CL205: AI & DS
Probability and Random Variables

Mani Bhushan

Department of Chemical Engineering,
Indian Institute of Technology Bombay,
Mumbai, India-400076

mbhushan@iitb.ac.in

1 Aug 2024

Probability

- Option 1: Intuitive approach to define probabilities of events of interest in terms relative frequencies of occurrence
  - If the event A is observed to occur $N(A)$ times in N trials, then $P(A)$ is defined as

$$P(A) \triangleq \lim_{N \to \infty} \frac{N(A)}{N}$$

    assuming the limit exists
  - Conceptually appealing, but not mathematically rigorous.
- Option 2: Axiomatic approach to probability: basis of modern probability theoretically.
  - We use this.

[Maybeck, 1979]

Sample Space

- Ω: Fundamental sample space containing all possible outcomes of the experiment conducted
- ω: single elementary outcome of the experiment, i.e. $\omega \in \Omega$.
- A: a specific event of interest, a specific set of outcomes of the experiment. $A \subset \Omega$.
- A is said to occur if the observed outcome ω is an element of A, i.e. $\omega \in A$.
- Discrete sample space: eg. coin toss experiments.
- Continuous sample space: eg. values measurement noise can take in a temperature sensor.

Coin Toss Example

Tossing two coins simultaneously.

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$$

$$\omega_1 = \{H, H\} \qquad \omega_2 = \{H, T\}$$

$$\omega_3 = \{T, H\} \qquad \omega_4 = \{T, T\}$$

A: Event that there is at least one head: $\{\omega_1, \omega_2, \omega_3\}$

Before we Assign Probabilities

- Notion of events important
- We may not be interested in each and every elementary outcome but rather in some events.
- Not necessary to assign probability to each elementary outcome.
- Instead assign probability to a collection of events which follows from the events of interest.

σ-algebra Field

σ-algebra $\mathcal{F}$ (σ-algebra field) is a non-empty collection of sets A_i ($A_i \subset \Omega$) such that if $A_i \in \mathcal{F}$ then:

1. $A_i^* \in \mathcal{F}$, where A_i^* is complement of A_i, i.e. $A_i^* = \Omega - A_i$.
2. If $A_1, A_2, ... \in \mathcal{F}$, then their union and intersection are also in $\mathcal{F}$, i.e.

$$\bigcup_i A_i \in \mathcal{F}, \quad \bigcap_i A_i \in \mathcal{F}$$

- Note: Ω, Φ always part of field $\mathcal{F}$

Example 1

- Experiment of rolling a die once.
- Sample space:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- σ-algebra Field can be defined in multiple ways.
- Case 1: Field $\mathcal{F}$ can be taken as set of all subsets of Ω (i.e. power set of Ω):

$$\mathcal{F} = \{\Phi, \{1\}, \{2\}, ..., \{1, 2\}, \{1, 3\}, ...\}$$

  - Number of elements of $\mathcal{F}$: $2^6 = 64$.

Example 1 (2)

- Case 2: If we are interested in betting on only odd and even events, then:

$$\mathcal{F} = \{\Phi, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\}$$

- Is this a field? Ans: Yes

Example 1 (3)

- Another case:

$$\mathcal{F} = \{\Phi, \{2\}, \{1,3,5\}, \{2,4,6\}, \Omega\}$$

- Is this a field? : No
  since $\{2\} \cup \{1,3,5\} = \{1,2,3,5\} \notin \mathcal{F}$
  Note: $\{1,2,3,5\} \subset \Omega$ yet it is not an element of $\mathcal{F}$ even though Ω is an element of $\mathcal{F}$.

Probability Function

The probability function (or probability measure) $P(.)$ is a real scalar-valued function defined on a σ-algebra $\mathcal{F}$ that assigns a value, $P(A_i)$, to each $A_i \in \mathcal{F}$ such that:

1. $P(A_i) \geq 0 \;\; \forall A_i \in \mathcal{F}$
2. $P(\Omega) = 1$
3. If $A_1, A_2, ..., A_N$ are disjoint or mutually exclusive then

$$P\left(\bigcup_{i=1}^{N} A_i\right) = \sum_{i=1}^{N} P(A_i)$$

   for all finite and countably infinite N.

Consistent with intuition of probability gained through the concept of relative frequency of occurrence. **HW**: To check

Probability Function: Intuition

Using the above properties,

- Q Show that $P(A_i) \leq 1$
  Answer: $\Omega = A_i \bigcup A_i^*$
  $P(\Omega) = 1 = P(A_i) + P(A_i^*)$
  $\Rightarrow \; P(A_i) \leq 1$
- Q Show that if $A_1 \subset A_2$ then $P(A_1) \leq P(A_2)$
  Answer: $A_2 = A_1 \bigcup (A_1^* \bigcap A_2)$
  $\Rightarrow P(A_2) = P(A_1) +$ a non-negative number

Probability Space

Probability Space: Defined by triplet $(\Omega, \mathcal{F}, P)$ of the sample space, the underlying σ-algebra, and the probability function.

Example 1: Rolling of a die

- Sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Case 1: σ-algebra $\mathcal{F}_1 \equiv$ power set of Ω.
  - Consider sets $A_i = \{i\}$ for $i = 1, 2, ..., 6$.
  - If we set $P_1(A_i) = 1/6$ for $i = 1, 2, ..., 6$, then we can find probability of any event in $\mathcal{F}$.
- Triplet $(\Omega, \mathcal{F}_1, P_1)$ forms a probability space.

Example continued

- Case 2: We are interested in only odd and even events
  - σ-algebra: $\mathcal{F}_2 = \{\Phi, \{1,3,5\}, \{2,4,6\}, \Omega\}$
  - Define: $P_2(\{1,3,5\}) = p$ with $1 \geq p \geq 0 \implies$

$$P_2(\{2,4,6\}) = 1 - p, \quad P_2(\Phi) = 0, \quad P_2(\Omega) = 1$$

  - Triplet $(\Omega, \mathcal{F}_2, P_2)$ also forms a probability space

Comments

- Definition of a probability space for a given experiment is NOT unique.
- Triplet $(\Omega, \mathcal{F}, P)$ must be specified

Computing Probabilities

- Derive probability of a union of non-disjoint events
- Let A, B be two events $\in \mathcal{F}$ that are NOT disjoint. Then,

$$\begin{aligned} A &= (A \cap B) \cup (A \cap B^*) \\ \text{Then, } P(A) &= P(A \cap B) + P(A \cap B^*) \quad \text{Why?} \end{aligned} \tag{1}$$

Similarly,

$$\begin{aligned} A \cup B &= ((A \cup B) \cap B) \cup ((A \cup B) \cap B^*) \\ &= (B) \cup (A \cap B^*) \\ \text{Then, } P(A \cup B) &= P(B) + P(A \cap B^*) \end{aligned}$$

Eliminating $P(A \cap B^*)$ from above using Eqn. (1):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Example (Ross, 2009)

A total of 28 percent of American males smoke cigarettes, 7 percent smoke cigars, and 5 percent smoke both cigars and cigarettes. What percentage of males smoke neither cigars nor cigarettes?

- Event A: a randomly chosen male is a cigarette smoker
- Event B: a randomly chosen male a cigar smoker.

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.28 + 0.07 - 0.05 = 0.3 \end{aligned}$$

- Thus the probability that the person is not a smoker is .7, implying that 70 percent of American males smoke neither cigarettes nor cigars.

Probability Fundamentals: What Else?

- Once probability space $(\Omega, \mathcal{F}, P)$ is properly defined for a given problem, theoretically we are done.
- However, sample space need not consist of elements that are numbers.
  - Example: A die with 6 faces painted with 6 different colours; Candidates appearing in an election in a constituency.
  - How to perform numerical computations involving such sample spaces
- It will help to have some mapping that enables us to deal with real numbers

Random Variable (RV)

A scalar random variable $\boldsymbol{x}(\omega)$

- Is a real-valued point function,
- Assigns a real scalar value to each point ω in Ω, denoted as $\boldsymbol{x}(\omega) = x$, such that
  every set A of the form $A = \{\omega : \boldsymbol{x}(\omega) \leq \xi\}$ for any value ξ on the real line ($\xi \in \mathbb{R}$) is an element of the σ-algebra $\mathcal{F}$. Hence, its probability of occurrence can be defined through probability function P.
- Important: Random variable is actually a function!!

$\boldsymbol{x}(\omega)$: Random variable (function),
x : a realization of the random variable (i.e. the value that this function assumes for a particular ω).

Random Variable: Example

Experiment: rolling a die once

- Sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Events of interest: A_1: a 1 or 2 was thrown, and A_2: a 3 was thrown.
- σ-algebra $\mathcal{F}$ has the following elements:

$$\Phi,\ \Omega = \{1,2,3,4,5,6\}, A_1 = \{1,2\},\ A_2 = \{3\},$$
$$A_1^* = \{3,4,5,6\},\ A_2^* = \{1,2,4,5,6\},\ A_1 \cup A_2 = \{1,2,3\},$$
$$\{A_1 \cup A_2\}^* = \{4,5,6\}$$

- Define RV: $\boldsymbol{x}(\omega_i) = i$.
- Consider set $A = \{\omega : \boldsymbol{x}(\omega) \leq \xi\}$:
  - $\xi = 1 \implies A = \{1\} \notin \mathcal{F}$.
- Random variable choice incorrect

Example (Cont.)

- Consider $x(\omega)$ to be:

$$x(\omega) = \begin{cases} 1, & \omega = 1, 2 \\ 2, & \omega = 3 \\ 3, & \omega = 4, 5, 6 \end{cases}$$

- Consider set $A = \{\omega : x(\omega) \leq \xi\}$:
  - $\xi = 1 \implies A = \{1, 2\} \in \mathcal{F}$.
  - $\xi = 2 \implies A = \{1, 2, 3\} \in \mathcal{F}$.
  - $\xi = 3 \implies A = \{1, 2, 3, 4, 5, 6\} \in \mathcal{F}$.
  - $\xi < 1 \implies A = \Phi \in \mathcal{F}$.
  - $\xi > 3 \implies A = \Omega \in \mathcal{F}$.
- A valid, though not unique, random variable

Advantage

- Once we define a random variable $\boldsymbol{x}(\omega)$ on an original sample space, say Ω, we can start working with the generic sample space $\Omega_R \equiv \mathbb{R}$.
- Original σ-algebra $\mathcal{F}$ is replaced by generic σ-algebra $\mathcal{F}_R$
  - Consisting of all sub-intervals of $\mathbb{R}$
  - Is called a Borel field.

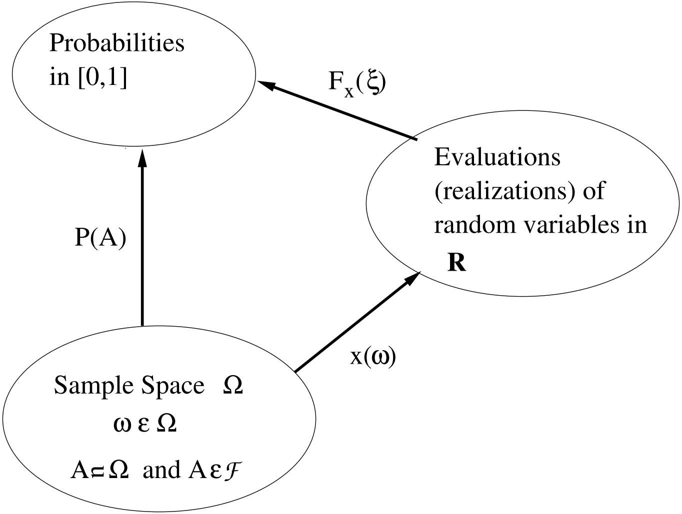
Advantage (Cont.)

- A generic probability function, $P(.)$ is defined for all events of the form $A = (-\infty, \xi]$, where ξ is a real number

$$F_{\boldsymbol{x}}(\xi) = P(\omega : \ \boldsymbol{x}(\omega) \leq \xi)$$

- Probability would have been defined through probability function in original sample space.
- $F_{\boldsymbol{x}}(\xi)$: Probability distribution function, or cumulative distribution function (cdf)
- $F_{\boldsymbol{x}}(\xi)$: $\mathbb{R} \to [0, 1]$.
- $F_{\boldsymbol{x}}(\xi)$ has all the information about probabilities
- From now on, we work with Probability Space: $(\mathbb{R}, \mathcal{F}_R, F_{\boldsymbol{x}}(\xi))$.

Schematic Illustration



[Maybeck, 1979]

Example: Coin Toss

- Consider the coin toss experiment with triplet:

$$\Omega = \{H, T\}; \mathcal{F} = \{\Phi, H, T, \Omega\}$$
$$P(H) = p, \; p(T) = q, \quad p + q = 1, \; 0 < p, q < 1$$

- Define a discrete random variable x:

$$x(H) = 1, \quad x(T) = 0$$

- New sample space is $\mathbb{R}$ and associated Borel field is $\mathcal{F}_R$.
- New distribution function:

$$F(\xi) = \begin{cases} 0 & \text{for } -\infty < \xi < 0 \\ q & \text{for } 0 \leq \xi < 1 \\ 1 & \text{for } 1 \leq \xi < \infty \end{cases}$$

- Sketch the distribution function!

Example: Phone Calls (Jazwinski, 1970)

- Telephone calls occur at random during the time interval $[0, T]$
- ω is the time at which phone call occurs
- $\Omega = [0, T]$
- Let $P(t_1 \leq \omega \leq t_2) = \frac{t_2 - t_1}{T}, \ t_1, t_2 \in [0, T]$
- Define RV as $\mathbf{x}(\omega) = \omega$.
- Compute the probability distribution function:
  - For $\xi \geq T$, $\{\omega : \mathbf{x}(\omega) \leq \xi\} = \Omega \Rightarrow F_{\mathbf{x}}(\xi) = 1$.
  - For $0 \leq \xi \leq T$, $\{\omega : \mathbf{x}(\omega) \leq \xi\} = \{0 \leq \omega \leq \xi\} \Rightarrow F_{\mathbf{x}}(\xi) = \frac{\xi}{T}$.
  - For $\xi < 0$, $\{\omega : \mathbf{x}(\omega) \leq \xi\} = \Phi \Rightarrow F_{\mathbf{x}}(\xi) = 0$.

Phone Call Example (Cont.)

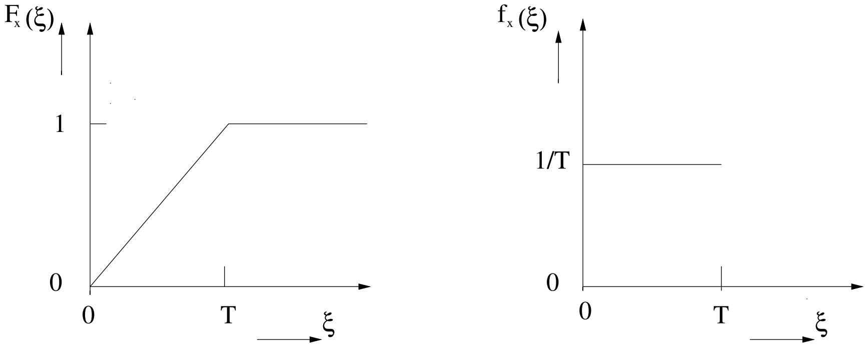


Figure: Distribution and density function for phone call example

Properties of Probability Distribution Function

Notation:

$$F_{\boldsymbol{x}}(\xi^+) \triangleq \lim_{\epsilon \to 0} \quad F_{\boldsymbol{x}}(\xi + \epsilon); \quad F_{\boldsymbol{x}}(\xi^-) \triangleq \lim_{\epsilon \to 0} \quad F_{\boldsymbol{x}}(\xi - \epsilon);$$

for $\epsilon > 0$. [Ignoring subscript of F from now on]

- Property 1: $F(+\infty) = 1$, and $F(-\infty) = 0$.
- Property 2: $F(\xi)$ is a non-decreasing function of ξ, i.e. if $\xi_1 < \xi_2 \implies F(\xi_1) \leq F(\xi_2)$.
- Property 3: If $F(\xi_0) = 0$ then $F(\xi) = 0 \; \forall \xi < \xi_0$.
- Property 4: Function $F(\xi)$ is continuous from the right, i.e. $F(\xi^+) = F(\xi)$.
- Property 5: $P(\boldsymbol{x}(\omega) = \xi) = F(\xi) - F(\xi^-)$.
- Property 6: $P(\xi_1 \leq \boldsymbol{x}(\omega) \leq \xi_2) = F(\xi_2) - F(\xi_1^-)$

Types of Random Variables

- *Continuous Random variable*: If the distribution function $F_{\boldsymbol{x}}(\xi)$ is continuous.
  Example: Phone call example
- *Discrete Random variable*: If the distribution function $F_{\boldsymbol{x}}(\xi)$ is constant except for a finite number of jump discontinuities i.e. it is piecewise constant, step type.
  Example: Coin-toss
- Mixed type random variable is also possible.

Probability of Other Events in $\mathcal{F}_R$

- Probability of event $\{\boldsymbol{x} > \xi\}$ i.e. set (ξ, ∞) is

$$P((\xi, \infty)) = 1 - P((-\infty, \xi]) = 1 - F(\xi) : \quad \text{why?}$$

- Probability of event $\{\xi_1 < x \leq \xi_2\}$ i.e. set $(\xi_1, \xi_2]$:

$$\begin{aligned}
(-\infty, \xi_2] &= (-\infty, \xi_1] \cup (\xi_1, \xi_2] \\
\implies \quad F(\xi_2) &= F(\xi_1) + P((\xi_1, \xi_2]) : \quad \text{why?} \\
\implies P((\xi_1, \xi_2]) &= F(\xi_2) - F(\xi_1)
\end{aligned}$$

Example: Rolling a Die

- Consider experiment of rolling a die once with 6 faces painted with 6 different colours.
- Sample space $\Omega = \{C_1, C_2, C_3, C_4, C_5, C_6\}$.
- Let σ-algebra $\mathcal{F}$ contain power set of Ω.

Rolling a Die (Contd.)

- Define a random variable $x(C_i) = 10i$.
- New sample space is $\mathbb{R}$.
- An event, say A in the σ-algebra $\mathcal{F}$ can now be described using an event $\tilde{A}$ in $\mathcal{F}_R$.
- Events in $\mathcal{F}_R$ are intervals of the form:

$$\tilde{A} = \{x \le \xi\} = (-\infty, \xi], \text{ where } \xi \in \mathbb{R}, \textit{or}$$
$$\tilde{A} = \{\xi_1 \le x \le \xi_2\} = [\xi_1, \xi_2], \text{ where } \xi_1, \xi_2 \in \mathbb{R}, \xi_1 \le \xi_2$$

etc.

Rolling a Die (Contd. 2)

- Event $\{x \leq 35\}$ in $\mathcal{F}_R \equiv$ Event $\{C_1, C_2, C_3\}$ in $\mathcal{F}$ since $x(C_i) \leq 35$ only if $i = 1, 2, 3$.
- Event $\{17.5 \leq x \leq 46.8\}$ in $\mathcal{F}_R \equiv$ Event $\{C_2, C_3, C_4\}$ in $\mathcal{F}$ since $17.5 \leq x(C_i) \leq 46.8$ only for i=2,3,4.
- Event $\{x \leq 6\}$ in $\mathcal{F}_R \equiv$ Event $\{\Phi\}$ in $\mathcal{F}$ since there is no outcome s.t. $x(C_i) \leq 6$.

Rolling a Die: Probability Measure

- A probability measure on $\mathcal{F}$ is: $P(C_i) = 1/6$ for $i = 1, 2, ..., 6$.
- An equivalent probability measured on $\mathcal{F}_R$ is:

$$F(\xi) = P((-\infty, \xi]) \equiv (1/6) \times \{\text{No. of pts. 10,20,...,60 which } (-\infty, \xi] \text{ contains}\}$$

Probability Mass Function (PMF)

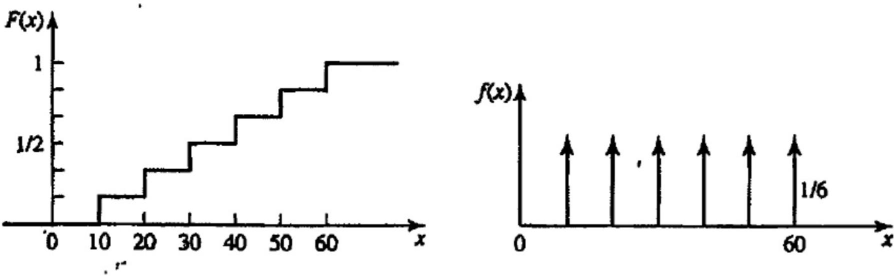


Figure: Probability Distribution and Probability Mass Function

PMF

- PMF (Probability Mass Function) of a discrete RV $\boldsymbol{x}$ is a function $f_{\boldsymbol{x}}(.) : \mathbb{R} \rightarrow [0, 1]$ defined by $f_{\boldsymbol{x}}(a) = P(\boldsymbol{x} = a)$ for $-\infty < a < \infty$.
- If $\boldsymbol{x}$ takes values $a_1, a_2, ...$ then we have
  $f_{\boldsymbol{x}}(a_i) = P(\boldsymbol{x} = a_i) = p_i > 0$,
  and $p_1 + p_2 + ... = 1$ with $0 < p_1, p_2, ... < 1$.
  and $f_{\boldsymbol{x}}(\xi) = P(\boldsymbol{x} = \xi) = 0$ for all other $\xi \in \mathbb{R}$.
- The probability/cumulative distribution function of a discrete RV, $\boldsymbol{x}$, is related to the probability mass function of $\boldsymbol{x}$ as follows: $F(\xi) = P(\boldsymbol{x} \leq \xi) = \sum_{i, a_i \leq \xi} f_{\boldsymbol{x}}(a_i)$
- Write the PMF for the die example

Example: Data Transmission

There is a chance that a bit transmitted through a digital transmission channel is received in error.

- Random variable $x =$ number of bits in error in the next four bits transmitted.
- Original sample space $\Omega = \{0, 1, 2, 3, 4\}$.
- Original sigma-algebra = Power set of Ω.
- Probabilities defined with reference to Ω:

$$P(\omega_1 = 0) = 0.6561, \quad P(\omega_2 = 1) = 0.2916, \quad P(\omega_3 = 2) = 0.0486$$
$$P(\omega_4 = 3) = 0.0036, \quad P(\omega_5 = 4) = 0.0001$$

Example (Cont.)

- Define discrete random variable $x(\omega_i) = i - 1$ for $i = 1, 2, 3, 4, 5$.
- New sample space$= \mathbb{R}$.
- Probability mass function:

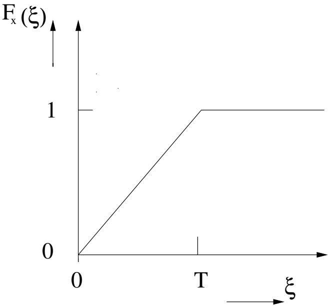
$$f_x(0) = 0.6561, \quad f_x(1) = 0.2916, \quad f_x(2) = 0.0486,$$
$$f_x(3) = 0.0036, \quad f_x(4) = 0.0001$$

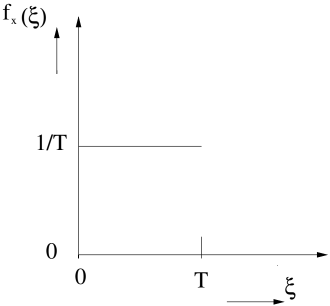
- Probability/Cumulative Distribution Function:

$$F_x(\xi) = \begin{cases} 0 & \text{for } -\infty < \xi < 0, \\ 0.6561 & \text{for } 0 \leq \xi < 1, \\ 0.9477 & \text{for } 1 \leq \xi < 2, \\ 0.9963 & \text{for } 2 \leq \xi < 3, \\ 0.9999 & \text{for } 3 \leq \xi < 4, \\ 1 & \text{for } 4 \leq \xi < \infty \end{cases}$$

Continuous Random Variable: Phone Call Example Revisited

- Telephone calls occur at random during the time interval $[0, T]$
- ω is the time at which phone call occurs
- $\Omega = [0, T]$
- Let $P(t_1 \leq \omega \leq t_2) = \frac{t_2 - t_1}{T}$, $t_1, t_2 \in [0, T]$
- Define RV as $\mathbf{x}(\omega) = \omega$.





Continuous Random Variable

- Distribution function is continuous
- Time of phone call: A continuous random variable
- For a continuous random variable, a scalar valued function, $f_{\mathbf{x}}(.)$ exists such that
$$F_{\mathbf{x}}(a) = \int_{-\infty}^{a} f_{\mathbf{x}}(\xi) d\xi$$
holds for all values of $a \in \mathbb{R}$
- $f_{\mathbf{x}}(\xi)$: called probability density function of $\mathbf{x}$.
- In particular, for any $a, b \in \mathbb{R}$ with $a \leq b$:
$$P(a \leq \mathbf{x} \leq b) = F_{\mathbf{x}}(b) - F_{\mathbf{x}}(a) = \int_{a}^{b} f_{\mathbf{x}}(\xi) d\xi$$
or in simplified notation,
$$P(a \leq \mathbf{x} \leq b) = F(b) - F(a) = \int_{a}^{b} f(\xi) d\xi$$

Continuous Random Variable (Cont.)

- Probability density function is derivative of probability distribution function whenever the derivative exists:

$$f_{\mathbf{x}}(\xi) = \frac{dF_{\mathbf{x}}(\xi)}{d\xi}$$

- Since derivative may not exist at all points, probability density function may have discontinuities at isolated points on the real line: This does not cause any problem.
- Properties of $f_{\mathbf{x}}(\xi)$:

$$f_{\mathbf{x}}(\xi) \geq 0$$

$$\int_{-\infty}^{\infty} f_{\mathbf{x}}(\xi) d\xi = 1$$

$$P(a \leq \mathbf{x} \leq b) = \int_{a}^{b} f_{\mathbf{x}}(\xi) d\xi, \quad \forall a, b \in \mathbb{R} \text{ with } a \leq b$$

Recall Phone Call Example with slight modification

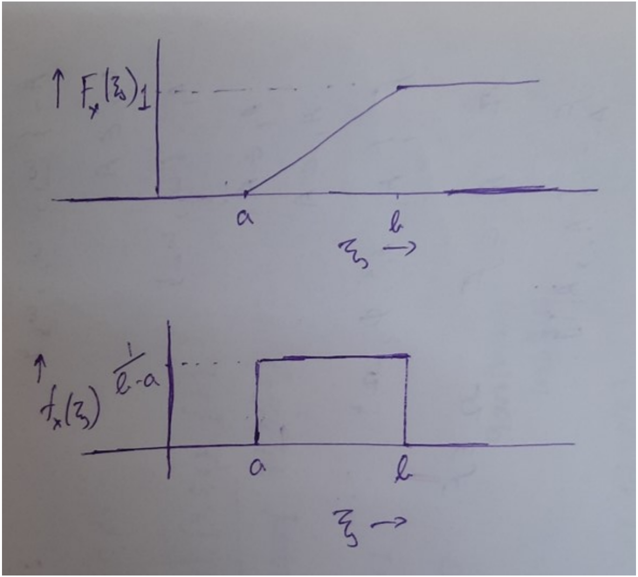
- Telephone calls occur at random during the time interval $[a, b]$
- ω is the time at which phone call occurs
- $\Omega = [a, b]$
- Let $P(t_1 \leq \omega \leq t_2) = \frac{t_2 - t_1}{b - a},\ t_1, t_2 \in [a, b]$
- Define RV as $\mathbf{x}(\omega) = \omega$.

Probability Distribution and Density Functions

$$F_{\mathsf{x}}(\xi) = \begin{cases} 0, & \xi < a \\ \frac{\xi - a}{b - a} & a \leq \xi \leq b \\ 1, & b < \xi \end{cases}$$

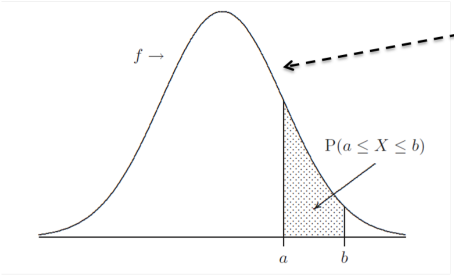
$$f_{\mathsf{x}}(\xi) = \begin{cases} 0, & \xi < a \\ \frac{1}{b - a} & a \leq \xi \leq b \\ 0, & b < \xi \end{cases}$$

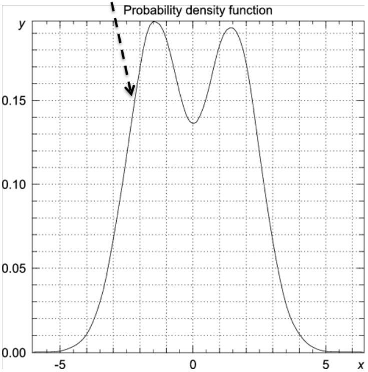
Phone Call Example (Cont.)



Probability Density Function Illustration

Typical $f_x(x)$





Area under the probability density function $f_x(x)$ in $[a,b]$ $=$ probability that the RV x will lie in $[a,b]$

An Important Observation

- As the interval gets progressively smaller, the probability tends to 0.
- For $\epsilon > 0$, we have:

$$P(a - \epsilon \leq \mathbf{x} \leq a + \epsilon) = \int_{a-\epsilon}^{a+\epsilon} f_{\mathbf{x}}(\xi) d\xi$$

- As $\epsilon \to 0 \implies P(a) = 0$, i.e. probability of random variable $\mathbf{x}$ taking a value **exactly equal to** a is 0.
- For a continuous random variable, we need not be precise about the inclusion/exclusion of interval end-points:

$$P(a \leq \mathbf{x} \leq b) = P(a < \mathbf{x} \leq b) = P(a < \mathbf{x} < b) = P(a \leq \mathbf{x} < b)$$

Another Important Observation

- For an arbitrarily small $\epsilon > 0$:

$$P(a - \epsilon \leq x \leq a + \epsilon) = \int_{a-\epsilon}^{a+\epsilon} f_{\mathbf{x}}(\xi) d\xi \approx 2\epsilon f_{\mathbf{x}}(a)$$

- $f_{\mathbf{x}}(a)$ can be interpreted as a relative measure of how likely it is that RV (random variable) **x** will be around a.
- In general, $f_{\mathbf{x}}(a)$ can have a very large value. It is NOT the probability of **x** at a.

Histogram as an Approximation of Probability Density Function for a Continuous RV

- Histogram of a continuous RV: An approximation of the probability density function.
- Relative frequency: estimate of the probability that a measurement falls in the interval.

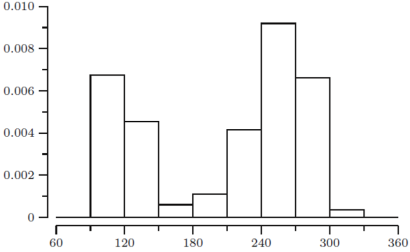


Figure: A Typical Histogram

Summary

- The modern axiomatic definition of the probability facilitates rigorous mathematical treatment of the random phenomenon.
- A probability space consists of the triplet (i) sample space, (ii) a σ-algebra field defined on the sample space and (iii) a probability measure defined on each event in the σ-algebra.
- The concept of a random variable is introduced because, we need a mapping from the sample space to the set of real numbers for carrying out quantitative analysis through a unified mathematical framework.

Summary (Cont.)

- After the random variable mapping
  - Sample Space: The real line
  - Field of interest: All possible intervals, point values of the real line
  - Probability Information: Captured via Cumulative Distribution Function $F_{\mathbf{x}}(\xi)$.
- Probability or cumulative distribution function always exists
- Equivalently:
  - Probability mass function for discrete random variable
  - Probability density function for continuous random variable

Thank You