EE325 Short Notes

Aditya Byju

Course Professors: Prof. Bikash Kumar Dey & Prof. D. Manjunath
Ref: Prof's video lectures
Couldn't complete ☹

Probability and Random Processes

September 2021

Contents

Set Theory

- **set** - a collection of well-defined objects
- **Russell's paradox** - the paradox defines the set S of all sets that are not members of themselves, but note that:
  - if S contains itself, then S must be a set that is not a member of itself by the definition of S, which is contradictory
  - if S does not contain itself, then S is one of the sets that is not a member of itself, and is thus contained in S by definition - also a contradiction

  this contradiction is called Russel's paradox
- **De-morgan's laws:**
  - $$\overline{A \cup B} = \overline{A} \cap \overline{B}$$
  - $$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

  these laws apply to any number of sets
- $$A \subseteq B \leftrightarrow \overline{B} \subseteq \overline{A}$$
- **symmetric difference:**
$$\begin{aligned} A \,\Delta\, B &= (A - B) \cup (B - A) \\ &= (A \cup B) - (A \cap B) \end{aligned}$$
- **cartesian product:**
$$A \times B = \{\ (a,b) \mid a \in A,\ b \in B\ \}$$
- A^∞ is the set of sequences of elements from A
- In general, $A^B =$ set of maps of B into A
- $|A|$ denotes the cardinality of a set
- Power set, $P(A) =$ set of subsets of A
$$|P(A)| = 2^{|A|}, \text{where } |A| \text{ is finite}$$
- A relation of A into B is a subset $R \subseteq A \times B$. If $(a,b) \in R$, then we write aRb.
- **equivalence relation** - a relation R of A into A is called an equivalence relation if:
  - It is reflexive i.e. $(a,a) \in R\ \forall\ a \in A$
  - It is symmetric i.e. $\forall a, b \in A$, if $(a,b) \in R$ then $(a,b) \in R$
  - It is transitive i.e. $\forall a, b, c \in A$, if $(a,b),\ (b,c) \in R$ then $(a,c) \in R$
- **equivalence class** - any equivalence relation partitions the set into a disjoint union of subsets, which are called equivalence classes, such that two elements are related iff they are in the same equivalence class
- Given any partition of A, i.e $P_i \subseteq A$; $i \in \mathbb{I}$ such that $\cup_{i\,\in\,\mathbb{I}}\, P_i = A\ \ \&\ \ P_i \cap P_j = \varnothing\ \ \forall i,\ j \in \mathbb{I},\ i \neq j$, one can define an equivalence relation R ($(a,b) \in R$) iff $a,\ b \in P_i$ for some i
- A function or mapping or map f of A into B is a relation such that $\forall a \in A,\ \exists$ a unique $b \in B$ such that $(a,b) \in f$. Here b is called the image of a, and a is called the pre-image of b.

Cardinality

- **one-to-one (injective):** $f : A \to B$ is said to be injective if every element in the range R has a unique pre-image
- **onto (surjective):** $f : A \to B$ is said to be surjective if Range, $(R) = B$, i.e., every element in B has a pre-image in A
- **bijective:** $f : A \to B$ is said to be bijective if it is both injective and surjective
- **cardinality of a set:** is the number of elements in the set
- Comparing cardinality of two sets:
  - two sets A and B are said to be equicardinal if there exists a bijective function from A to $B \rightarrow |A| = |B|$
  - set B has cardinality greater than or equal to set A if there exists a one-to-one function from A to $B \rightarrow |B| \geq |A|$
  - set B has cardinality strictly greater than set A if there exists a one-to-one function from A to B, but no bijective function $\rightarrow |B| > |A|$
- A set is said to be countably infinite if it is equicardinal with $\mathbb{N}$
- A set is said to be countable if it is finite or countably infinite
- Countable union of countable sets is countable
- A set is said to be uncountable if its cardinality is strictly greater than that of $\mathbb{N}$
- **Lemma:** the set of all infinite length binary strings $\{0, 1\}^\infty$ is uncountable. It's proof is given by Cantor's diagonalization argument.
- The sets $[0, 1]$, $\mathbb{R}$, $\mathbb{R} \backslash \mathbb{Q}$ are uncountable
- Dyadic rational is a rational number of the form $\frac{a}{2^b}$

Basics of probability

- **sample space (S or Ω)** - a set of outcomes of a random experiment
- If the sample space is finite or countably infinite, then we say that it is discrete
- **event** - a subset of a sample space
- For a sample space Ω, the set Ω is called sure event and the null set $(\emptyset)$ is called impossible event
- For discrete sample spaces we usually take $P(\Omega)$, the power set of Ω as the set of events
- If $A \cap B = \emptyset$ then A and B are called disjoint events
- Properties of events/subsets:
  - commutativity - $A \cup B = B \cup A$ & $AB = BA$
  - associativity - $(A \cup B) \cup C = A \cup (B \cup C)$ & $(A \cap B) \cap C = A \cap (B \cap C)$
  - distributivity - $(A \cup B)C = AC \cup BC$ & $AB \cup C = (A \cup C)(B \cup C)$
- Let the event space denoted by $\mathcal{F}$. Then, a probability measure on $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to [0, 1]$ satisfying:

- $P(\emptyset) = 0,\ P(\Omega) = 1$
- If $A_1,\ A_2, \ldots$ is a collection of disjoint events, i.e., $A_\text{i} \cup A_\text{j} = \emptyset\ \ \forall\ \text{i} \neq \text{j}$, then $P(\cup_{\text{i}\,=\,1}^{\infty} A_i) = \sum_{\text{i}\,=\,1}^{\infty} P(A_\text{i})$

- The triplet $(\Omega,\ \mathcal{F},\ P)$ is called probability space
- Properties of probability measure P:
  - $P(\overline{A}) = 1 - P(A)$
  - If $B \supseteq A$, then $P(A) \geq P(B)$
  - $P(A \cup B) = P(A) + P(B) - P(A \cup B)$
  - $P(\cup_{\text{i}\,=\,1}^{n} A_\text{i}) = \sum_{\text{j}\,=\,1}^{n} (-1)^{\text{j}-1} (\sum_{s\,\subseteq\,[1:n],\ |s|\,=\,\text{j}} P(\cup_{\text{i}\,\in\,s} A_\text{i}))$
- **Lemma:**
  - let $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ be an increasing sequence of events and let $A = \lim_{\text{i}\to\infty} A_\text{i} = \cup_{\text{i}=1}^{\infty} A_\text{i}$. Then $P(A) = \lim_{\text{i}\to\infty} P(A_\text{i})$.
  - let $B_1 \supseteq B_2 \supseteq B_3 \supseteq \ldots$ be a decreasing sequence of events and let $B = \lim_{\text{i}\to\infty} B_\text{i} = \cap_{\text{i}=1}^{\infty} B_\text{i}$. Then $P(B) = \lim_{\text{i}\to\infty} P(B_\text{i})$.
- **inclusion-exclusion principle (inclusion-exclusion bounds)** - If $E_1,\ E_2,\ E_3, \ldots,\ E_n$ are n events. Then:

$$P_r(\bigcup_{\text{i}\,=\,1}^{n} E_\text{i}) = \sum_{\text{i}\,=\,1}^{n} P_r(E_\text{i}) - \sum_{\{\text{i},\ \text{j}\}\ \subseteq\ [1:n]} P_r(E_\text{i}E_\text{j}) + \sum_{\{\text{i},\ \text{j},\ \text{k}\}\ \subseteq\ [1:n]} P_r(E_\text{i}E_\text{j}E_k) - \ldots (-1)^{n-1} P_r(E_1 \ldots E_n)$$

- **union bound:**

$$P_r(\bigcup_{\text{i}\,=\,1}^{n} E_\text{i}) \leq \sum_{\text{i}\,=\,1}^{n} P_r(E_\text{i})$$

- For random experiments that have equiprobable outcomes the probability of any event E is:

$$P_r(E) = \frac{|E|}{|\Omega|}$$

- Uniform probability measure means all outcomes are equally likely

Conditional probability

- Conditional probability is defined as:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \quad if\ P(B) > 0$$

- For any event B, such that $P(B) > 0$, $P_B(A) = P(A/B)\ \ \forall A \subseteq \Omega$. $P_B(A)$ is a valid probability measure, i.e., it satisfies all the properties of a probability measure.
- **Multiplication rule:**
  - $P(A \cap B) = P(B)\,P(A/B)$
  - $P(A \cap B \cap C) = P(B)\,P(A/B)\,P(C/A \cap B)$
  - in general $P(\cap_{\text{i}\,=\,1}^{n} A_\text{i}) = P(A_1) \prod_{\text{i}\,=\,2}^{n} P(A_\text{i}/A_1 \cap A_2 \cap \ldots \cap A_{\text{i}-1})$

- **Law of total probability:** let A be an event and $\{B_{\text{i}},\ \text{i} \in \mathbb{I}\}$ be countable collection of events that partition Ω $(P(B_i) > 0,\ \forall \text{i})$. Then:

$$P(A) = \sum_{\text{i} \in \mathbb{I}} P(B_{\text{i}})P(A/B_{\text{i}})$$

- **Bayes' theorem:** let A be an event and $\{B_{\text{i}},\ \text{i} \in \mathbb{I}\}$ be countable collection of events that partition Ω $(P(B_i) > 0,\ \forall \text{i})$. Then:

$$P(B_{\text{i}}/A) = \frac{P(B_{\text{i}})P(A/B_{\text{i}})}{P(A)}$$

- **independence** - two events A and B are said to be independent (under probability measure P) if $P(A \cap B) = P(A)P(B)$
- If A, B are independent events, and $P(B) > 0$ then $P(A/B) = P(A)$
- **Lemma:** if A & B are independent events then A & $\overline{B}$ are also independent events
- The events $A_1,\ A_2, \dots,\ A_n$ are said to be independent if for all non-empty subsets $I \in \{1, 2, 3, \dots, n\}$ we have:

$$P(\bigcap_I A_{\text{i}}\ ) = \prod_I A_{\text{i}}$$

- Pairwise independence does not imply independence
- **conditional independence** - A & B are conditionally independent given C $(P(C) > 0)$ if $P((A \cap B)/C) = P(A/C)P(B/C)$
- Conditional independence does not imply independence and vice versa

Borel-Cantelli Lemma

- Let $\{A_n\}$ be a sequence of events over a sample space Ω, and a probability measure P. Then the event $A(i.o.) = \{A_n$ occurs for infinitely many $n\}$ is given by:

$$A(i.o.) = \cap_{n=1}^{\infty} \cup_{m=n}^{\infty} A_m$$

- **First Borel-Cantelli lemma:** let $\{A_n\}$ be a sequence of events over a sample space Ω, and a probability measure P. If:

$$\sum_{n=1}^{\infty} P(A_n) < \infty$$

then $P(A(i.o.)) = 0$

- **Second Borel-Cantelli lemma:** let $\{A_n\}$ be a sequence of events over a sample space Ω, and a probability measure P. If:

$$\sum_{n=1}^{\infty} P(A_n) = \infty$$

then $P(A(i.o.)) = 1$

- **Stirling's formula:** $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$, when $n \to \infty$

Probability measures

- $\nexists$ a uniform probability measure $\mu : P(\Omega) \to [0,1]$ that satisfies the following two conditions:
  - for $0 \le a < b \le 1$, $\mu([a,b]) = b - a$
  - translational invariance: for $A \subseteq [0,1]$, and $\forall x \in [0,1]$, $\mu(A) = \mu(A \oplus x)$, where $A \oplus x = \{a + x \mid a \in A,\ a + x \le 1\} \cup \{a + x - 1 \mid a \in A,\ a + x > 1\}$

  Here, $P(\Omega)$ is called the Vitali set.
- Let $\mathcal{F}$ be a collection of subsets of Ω. $\mathcal{F}$ is said to be a $\sigma-$algebra of Ω if it satisfies:
  - if $A \in \mathcal{F}$ then $\overline{A} \in \mathcal{F}$
  - if $A_\mathrm{i} \in \mathcal{F}$ ($\mathrm{i} \ge 1$) is a countable sequence of sets, then $\cup_\mathrm{i} A_i \in \mathcal{F}$
- A function $\mu : \mathcal{F} \to \mathbb{R} \cup \{-\infty, \infty\}$ is called a measure on $(\Omega, \mathcal{F})$ if it satisfies:
  - $\mu(A) \ge \mu(\emptyset) = 0 \quad \forall A \in \mathcal{F}$
  - if $A_\mathrm{i} \in \mathcal{F}$ is countable sequence of disjoint sets, then $\mu(\cup_\mathrm{i} A_\mathrm{i}) = \sum_\mathrm{i} \mu(A_\mathrm{i})$

  If $\mu(\Omega) = 1$, then μ is called a probability measure
- **Theorem:** let μ be a measure on $(\Omega, \mathcal{F})$:
  - if $A \subseteq B$, then $\mu(A) \le \mu(B)$ [monotone]
  - if $A \subseteq \cup_{m=1}^{\infty} A_m$ than $\mu(A) \le \sum_{m=1}^{\infty} \mu(A_m)$ [subadditive]
  - if$A_i \uparrow A$, i.e., $A_1 \subset A_2 \subset \ldots$ and $\cup_\mathrm{i} A_\mathrm{i} = A$, then $\mu(A_\mathrm{i}) \uparrow \mu(A)$ [continuity from below]
  - if$A_i \downarrow A$, i.e., $A_1 \supset A_2 \supset \ldots$ and $\cap_\mathrm{i} A_\mathrm{i} = A$, then $\mu(A_\mathrm{i}) \downarrow \mu(A)$ [continuity from above]
- Probability mass function is a $p : \Omega \to [0,1]$ such that $\sum_{\omega \in \Omega} p(\omega) = 1$. Now $P(A) = \sum_{\omega \in A} p(\omega)$ is a probability measure.
- For any collection of $\sigma-$fields $\mathcal{F}_\mathrm{i}\ :\ \mathrm{i} \in \mathbb{I}$, $\cup_i \mathcal{F}_i$ is a $\sigma-$field
- In general, $\cap_i \mathcal{F}_i$ is not a $\sigma-$field
- For any $A \subset P(\Omega)$, $\sigma(A) = \cap_{A \subset \mathcal{F}} \mathcal{F}$ (where $\mathcal{F}$ is a $\sigma-$field) is the smallest $\sigma-$field containing A, or the $\sigma-$field generated by A
- **Borel $\sigma-$field and Borel sets:** For $\mathbb{R}$, the $\sigma-$field $\mathcal{R}$ generated by the open sets is called the Borel $\sigma-$field. The sets in it are the Borel sets. Similarly, for $\mathbb{R}^d$ the $\sigma-$field $\mathcal{R}^d$ generated by the open sets is called the Borel $\sigma-$field.
- **Stieltjes measure function (SMF):** let $F : \mathbb{R} \to \mathbb{R}$ be a function such that:
  - F is non-decreasing
  - F is right-continuous, i.e., $\lim_{y \downarrow x} F(y) = F(x)$
  - $F(\infty) = \lim_{x \to \infty} = 1$, and $F(-\infty) = 0$ (For probability measure)
- **Theorem:** for every Stieltjes measure function F there is a unique measure μ on $(\mathbb{R}, \mathcal{R})$ with $\mu(\,(a,b]\,) = F(b) - F(a)$. If F satisfies the third property above, then μ is a probability measure.
- **Lebesgue measure:** this is the natural length measure on $\mathbb{R}$. This corresponds to $F(x) = x$.
- **Generalization of Stieltjes measures:** in $\mathbb{R}^d$, the three SMF conditions do not ensure a measure corresponding to the function. To ensure a measure we need an extra condition:
  - $\Delta_A F \ge 0$ for all rectangles A, where $\Delta_A F = \sum_{v \in V} sgn(v) F(v)$
- **Theorem:** suppose $F : \mathbb{R}^d \to [0,1]$ satisfies the four SMF properties, then $\exists$ a unique measure μ on $(\mathbb{R}^d, \mathcal{R}^d)$ so that $\mu(A) = \Delta_A F$ for all finite rectangles

Random variables

- **Random variables:** given a probability space $(\Omega, \mathcal{F}, P)$, a function $X : \Omega \to \mathbb{R}$ is said to be a random variable if it is measurable, i.e., for every Borel set $B \subset \mathbb{R}$, $X^{-1}(B) = \{\omega | X(\omega) \in B\} \in \mathcal{F}$
- For a discrete probability space with $\mathcal{F} = P(\Omega)$, every map $X : \Omega \to \mathbb{R}$ is measurable and so is a random variable
- For any event $A \in \mathcal{F}$, the indicator function:

$$1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

  is a random variable
- A random variable X induces a probability measure on $(\mathbb{R}, \mathcal{R}) : \mu(B) = P(x \in B) = P(X^{-1}(B))$. This induced probability measure is called the distribution (probability distribution) of X
- The function $F(x) = P(X \leq x)$ is called the cumulative distribution function (CDF) of X
- **Theorem:** any distribution function F has the following properties:
  - F is non-decreasing
  - $\lim_{x \to \infty} F(x) = 1, \;\; \lim_{x \to -\infty} F(x) = 0$
  - F is right continuous, i.e., $\lim_{y \uparrow x} F(y) = F(x)$
  - If $F(x^-) = \lim_{y \uparrow x} F(y)$, then $F(x_-) = P(X < x)$
  - $P(X = x) = F(x) - F(x^-)$
- **Theorem:** if F satisfies the first three points above, then it is the distribution function of some random variable
- **uniform random variable in** $(0,1)$ $(\mathbb{R}, \mathcal{R}, \mu)$**:**
  - $X(\omega) = \omega$
  - $\mu((a.b)) = b - a$
  - $F(x) = x$

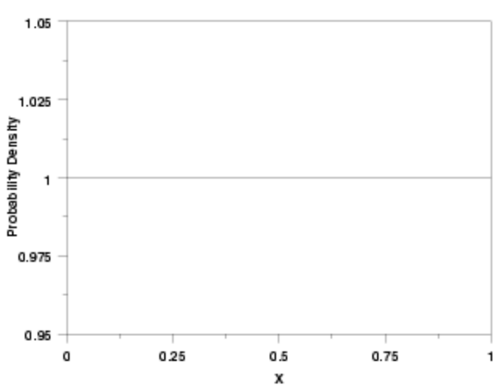

Figure : PDF of uniform r.v. in (0,1)

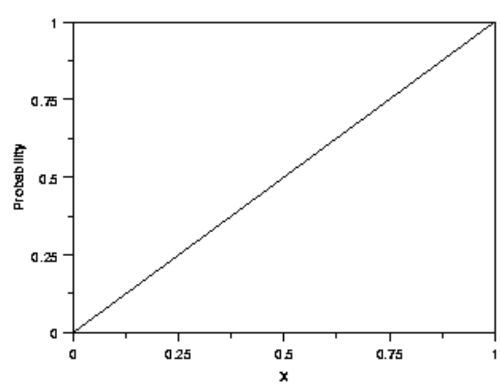

Figure : PDF of uniform r.v. in (0,1)

- Two random variables X and Y are said to be equal in distribution if they have the same distribution function (i.e., they induce the same measure μ on $(\mathbb{R}, \mathcal{R})$). We then write $X \overset{d}{=} Y$ or $X =_d Y$

- For continuous random variables, when $F(x)$ ha the form $F(x) = \int_{-\infty}^{x} f(y)dy$ for some function f, we say that X has a density function f
- Some distributions:
  - **exponential distribution:**

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

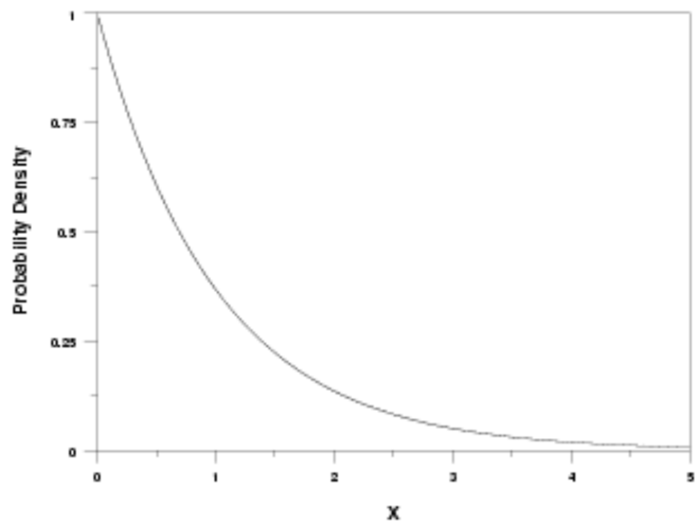

Figure : PDF of exponential distribution

  - **standard normal/Gaussian distribution:**

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

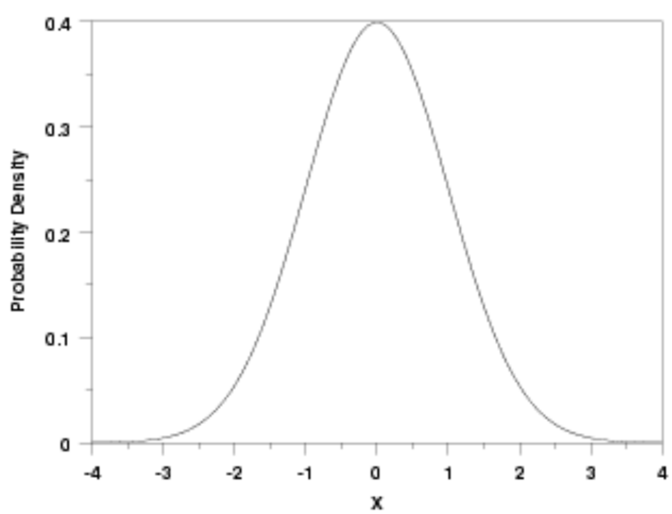

Figure : PDF of standard normal/Gaussian distribution

- Properties of joint CDF:
  - $\lim_{x,y \to \infty} F_{XY}(x,y) = 1$ and $\lim_{x,y \to -\infty} F_{XY}(x,y) = 0$
  - if $x_1 \le x_2,\ y_1 \le y_2$, then $F_{XY}(x_1,y_1) \le F_{XY}(x_2,y_2)$
  - $F_{XY}(x,y)$ is continuous from above, i.e., $\lim_{u,v \downarrow 0} F_{XY}(x+u, y+v) = F_{XY}(x,y)$
  - the marginal CDF's are given by $F_X(x) = \lim_{y \to \infty} F_{XY}(x,y)$ and $F_Y(y) = \lim_{x \to \infty} F_{XY}(x,y)$
- Joint CDF uniquely determines the marginal CDF's but not vice-versa
- Random variables $X_1, X_2, \ldots, X_n$ are said to be independent if $F_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n) \quad \forall x_1, x_2, \ldots, x_n \in \mathbb{R}^n$
- **Def:** joint PMF for two random variables (X, Y) is given by:

$$P_{XY}(x,y) = P(X = x, Y = y) \qquad \forall (x,y) \in C_X \times C_Y$$

  Properties of joint PMF:
  - $\sum_{x \in C_X,\ y \in C_Y} P_{XY}(x,y) = 1$
  - $P_X(x) \quad (= P(X = x)) = \sum_{y \in C_Y} P_{XY}(x,y)$ and $P_Y(y) \quad (= P(Y = y)) = \sum_{x \in C_X} P_{XY}(x,y)$
  - joint PMF uniquely determines joint CDF
- **Def:** conditional PMF of X given Y is defined as:

$$P_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{P_{XY}(x,y)}{P_Y(y)}, \quad when\ P_Y(y) > 0$$

  Properties of conditional PMF:
  - $\sum_{x \in C_X,\ y \in C_Y} P_{XY}(x,y) = 1$
  - $\sum_{y \in C_Y} P_Y(y) P_{X|Y}(x \mid y) = P_X(x)$
- For two discrete random variables X and Y the following statements are equivalent:
  - $X,\ Y$ are independent
  - $P_{XY}(x,y) = P_X(x)P_Y(y)$
  - $P_{X|Y}(x \mid y) = P_X(x), \quad when\ P_Y(y) > 0$
- **Def:** random variables X and Y are jointly continuous if there exists a non-negative function $f_{XY} : \mathbb{R}^2 \to (0, \infty)$ such that:

$$P(X \le x, Y \le y) = F_{XY}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(s,t)\, ds\, dt$$

  f_{XY} is called the joint probability function

Expectation

- Expectation is also known by "mean" and "expected value"
- Expectation for a discrete random variable X with PMF $p(x)$ is:

$$EX\ or\ E[X] = \sum_{x \ \mid\ p(x) > 0} x p(x)$$

- Expectation for a continuous random variable X with density function $f(x)$ is:

$$EX \ or \ E[X] = \int x f(x) dx$$

- Examples of expectation in discrete cases:
  - for uniform random variable, $X \in \{x_1, x_3, \dots, x_M\}$:

$$p_X(x_i) = \frac{1}{M}$$

$$E[X] = \frac{1}{M} \sum_{i=1}^{M} x_i \quad \rightarrow \quad arithmetic \ average$$

  - for indicator random variable, for $A \in \mathcal{F}$, $I_A = \begin{cases} 1 & if \ \ w \in A \\ 0 & if \ w \notin A \end{cases}$:

$$p(1) = P(A) \qquad p(0) = P(\overline{A})$$

$$E[I_A] = P(A)$$

  - for binary random variable $X \in \{a, b\}$, $a, b \in \mathbb{R}$

$$E[X] = ap(a) + bp(b) \quad \rightarrow \quad weighted \ average$$

  - for Poisson distribution, for $\lambda > 0$:

$$PMF \ : \ f_\lambda(k) = P_r(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \dots$$

$$E[X] = \lambda$$

- Examples of expectation in continuous cases:
  - for uniform random variable:

$$density \ function \ \ f_X(x) = \frac{1}{b-a} \ \ for \ a < x < b$$

$$E[X] = \frac{a+b}{2}$$

  - for Gaussian/normal random variable:

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

$$E[X] = \mu$$

- **Expected utility:** consider actions $a_1, a_2, \dots, a_n$. Assume action a_i results in consequences $C_1, C_2, \dots, C_M$ with probabilities $p_{i1}, p_{i2}, \dots, p_{iM}$ respectively. Also assume that C_j has utility (or negative cost) u_j. Then a_i has expected utility:

$$U_i = \sum_{j=1}^{M} p_{ij} u_j$$

  Now, we can choose the action that has the maximum utility.

- If X is a discrete random variable, and g is a real valued function, then:

$$E[g(x)] = \sum_i g(x_i)p(x_i)$$

- **Lemma:** for a non-negative random variable X:

$$E[X] = \int_0^\infty PX > xdx = \int_0^\infty (1 - F_X(x))dx$$

- **Lemma:** for a random variable X:

$$E[X] = \int_0^\infty P(X > x)dx - \int_0^\infty P(X < -x)dx$$

- If X is a continuous random variable, and g is a real valued function, then:

$$E[g(x)] = \int_{-\infty}^\infty g(x)f(x)dx$$

- **Theorem:** suppose $X,\ Y > 0$ or $E[X],\ E[Y] < \infty$ (i.e. $E[X],\ E[Y]$ are well-defined), then:
  - $E(X + Y) = E[X] + E[Y]$
  - $E[aX + b] = aE[X] + b, \quad \forall a, b \in \mathbb{R}$
  - if $X \geq Y$ then $E[X] \geq E[Y]$
- **variance ($var(x)$ or σ^2):** measures the variation or spread of X around $E[X]$:

$$\sigma^2 = E[(X - E[X])^2]$$

- **standard deviation:** $\sqrt{var(x)}$
- The i^{th} moment of X is given by $E[X^i]$. The 1^{st} moment is the mean.
-

$$var(x) = E[X^2] - (E[X])^2$$

- Shifting X to $X + b$ does not change its variance. Scaling X by a scales the variance by a^2.
- **Lemma:** if $X,\ Y$ are independent, then $E[XY] = E[X]E[Y]$
- If $X_1, X_2, \ldots, X_n$ are independent with variance σ^2, then $X = X_1 + X_2 + \ldots + X_n$ has variance $n\sigma^2$
- If $X,\ Y$ are independent, then for any two functions f and g, $f(X),\ g(Y)$ are also independent
- Covariance is defined as:

$$cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

- If X and Y are independent then $cov(X, Y) = 0$
- Properties of covariance:
  - $cov(Y, X) = cov(X, Y)$
  - $cov(X, X) = var(X)$
  - $cov(aX, Y) = a\ cov(X, Y)$
  - $cov(X + Y, Z) = cov(X, Z) + Cov(Y, Z)$

- $cov(\sum_{i=1}^{n} X_i, \ \sum_{j=1}^{m} Y_j) = \sum_i \sum_j cov(X_i, Y_j)$

- $$var(\sum_{i=1}^{n} X_i) = \sum_i var(X_i) + 2\sum_{i<j} cov(X_i, X_j)$$

- $$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{var(x)var(y)}}$$

- **Cauchy-Schwarz inequality:**
$$cov(X,Y)^2 \le var(X)\, var(Y) \quad (i.e., \ -1 \le \rho(X,Y) \le 1)$$

- If $\rho(X,Y) = 0$, then $X, \ Y$ are said to be uncorrelated

- For a random vector $\underline{X} = (X_1, X_2, \ldots, X_n)$, its mean is given by $E[\underline{X}] = (E[X_1], E[X_2], \ldots, E[X_n])$

- **Covariance matrix:**
$$K_{\underline{X}} = E[(\underline{X}^T - E[\underline{X}])(\underline{X}^T - E[\underline{X}])] = (cov(x_i, x_j))_{i,j}$$

- **Def:** let $\psi(y) = E[X|Y = y]$, then the conditional expectation of X given Y is defined as:
$$E[X|Y] = \psi(Y)$$

- **Tower property or law of iterated expectation:**
$$E[X] = E[E[X|Y]]$$

- **Def:** let $\psi(y) = E[g(X,Y)|Y = y]$, then the generalized conditional expectation of $g(X,Y)$ given Y is defined as:
$$E[g(X,Y)|Y] = \psi(Y)$$

- **Generalized Tower property:**
$$E[g(X,Y)] = E[E[g(X,Y)|Y]]$$

- Conditional variance of X given $Y = y$ is the variance of conditional PMF $P_{X|Y}(.|y)$ and is given by:
$$var(X|Y) = E[X^2|Y] - (E[X|Y])^2$$

- **Law of conditional variances:**
$$var(X) = E[var(X|Y)] + var(E[X|Y])$$

- **random sums:** let $\{X_i\}_{i\ge1}$ are independent and identically distributed random variables with $E[X] = \mu_X, \ var(X) = \sigma_X^2$, then its random sum is given by:
$$S_N = \sum_{i=1}^{N} X_i$$
Here $E[S_N] = \mu_X \mu_N$. This is called Wald's identity. Also $var(S_N) = \sigma_X^2 \mu_N + \mu_X^2 \sigma_N^2$

- **Lemma:**
$$E[(X - E[X|Y])^2] \le E[(X - g(Y))^2] \quad \forall g$$

- $E[X]$ is the MMSE estimate of X
- Properties of MMSE:
  - unbiased estimator, i.e., $E[\hat{X}] = E[X]$
  - $E[(X - \hat{X})h(Y)] = 0$
  - the MSE corresponding to $\hat{X}$ is given by $E[var(X|Y)]$
  - from law of total variance, $var(X) = var(\hat{X}) + MSE(\hat{X})$
  - if X, Y are independent, then $\hat{X} = E[X]$

Generating functions

- For a discrete random variable X, taking non-negativve integer values $\{0, 1, 2, \dots\}$, the probability generating function (PGF) is given by:

$$G(s) = E[s^X], \quad s \in \mathbb{C}$$

$$= \sum_{i=0}^{\infty} s^i p(i), \quad X \ has \ PMF \ p(.)$$

- Examples of probability generating functions:
  - for Bernoulli distribution:
  $$G(s) = E[s^X] = 1 - p + ps$$
  - for geometric distribution, $P(X = k) = p(1-p)^{k-1}, \quad k \geq 1$:
  $$G(s) = E[s^X] = \frac{ps}{1 - s(1-p)}$$
  - for Poisson distribution, $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$:
  $$G(s) = E[s^X] = e^{\lambda(s-1)}$$
- Moment generating function (MGF) is given by:

$$M(t) = E[e^{tX}] \quad (= G(e^t)), \quad t \in \mathbb{R}$$

$$= \begin{cases} \sum_i e^{ti} p(i), & X \ - \ discrete \ with \ PMF \ p(.) \\ \int e^{tx} f(x) dx, & X \ - \ continuous \ with \ density \ function \ f(.) \end{cases}$$

- Examples of moment generating functions:
  - for Bernoulli distribution:
  $$M(t) = E[e^{tX}] = 1 - p + pe^t$$
  - for binomial distribution, $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$:
  $$M(t) = E[e^{tX}] = (1 - p + pe^t)^n$$
  - for Poisson distribution, $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$:
  $$M(t) = E[e^{tX}] = e^{\lambda(e^t - 1)}$$

- for normal distribution, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$:

$$M(t) = E[e^{tX}] = e^{t^2/2}$$

- Joint moment generating function of $X_1, X_2, \ldots, X_n$ is given by:

$$M(t_1, t_2, \ldots, t_n) = E[e^{t_1 x_1 + t_2 x_2 + \ldots + t_n X_n}]$$

- The characteristic function is a function from $\mathbb{R} \to \mathbb{C}$ defined as:

$$\phi(t) = E[e^{itX}], \quad i = \sqrt{-1}$$

- For a continuous random variable with the density function $f(x)$:

$$\phi(t) = \int e^{itx} f(x) dx$$

  is the Fourier transform of $f(x)$

- **Theorem:** if X, Y are independent, then:

$$\phi_{X+Y}(t) = \phi_X(t) \phi_Y(t)$$

- **Theorem:** if $a, b \in \mathbb{R}$ and $Y = aX + b$, then:

$$\phi_Y(t) = e^{itb} \phi_X(at)$$

- **Inverse Fourier transform theorem:** if X has density functiion f and characteristic function ϕ, then:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) dt$$

  at every x where $f(x)$ is differentiable

- **Theorem:** X and Y have the same characteristic function iff they have the same distribution function

- **Continuity theorem:** suppose that $F_1, F_2, \ldots$ is a sequence of distribution fnuctions with corresponding characteristic functions $\phi_1, \phi_2, \ldots$. Then:
  - if $F_n \to F$ for some distribution function F with characteristic function ϕ, then $\phi_n(t) \to \phi(t), \quad \forall t$
  - if $\phi(t) = \lim_{n \to \infty} \phi_n(t)$ exists and is continuous at $t = 0$, then ϕ is the characteristic function of some distribution function F, and $F_n(x) \to F(x), \quad \forall x$

- Joint characteristic function:

$$\phi_{X_1, X_2, \ldots, X_n}(t_1, \ldots, t_n) = E[e^{i(t_1 X_1 + t_2 X_2 + \ldots + T_n X_n)}]$$

- **Theorem:** X, Y are independent iff:

$$\phi_{X,Y}(t_1, t_2) = \phi_X(t_1) \phi_Y(t_2)$$

- Quadratic form is defined as:

$$\sum_{1 \le i,j \le n} a_{ij} x_i x_j = \underline{x} A \underline{x}^T$$

- A $n \times n$ matrix A is:

- +ve definite if: $\underline{x}A\underline{x}^T > 0 \quad \forall \underline{x} \neq \underline{0}$
- +ve semidefinite if: $\underline{x}A\underline{x}^T \geq 0 \quad \forall \underline{x} \neq \underline{0}$
- −ve definite if: $\underline{x}A\underline{x}^T < 0 \quad \forall \underline{x} \neq \underline{0}$
- −ve semidefinite if: $\underline{x}A\underline{x}^T \leq 0 \quad \forall \underline{x} \neq \underline{0}$

- A is +ve definite $\iff$ all eigen values $> 0 \implies \det A > 0$
- A is +ve semidefinite $\iff$ all eigen values $\geq 0 \implies \det A \geq 0$
- Linear transformation of jointly Gaussian is jointly Gaussian. That is, for any $n \times m$ matrix B, where $m \leq n$, $\text{rank}(B) = m$, $\underline{Y} = \underline{Z}B + \underline{C}$ is jointly Gaussian.
- **Theorem:** let $Y_1, Y_2, \ldots, Y_n$ be random variables such that:
  - $\sum_{i=1}^{n} a_i Y_i$ is a normal random variable $\forall a_1, a_2, \ldots, a_n \in \mathbb{R}$, and
  - $\det(K_Y) \neq 0$

  then $Y_1, Y_2, \ldots, Y_n$ are jointly Gaussian

More concepts of random variables

- **Markov's inequality:** if X is a non-negative random variable with finite mean, then $P(X \geq a) \leq \frac{E[X]}{a}, \quad for\ a > 0$
- **Chebyshev's inequality:** if X is a random variable with finite mean μ and finite variancce σ^2, then for any $a > 0$:
$$P(\ |X - \mu|\ \geq a\ ) \leq \frac{\sigma^2}{a^2}$$
- **Chernoff bound:** for any random variable X, we have:
$$P(X \geq a) \leq \inf_{s \geq 0} e^{-sa} M_X(s)$$
  where $M_X(s) = E[e^{sX}]$ is the MGF of X