# S&DS 241 Lecture 13

Continuous random variables, cumulative distribution function (CDF), probability density function (PDF)

B-H 3.6, 5.1

# Discrete random variables

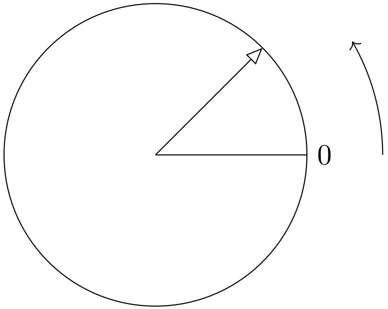
So far we have been focusing on discrete random variables/distributions

- Finite number of values: Coin/dice, Bernoulli, Binomial
- Countably infinite number of values: Geometric, Poisson

# Motivations of continuous random variables: (1)

Many physical quantities naturally take a continuum of values

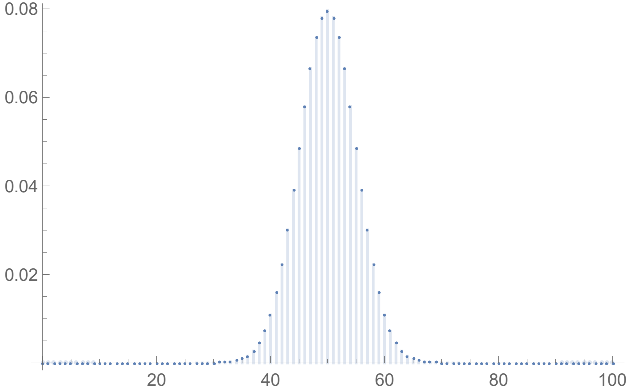
- Position of a randomly spinned wheel




- Arrival time of the next Yale shuttle
- Temperature tomorrow
- Location of a dust particle in water

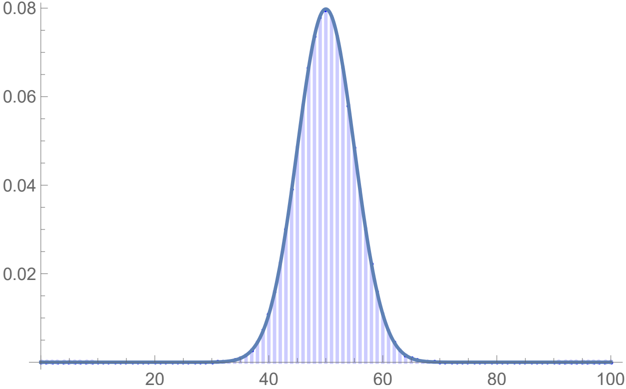
# Motivations of continuous random variables: (2)

Continuous approximation is useful!



# Motivations of continuous random variables: (2)

Continuous approximation is useful!



# Cumulative distribution function

# CDF

- The cumulative distribution function (CDF) of a random variable $X$ is

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

  This defines a function $F_X : \mathbb{R} \to [0, 1]$.

# CDF

- The cumulative distribution function (CDF) of a random variable $X$ is

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

This defines a function $F_X : \mathbb{R} \to [0, 1]$.

- General properties:
  - $F_X$ is a non-decreasing function.
  - Limits at infinity:

$$\underbrace{\lim_{x \to -\infty} F(x)}_{F(-\infty)} = 0, \quad \underbrace{\lim_{x \to \infty} F(x)}_{F(\infty)} = 1.$$

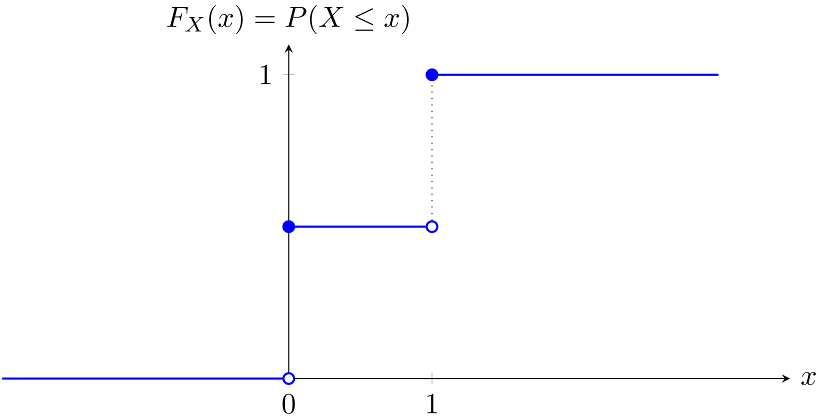
  - Right-continuous: $F(x) \to F(a)$ as $x \to a$ from the right.

# CDF of discrete random variables

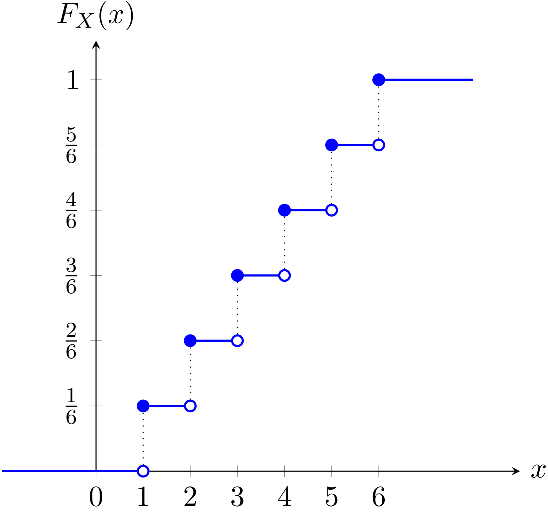
CDF = sum of PMFs:

$$F_X(x) = \sum_{t \leq x} \underbrace{p_X(t)}_{P(X=t)}$$

# CDF of $X \sim \text{Bern}(1/2)$



# CDF of $X$ = outcome of a fair die



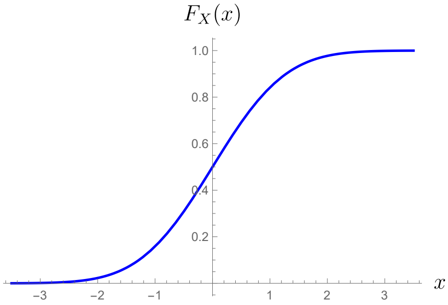
# Summary

- CDF of discrete random variables has jumps
  - Piecewise constant
  - Amount of jump = PMF
  - Not very convenient to use (contains the same information as PMF)

# Summary

- CDF of discrete random variables has jumps
  - Piecewise constant
  - Amount of jump = PMF
  - Not very convenient to use (contains the same information as PMF)
- CDF of continuous random variables have no jumps!

  For example:

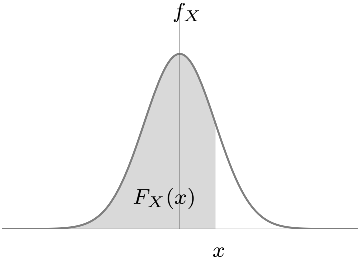


# Definition of continuous random variables

# Continuous random variable and PDF

A random variable $X$ is continuous if its CDF can be expressed as an integral, i.e., there exists a nonnegative function $f_X : \mathbb{R} \to \mathbb{R}$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \quad \text{for all } x$$

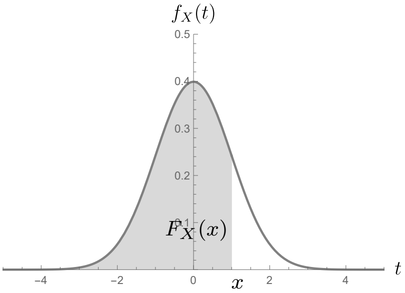


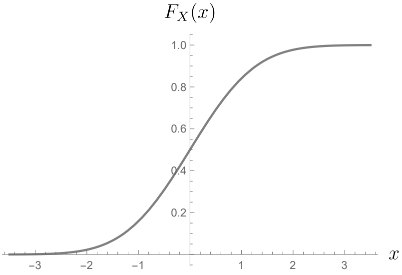

We call $f_X$ the probability density function (PDF) of $X$: by fundamental theorem of calculus,

$$f_X = F_X'$$

# PDF and CDF

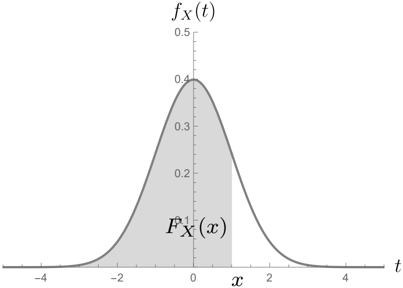
$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \qquad \text{for all } x$$

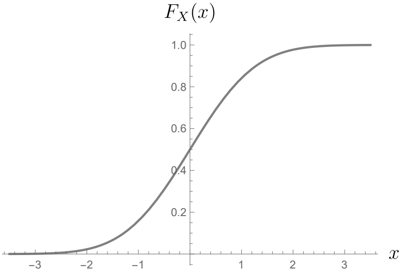


# PDF and CDF

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \quad \text{for all } x$$



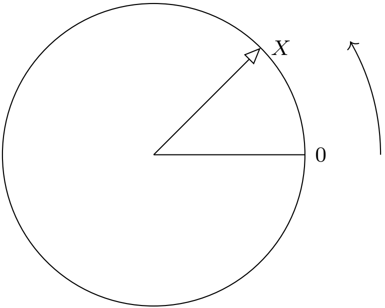




Since $f_X = F_X'$, by fundamental theorem of calculus

$$\int_{-\infty}^{x} f_X(t)dt = F_X(x) - \underbrace{F_X(-\infty)}_{0}$$

# Example: spinning wheel

Consider a wheel of unit radius. Let $X \in [0, 2\pi)$ denotes the angle.



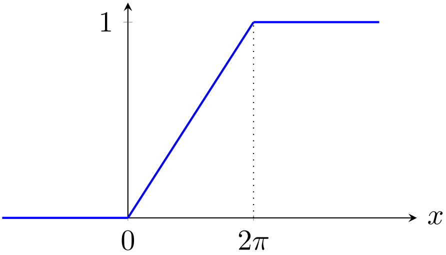

CDF of $X$

$$F_X(x) = P(X \leq x) = \frac{\text{arc length}}{\text{circumference}} = \frac{x}{2\pi},$$

provided that $0 \leq x < 2\pi$. If $x < 0$, $F_X(x) = 0$; if $x \geq 2\pi$, $F_X(x) = 1$.

# Example: spinning wheel

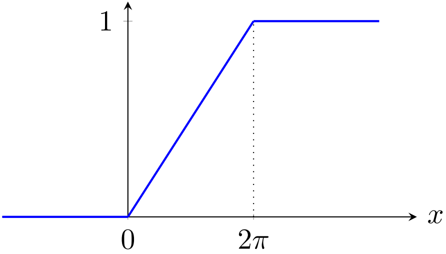
- CDF of $X$:




$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2\pi} & 0 \leq x < 2\pi \\ 1 & x \geq 2\pi \end{cases}$$

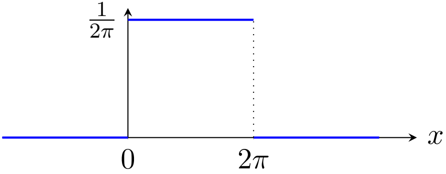
# Example: spinning wheel

- CDF of $X$:




$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2\pi} & 0 \le x < 2\pi \\ 1 & x \ge 2\pi \end{cases}$$

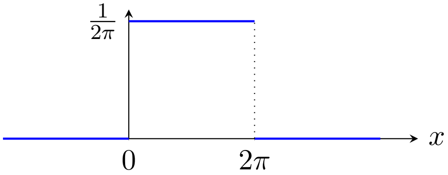
- PDF of $X$:




$$f_X(x) = \begin{cases} \frac{1}{2\pi} & 0 \le x < 2\pi \\ 0 & \text{otherwise} \end{cases}$$

# Support

- The support of $X$ is a set consisting of all $x$ where $f_X(x) > 0$
- This is the set of possible values that $X$ can take
- Example:




Support is $[0, 2\pi]$.

# Summary: CDF & PDF

- For continuous random variables,

$$\text{PDF} \underset{\text{differentiate}}{\overset{\text{integrate}}{\rightleftharpoons}} \text{CDF}$$

- In most problems, to find PDF, we first find CDF, then take derivative.

# Properties of continuous random variables

1. $F_X(x)$ is continuous in $x$ (no jumps)

# Properties of continuous random variables

1. $F_X(x)$ is continuous in $x$ (no jumps)
2. For any given $x \in \mathbb{R}$,

$$P(X = x) = 0$$

# Properties of continuous random variables

1. $F_X(x)$ is continuous in $x$ (no jumps)
2. For any given $x \in \mathbb{R}$,
$$P(X = x) = 0$$
3. For any $a \leq b$,
$$P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$$

(In the LHS whether we include the endpoints or not does not matter.)

# Properties of continuous random variables

1. $F_X(x)$ is continuous in $x$ (no jumps)
2. For any given $x \in \mathbb{R}$,
$$P(X = x) = 0$$
3. For any $a \leq b$,
$$P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$$
(In the LHS whether we include the endpoints or not does not matter.)
4. For $A \subset \mathbb{R}$
$$P(X \in A) = \int_A f_X(x)dx$$

# Discussion of $P(X = x) = 0$

- Mathematically: for any $\epsilon > 0$,

$$0 \leq P(X = x) \leq P(x - \epsilon < X \leq x) = F_X(x) - F_X(x - \epsilon)$$

By continuity of $F_X$, sending $\epsilon \to 0 \implies P(X = x) = 0$

# Discussion of $P(X = x) = 0$

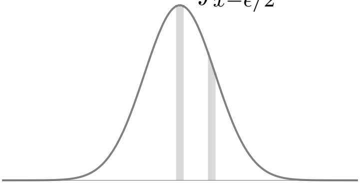
- Mathematically: for any $\epsilon > 0$,

$$0 \leq P(X = x) \leq P(x - \epsilon < X \leq x) = F_X(x) - F_X(x - \epsilon)$$

By continuity of $F_X$, sending $\epsilon \to 0 \implies P(X = x) = 0$

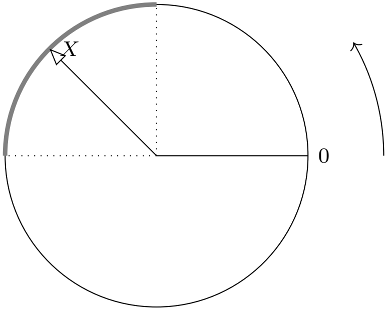
- Interpretation of PDF: for small $\epsilon$,

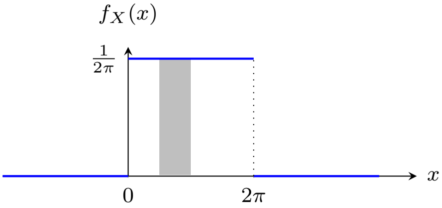
$$P(X \in (x - \epsilon/2, x + \epsilon/2)) = \int_{x-\epsilon/2}^{x+\epsilon/2} f_X(t)dt \approx f(x) \times \epsilon$$



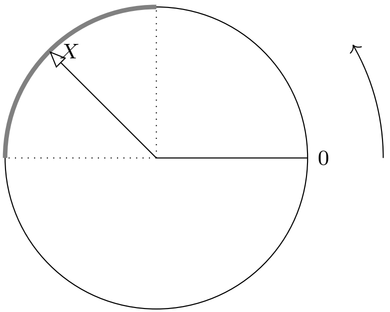
PDF tells us relatively how likely it is for $X$ to be near a given value

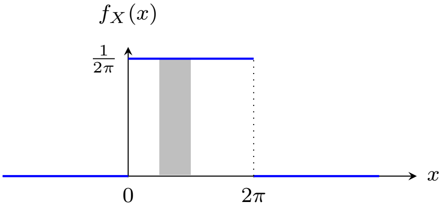
# Example: spinning wheel



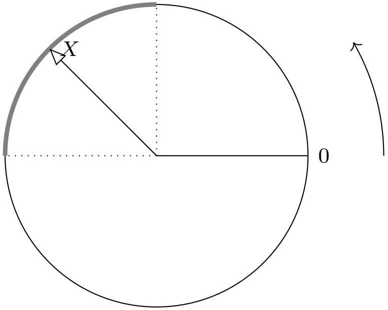
# Example: spinning wheel

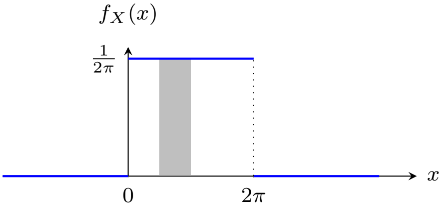



$$
\begin{aligned}
P(\text{land in the grey region}) &= P(\pi/2 \le X \le \pi) \\
&= \int_{\pi/2}^{\pi} f_X(x)dx = \int_{\pi/2}^{\pi} \frac{1}{2\pi}dx = \frac{1}{4}
\end{aligned}
$$

# Example: spinning wheel




$$
\begin{aligned}
P(\text{land in the grey region}) &= P(\pi/2 \le X \le \pi) \\
&= \int_{\pi/2}^{\pi} f_X(x)dx = \int_{\pi/2}^{\pi} \frac{1}{2\pi} dx = \frac{1}{4}
\end{aligned}
$$

Alternatively, $F_X(\pi) - F_X(\pi/2) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$.

# Properties of PDF

What function constitutes a valid PDF? Similarly to PMF,

1. Non-negativity:

$$f_X(x) \geq 0$$

2. Normalization:

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

since $P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f_X(x)dx = 1$

# PMF vs PDF

| PMF | PDF |
|---|---|
| $p_X(x) \geq 0$ | $f_X(x) \geq 0$ |

# PMF vs PDF

| PMF | PDF |
|---|---|
| $p_X(x) \geq 0$ | $f_X(x) \geq 0$ |
| $\sum_x p_X(x) = 1$ | $\int_{-\infty}^{\infty} f_X(x)dx = 1$ |

# PMF vs PDF

| PMF | PDF |
|---|---|
| $p_X(x) \geq 0$ | $f_X(x) \geq 0$ |
| $\sum_x p_X(x) = 1$ | $\int_{-\infty}^{\infty} f_X(x) dx = 1$ |
| $P(X \in A) = \sum_{x \in A} p_X(x)$ | $P(X \in A) = \int_A f_X(x) dx$ |

# PMF vs PDF

| PMF | PDF |
|---|---|
| $p_X(x) \geq 0$ | $f_X(x) \geq 0$ |
| $\sum_x p_X(x) = 1$ | $\int_{-\infty}^{\infty} f_X(x)dx = 1$ |
| $P(X \in A) = \sum_{x \in A} p_X(x)$ | $P(X \in A) = \int_A f_X(x)dx$ |
| $E(X) = \sum_x x p_X(x)$ | $E(X) = \int_{-\infty}^{\infty} x f_X(x)dx$ |

# PMF vs PDF

| PMF | PDF |
|---|---|
| $p_X(x) \geq 0$ | $f_X(x) \geq 0$ |
| $\sum_x p_X(x) = 1$ | $\int_{-\infty}^{\infty} f_X(x)dx = 1$ |
| $P(X \in A) = \sum_{x \in A} p_X(x)$ | $P(X \in A) = \int_A f_X(x)dx$ |
| $E(X) = \sum_x x p_X(x)$ | $E(X) = \int_{-\infty}^{\infty} x f_X(x)dx$ |
| $E(g(X)) = \sum_x g(x) p_X(x)$ | $E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)dx$ |

# PMF vs PDF

| PMF | PDF |
|---|---|
| $p_X(x) \geq 0$ | $f_X(x) \geq 0$ |
| $\sum_x p_X(x) = 1$ | $\int_{-\infty}^{\infty} f_X(x)dx = 1$ |
| $P(X \in A) = \sum_{x \in A} p_X(x)$ | $P(X \in A) = \int_A f_X(x)dx$ |
| $E(X) = \sum_x x p_X(x)$ | $E(X) = \int_{-\infty}^{\infty} x f_X(x)dx$ |
| $E(g(X)) = \sum_x g(x) p_X(x)$ | $E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)dx$ |

Distinctions:

- PMF: $p_X(x) = P(X = x) \in [0, 1]$
- PDF: it is possible $f_X(x) > 1$. We need to integrate PDF to get probability.

# Universal facts

The following apply to both discrete and continuous RVs ($\sum \to \int$):

- Linearity of expectation:

$$E(X+Y) = E(X) + E(Y)$$

- Sum of independent RVs:

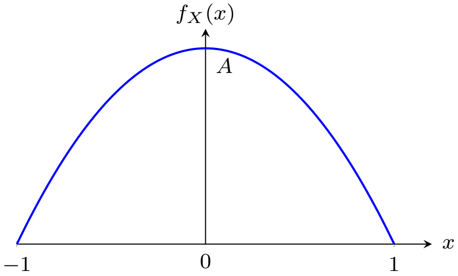
$$\mathrm{Var}(X+Y) \xlongequal{\text{independence}} \mathrm{Var}(X) + \mathrm{Var}(Y)$$

- Scaling rules:

$$\begin{aligned} E(aX+b) &= aE(X)+b \\ \mathrm{Var}(aX+b) &= a^2\mathrm{Var}(X) \\ \mathrm{SD}(aX+b) &= |a|\mathrm{SD}(X) \end{aligned}$$

- Markov's and Chebyshev's inequalities

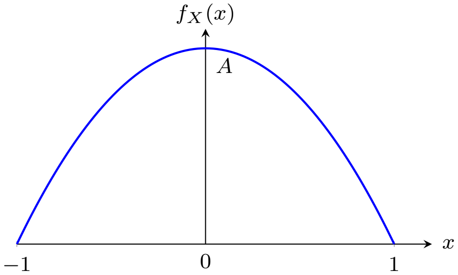
# Example




$$f_X(x) = \begin{cases} A(1-x^2) & |x| \le 1 \\ 0 & \text{else} \end{cases}$$

1. Find $A$:

# Example



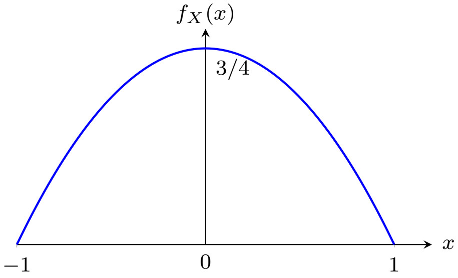

$$f_X(x) = \begin{cases} A(1-x^2) & |x| \leq 1 \\ 0 & \text{else} \end{cases}$$

1. Find $A$:

$$1 = \int_{-1}^{1} f_X(x)dx = A\int_{-1}^{1}(1-x^2)dx = A\left(x - \frac{x^3}{3}\Big|_{-1}^{1}\right) = \frac{4A}{3}$$

$$\implies A = 3/4$$

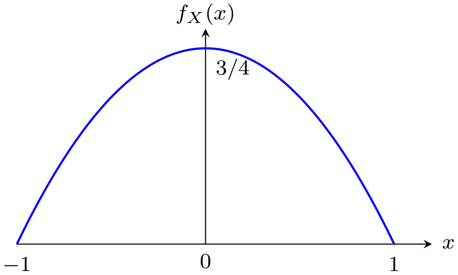
# Example




$$f_X(x) = \begin{cases} \frac{3}{4}(1 - x^2) & |x| \leq 1 \\ 0 & \text{else} \end{cases}$$

2. Find $E(X)$:

# Example

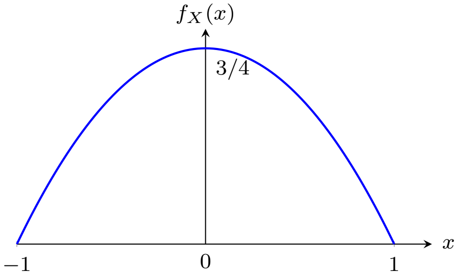



$$f_X(x) = \begin{cases} \frac{3}{4}(1 - x^2) & |x| \leq 1 \\ 0 & \text{else} \end{cases}$$

2. Find $E(X)$:

$$E(X) = \int_{-1}^{1} x \frac{3}{4}(1 - x^2) dx$$

# Example



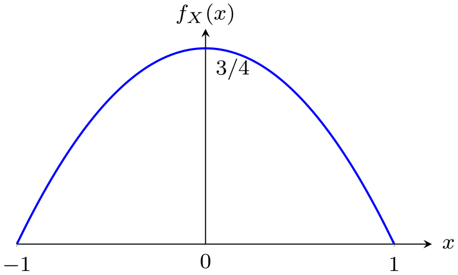

$$f_X(x) = \begin{cases} \frac{3}{4}(1 - x^2) & |x| \le 1 \\ 0 & \text{else} \end{cases}$$

2. Find $E(X)$:

$$E(X) = \int_{-1}^{1} x \frac{3}{4}(1 - x^2) dx = 0$$

because integral of an odd function over a symmetric interval is 0 (review: B-H A.2.3)

# Example

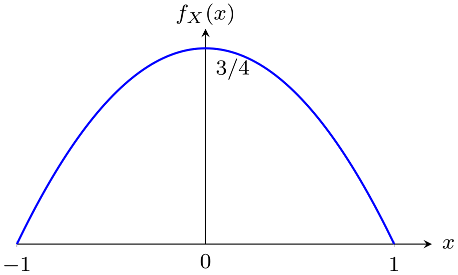



$$f_X(x) = \begin{cases} \frac{3}{4}(1 - x^2) & |x| \le 1 \\ 0 & \text{else} \end{cases}$$

3. Find $\mathrm{Var}(X)$:

$$\begin{aligned} \mathrm{Var}(X) = E(X^2) &\overset{\text{LOTUS}}{=\!=\!=} \int_{-1}^{1} x^2 \frac{3}{4}(1 - x^2)dx \\ &= \frac{3}{4} \times 2 \times \int_0^1 (x^2 - x^4)dx = \frac{3}{4} \times 2 \times \left(\frac{1}{3} - \frac{1}{5}\right) = \frac{1}{5} \end{aligned}$$

# Example

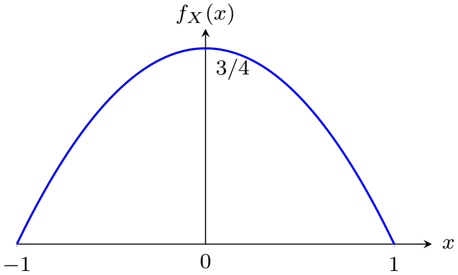



$$f_X(x) = \begin{cases} \frac{3}{4}(1-x^2) & |x| \le 1 \\ 0 & \text{else} \end{cases}$$

4. Find $P\left(X \ge \frac{1}{2} \middle| |X| \ge \frac{1}{2}\right)$:

$$P\left(X \ge \frac{1}{2} \middle| |X| \ge \frac{1}{2}\right) = \frac{P\left(X \ge \frac{1}{2}\right)}{P\left(|X| \ge \frac{1}{2}\right)} = \frac{1}{2}$$

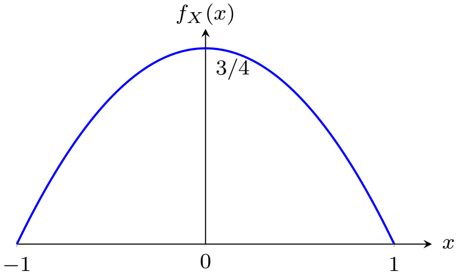
# Example




$$f_X(x) = \begin{cases} \frac{3}{4}(1 - x^2) & |x| \le 1 \\ 0 & \text{else} \end{cases}$$

5. Find CDF: $F_X(x) = P(X \le x)$.
   - $x > 1$: $F_X(x) = 1$
   - $x < -1$: $F_X(x) = 0$

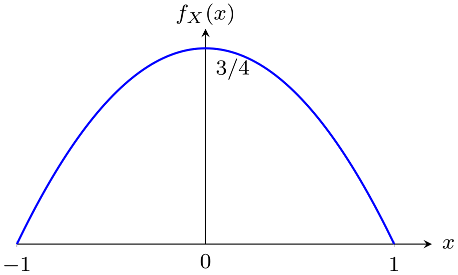
# Example




$$f_X(x) = \begin{cases} \frac{3}{4}(1 - x^2) & |x| \leq 1 \\ 0 & \text{else} \end{cases}$$

5. Find CDF: $F_X(x) = P(X \leq x)$.
   - $x > 1$: $F_X(x) = 1$
   - $x < -1$: $F_X(x) = 0$
   - $|x| \leq 1$: $F_X(x) = \int_{-1}^{x} f_X(t)dt = \int_{-1}^{x} \frac{3}{4}(1 - t^2)dt = \frac{2+3x-x^3}{4}$

# Example




$$f_X(x) = \begin{cases} \frac{3}{4}(1-x^2) & |x| \le 1 \\ 0 & \text{else} \end{cases}$$

5. Find CDF: $F_X(x) = P(X \le x)$.
   - $x > 1$: $F_X(x) = 1$
   - $x < -1$: $F_X(x) = 0$
   - $|x| \le 1$: $F_X(x) = \int_{-1}^{x} f_X(t)dt = \int_{-1}^{x} \frac{3}{4}(1-t^2)dt = \frac{2+3x-x^3}{4}$