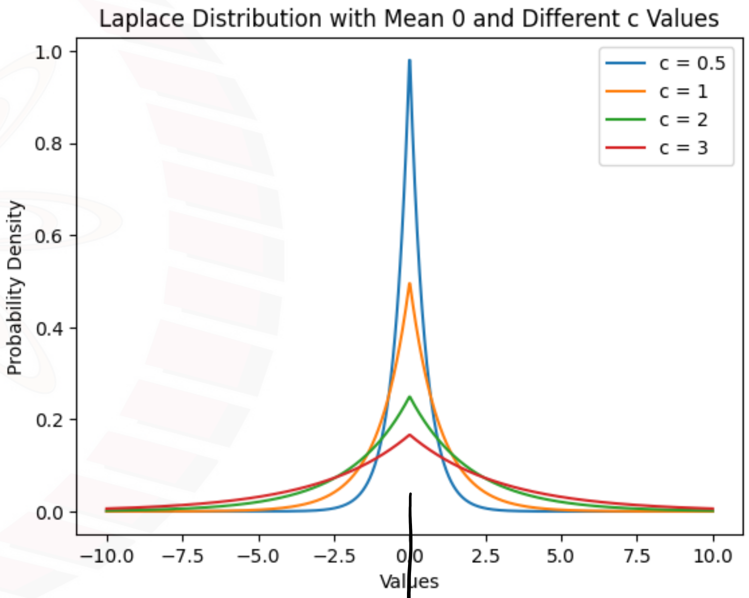
4. Laplacian distributed random variable

The PDF is given by

$$f_X(x) = \frac{c}{2} e^{-c|x|}$$

where,

scale parameter, $c > 0$



Examples of Continuous Random Variable

5. Exponentially distributed random variable

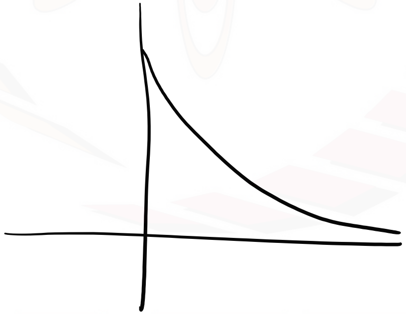
The PDF is given by

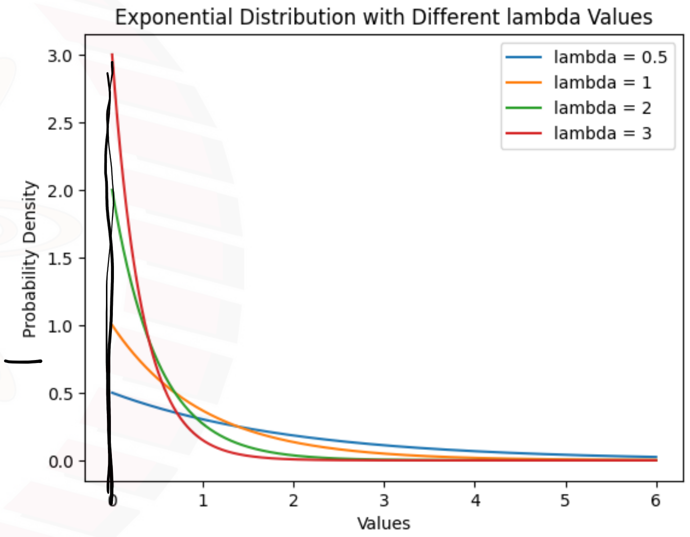
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

where,

scale parameter, $\lambda > 0$

- Mean: $\frac{1}{\lambda}$ ✓
- Variance: $\frac{1}{\lambda^2}$ ✓





Examples of Discrete Random Variable

1. Bernoulli distributed random variable

Prob. mass fn.

$P_X(K)$

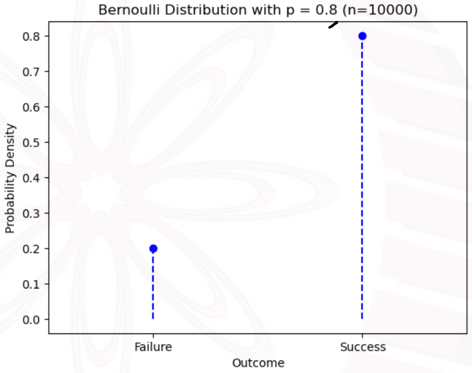
The ~~PDF~~ PMF is given by

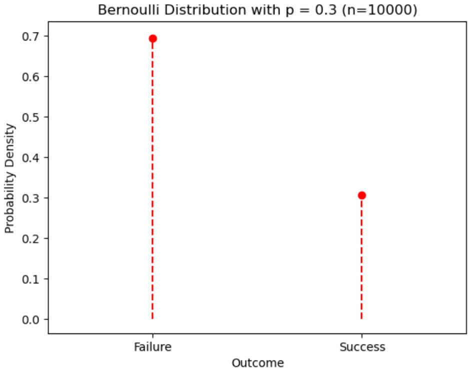
$$\cancel{f(x)}\ P_X(x) = \begin{cases} p, & x = 1 \\ q, & x = 0 \end{cases}$$

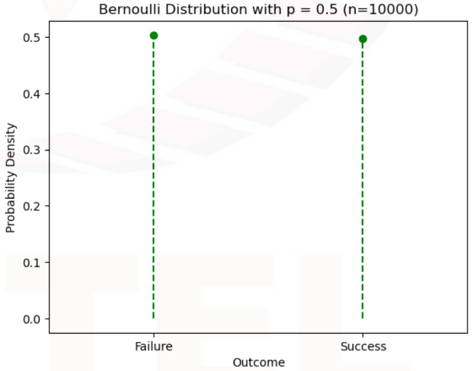
where,

p: probability of success
q: probability of failure

- Mean $= p$ ✓
- Variance $= p(1-p) = pq$







Examples of Discrete Random Variable

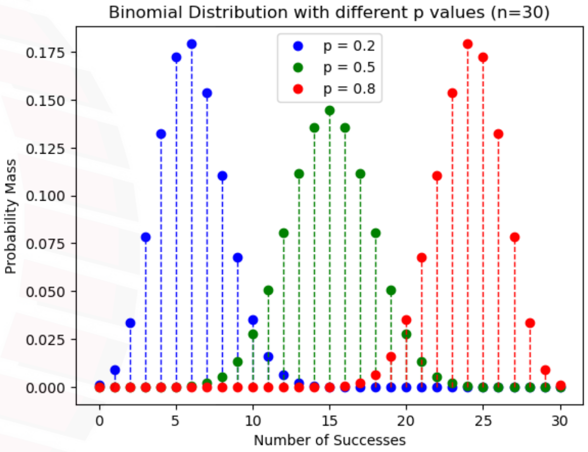
2. Binomial distributed random variable

PMF

The ~~PDF~~ is given by

$$P_X(X=K)$$

$$\cancel{f_X(X=k)} = {}^nC_k p^k q^{n-k}$$



where,

- p: probability of success ✓
- q: probability of failure ✓
- n: number of trials ✓
- k: number of successes ✓

$$\sum_{K=0}^{n} K P_X(X=K) = \mu$$

$$= np$$

- $p + q = 1$ ✓
- Mean $= np$ ✓
- Variance $= npq$ ✓

$$= \sum_{K=0}^{n} (K-\mu)^2 P_X(X=K)$$

Examples of Discrete Random Variable

3. Poisson distributed random variable

PMF

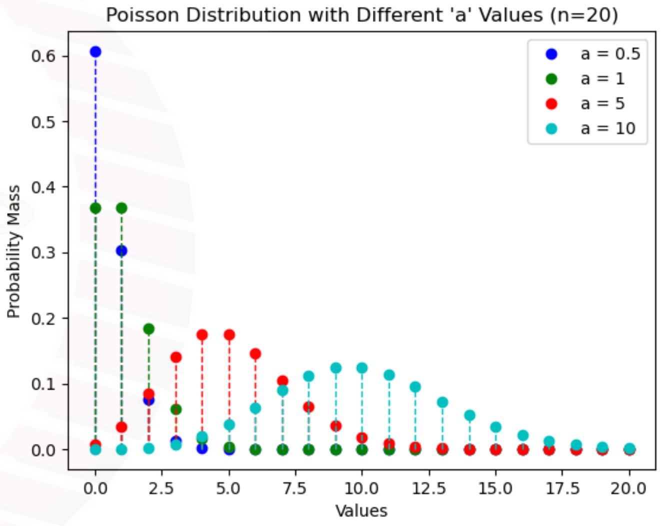
The ~~PDF~~ is given by

$$P_X(X = k) = \frac{e^{-a} a^k}{k!}$$

where,

$$k = 0, 1, \ldots$$
$$a > 0$$

- Mean = a
- Variance = a



Two Random Variables

Let there be two random variables X and Y

Then,

- $f_{XY}(x, y)$: Joint density function ✓
- $F_{XY}(x, y)$: Joint distribution function

$$F_{xy}(\infty, y) = Pr[\underbrace{\{X \leq \infty\}} \cap \{Y \leq y\}]$$
$$= Pr[\{Y \leq y\}]$$
$$= F_Y(y)$$

$$F_{XY}(+\infty, \infty) = 1$$

$F_{XY}(\infty, y) = F_Y(y)$ → [Marginal distribution function of Y]

$F_{XY}(x, \infty) = F_X(x)$ → [Marginal distribution function of X]

$$F_{XY}(-\infty, y) = 0 \checkmark$$
$$F_{XY}(x, -\infty) = 0 \checkmark$$
$$F_{xy}(-\infty, -\infty) = 0$$

- $f_{XY}(x, y) = \dfrac{\partial^2 F_{XY}(x,y)}{\partial x\, \partial y}$

pdf

$$F_{xy}(x, y) = Pr\left[\{X \leq x\} \cap \{Y \leq y\}\right]$$

$x = -\infty$

$$= Pr[\underbrace{\{X \leq -\infty\}} \cap \{Y \leq y\}]$$
$$= Pr[\phi \cap \{Y \leq y\}] = 0$$

Jointly Gaussian Random Variables

Example — Gaussian correlated r.v.

Correlated random variables:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma^2(1-\rho^2)}\{x^2+y^2-2\rho xy\}}$$

Where
ρ : Correlation Coefficient

i.e.,

$$\rho = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sqrt{E[(X-\mu_X)^2]\,E[(X-\mu_Y)^2]}}$$

normalized correlation coeff.

$$\mu_X = 0$$
$$\mu_Y = 0$$
$$\sigma_X^2 = E[(X-\mu_X)^2] = \sigma^2$$
$$\sigma_Y^2 = E[(Y-\mu_Y)^2] = \sigma^2$$

$$\rho = \frac{E[XY]}{\sqrt{\sigma^4}} = \frac{E[XY]}{\sigma^2}$$

if $\rho = 0$

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{-1}{2\sigma^2}(x^2+y^2)}$$

Cases of Joint Random Variables

- Case 1: Uncorrelated random variables ✓
- Case 2: Statistically independent random variables: ✓

Define A : $\{X \le x\}$

B : $\{Y \le y\}$

Proof:

If A and B are two independent events, then

$$P(AB) = P(A \cap B) = P(A)\,P(B)$$

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

$= P(A) \cdot P(B)$

$= F_X(x)\,F_Y(y)$

[S.I.R.V.]

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

$$\iint_{-\infty}^{x,y} f_{XY}(\xi,\eta)\,d\xi\,d\eta = \int_{-\infty}^{x} f_X(\xi)d\xi \int_{-\infty}^{y} f_Y(\eta)d\eta$$

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$ ✓

$E[XY] = 0$

- Case 3: Orthogonal random variables
- Case 4: Independent and identically distributed random variables

Jointly Gaussian Random Variables

Let there be two jointly Gaussian random variables $X \equiv \mathcal{N}(0, \sigma^2)$ and $Y \equiv \mathcal{N}(0, \sigma^2)$.
Then,

- Joint distribution function: $F_{XY}(x, y)$
- Joint density function:

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\left\{\frac{x^2+y^2}{\sigma^2}\right\}} = \left(\frac{1}{2\pi\sigma^2} e^{-\frac{x^2}{2\sigma^2}}\right)\left(\frac{1}{2\pi\sigma^2} e^{-\frac{y^2}{2\sigma^2}}\right)$$

$$= f_X(x) f_Y(y)$$

Thus, X & Y are jointly Gaussian random variables but they are statistically independent (S.I.) random variable.

I.I.D.: Independent and identically distributed random variable: Having identical density function & being statistically independent.

i.i.d

Example

- Central moments: $E[(\cancel{X}-\mu)], E[(\cancel{X}-\mu)^2], \ldots, E[(\cancel{X}-\mu)^i]$ ✓

where

$$E[(\cancel{X}-\mu)^i] = \underline{\int_{-\infty}^{\infty} (x-\mu)^i f_X(x)\, dx}$$

- Non-central moments: $E[X], E[X^2], \ldots, E[X^j]$

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

Central Limit Theorem

The normalized sum of a large number of independent random variables $X_1, X_2, X_3,\ldots,X_n$ each with mean zero and finite variance $\sigma_1{}^2, \sigma_2{}^2,\ldots, \sigma_n{}^2$ tends to be a normal random variable provided that the individual variances $\sigma_k{}^2;\ k=1,2,\ldots,n$ are small compared to s_n^2, where $s_n^2 = \sum_{i=1}^{n} \sigma_i^2$. This condition on variance is also called the Lindeberg condition.

Define

$$Z_n \triangleq \frac{X_1 + X_2 + \cdots + X_n}{s_n}$$

$\approx$

If $\varepsilon > 0$, n is sufficiently large, and each of the σ_k satisfy

$$\boxed{\sigma_k < \varepsilon s_n} \quad \text{for } k=1,2,\ldots\ n,$$

then $Z_n \approx \mathcal{N}(0, 1)$.

Vector Random Variables

$$\underset{\sim}{\boldsymbol{X}} = \begin{bmatrix} X_1 \\ \bullet \\ \bullet \\ \bullet \\ X_n \end{bmatrix} \qquad X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

= n-dimensional random vector with PDF $F_X(\mathbf{x})$

where $\underline{F_X(\mathbf{x})} = P[X_1 \leq x_1; X_2 \leq x_2; \ldots\ldots; X_n \leq x_n]$

$= P[\boldsymbol{X} \leq \mathbf{x}]$

Or, $F_X(\mathbf{x}) = F_{X_1 X_2 \ldots X_n}(\underline{x_1, x_2, \ldots, x_n})$

1) $P[\boldsymbol{X} \leq \infty] = P[X_1 \leq \infty; X_2 \leq \infty; \ldots\ldots; X_n \leq \infty] = 1.$

2) $P[\boldsymbol{X} \leq -\infty] = P[\overline{X_1 \leq -\infty}; X_2 \leq x_2; \ldots\ldots; X_n \leq x_n] = P(\underline{\phi} \cap A_2 \cap \ldots \cap A_n) = P(\phi) = 0$

Or $= P[X_1 \leq x_1; X_2 \leq -\infty; \ldots\ldots; X_n \leq x_n] = P(A_1 \cap \phi \cap \ldots \cap A_n) = P(\phi) = 0$

Or $= P[X_1 \leq x_1; X_2 \leq x_2; \ldots\ldots; X_n \leq -\infty] = P(A_1 \cap A_2 \cap \ldots \cap \phi) = P(\phi) = 0$

Any of the above expressions on the right side will have the probability equal to zero.

Vector Random Variables

$$f_X(\mathbf{x}) = \frac{\partial^n F_X(\mathbf{x})}{\partial \mathbf{x}^n} = \frac{\partial^n F_X(x_1 x_2 \ldots . x_n)}{\partial x_1 \partial x_2 \ldots . \partial x_n}$$

$$F_X(\mathbf{x}) = \iiint_{-\infty}^{\mathbf{x}} f_X(\mathbf{x})\mathbf{dx}$$

Mean of a random vector:

$$E[X] = \mu_X = \begin{bmatrix} \mu_1 \\ \bullet \\ \bullet \\ \bullet \\ \mu_n \end{bmatrix} = \text{mean vector}$$

where

$$\mu_1 = \iiint_{-\infty}^{\infty} x_1 f_X(\mathbf{x}) dx_1 \ldots dx_n$$

Covariance matrix of a random vector

Covariance matrix of a random vector $[X_1, \ldots\ldots, X_n]^T$:

$$\mathbf{K}=E[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T]$$

$$(X-\underline{\mu})(X-\underline{\mu})^T$$

$\uparrow$ outer product

$\underline{\mu}^T \underline{\mu} \rightarrow$ inner product

$$= \begin{bmatrix} (X_1-\mu_1)^2 & \cdots & (X_1-\mu_1)(X_n-\mu_n) \\ \vdots & \ddots & \vdots \\ (X_n-\mu_n)(X_1-\mu_1) & \cdots & (X_n-\mu_n)^2 \end{bmatrix}_{n \times n}$$

$$E\left[\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ X_3-\mu_3 \end{pmatrix} \begin{pmatrix} X_1-\mu_1 & X_2-\mu_2 & X_3-\mu_3 \end{pmatrix}\right]$$

$$= \begin{bmatrix} \sigma_1^2 & K_{12} & \cdots & \cdots & \cdots & K_{1n} \\ K_{21} & \underline{\sigma_2^2} & \cdots & \cdots & \cdots & K_{2n} \\ \vdots & & \ddots & & & \\ \vdots & & & \ddots & & \\ & & \cdots & \cdots & \ddots & \\ K_{n1} & K_{n2} & \cdots & \cdots & \cdots & \underline{\sigma_n^2} \end{bmatrix}$$

$$K_{12} = E[(X_1-\mu_1)(X_2-\mu_2)] = E[(X_2-\mu_2)(X_1-\mu_1)] = K_{21}$$

~~cross cor~~ covariance of X_1 & X_2

For complex random variables, $\mathbf{K}=E[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^H]$, where H denotes the conjugate transpose operation.

Properties of a covariance matrix

1) A covariance matrix **K** is a always a symmetric matrix. ✓

   For real-valued random variables, $\mathbf{K}=\mathbf{K}^T$. ✓
   For complex-valued random variables, **K** is Hermitian, i.e., $\underline{\mathbf{K}=\mathbf{K}^H}$.

   Verify if $K_{21}=K_{12}$? ✓

$\underline{z}^T K \underline{z} \geq 0 \;\; \forall \; \underline{z} \neq 0$

2) A covariance matrix is at least positive semi-definite (p.s.d.). ✓

3) A real symmetric covariance matrix **K** is similar to a diagonal matrix, i.e.,

$$\underline{\mathbf{U}^{-1}\mathbf{K}\mathbf{U} = \Lambda,}$$

   where the columns of U contain the ordered orthogonal unit eigenvectors $\phi_1, \phi_2, \ldots, \phi_n$ of **K**.

Multidimensional Gaussian Law

Given $\boldsymbol{X} = \begin{bmatrix} X_1 \\ \bullet \\ \bullet \\ \bullet \\ X_n \end{bmatrix}$, where Xi's are independent Gaussian random variables $X_i = \mathcal{N}(\mu_i, \sigma_i^2)$.

The probability density function of $\boldsymbol{X}$ is given by

$$f_{\mathbf{X}}(x_1, \dots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

In general, the probability density function of a Gaussian random vector $\boldsymbol{X}$ is given by

$$f_{\mathbf{X}}(x_1, \dots, x_n) = \frac{1}{(\det \mathbf{K})^{1/2}(2\pi)^{n/2}} exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

where $\boldsymbol{\mu}$ = mean vector of $\boldsymbol{X}$ and $\mathbf{K}$ is the covariance matrix of $\boldsymbol{X}$.

Some Theorems

Theorem: Let X be an n-dimensional normal random vector with a positive definite covariance matrix **K** and a mean vector $\boldsymbol{\mu}$. Let **A** be an mxn matrix of rank m.

Then, $\underbrace{Y}_{m\times 1}=\underbrace{\mathbf{A}}_{m\times n}\underbrace{X}_{n\times 1}$ has an m-dimensional normal pdf with positive definite covariance matrix **Q** and a mean vector $\boldsymbol{\beta}$ given by

$$\mathbf{Q}=\mathbf{AKA}^T \equiv \text{covariance matrix of } Y$$

$$\boldsymbol{\beta}=\mathbf{A}\boldsymbol{\mu} \equiv \text{mean vector of } Y$$

Theorem: Let X be an n-dimensional normal random vector with a positive definite covariance matrix **K** and a mean vector $\boldsymbol{\mu}$. Let **U** be the matrix consisting of the eigenvectors of **K** and **Z** be the diagonal matrix consisting of the corresponding eigenvalues at the diagonal of **Z**. Then, **Y** defined as below is a normal random vector with mean vector $\mathbf{A}\boldsymbol{\mu}$ and consists of uncorrelated Gaussian random variables of unit variance:

$$\mathbf{Y}=\mathbf{AX}, \qquad U^T K U = \Lambda$$

where $\mathbf{A}=\mathbf{Z}^{-1/2}\mathbf{U}^T$.

Learning Objectives

- Definition of a random variable
- Types- discrete and continuous
- Probability Distribution Function (PDF) OR Cumulative Distribution Function (CDF)
    - Examples
- Probability Density Function (pdf)
- Examples of famous random variables
- Expectation & Variance
- Joint distribution and density functions
- Central and non-central moments
- Independence, uncorrelatedness, orthogonality of random variables
- Central limit theorem
- Random vector: mean vector, covariance matrix

Reference Text.: Stark, Henry, and John William Woods. "Probability and random processes with applications to signal processing." (2002).