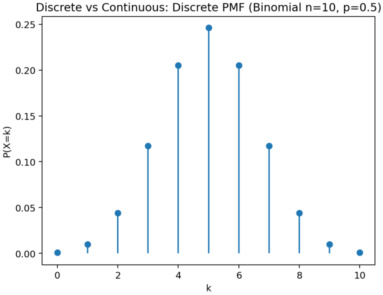
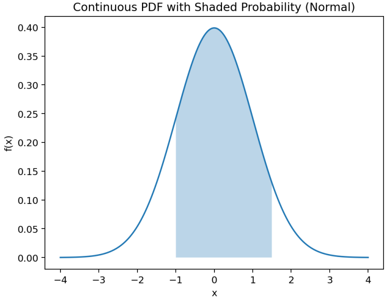
Continuous Probability

Discrete vs. Continuous (Quick Review)

- **Discrete:** outcomes are countable (e.g., # of bugs, coin flips).
- **Continuous:** outcomes vary over a continuum (e.g., time, height).
- Region/area carries probability for continuous; points have probability 0.



Discrete PMF (Binomial)



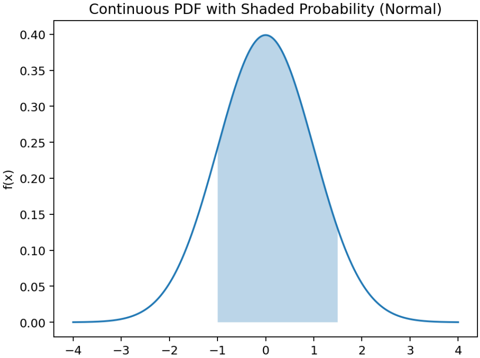
Continuous PDF with shaded

Probability Density Function (PDF)

The **probability density function** (pdf) describes relative likelihood:

$$f_X(x) \geq 0, \quad \int_{-\infty}^{\infty} f_X(x)\,dx = 1, \quad P(a \leq X \leq b) = \int_a^b f_X(x)\,dx.$$

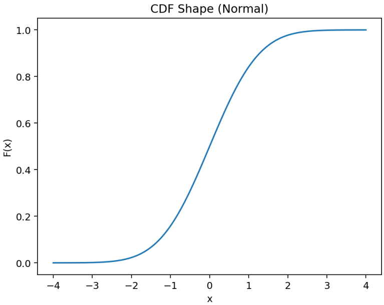
Important: For continuous X, $P(X = a) = 0$.



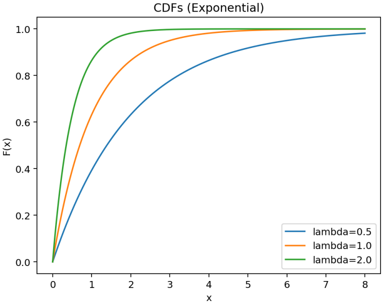
Cumulative Distribution Function (CDF)

The **CDF** accumulates probability from $-\infty$ to x:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt, \quad P(a \leq X \leq b) = F_X(b) - F_X(a).$$



Normal CDF



Exponential CDFs (various λ)

Using $F_X(b) - F_X(a)$

Suppose a random variable X has CDF

$$F_X(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leq x \leq 1, \\ 1, & x > 1\ , \end{cases}$$

which corresponds to a Uniform$(0, 1)$ distribution.

Probability via the CDF:

$$P(0.20 \leq X \leq 0.60) = F_X(0.60) - F_X(0.20) = 0.60 - 0.20 = 0.40.$$

Using $F_X(b) - F_X(a)$

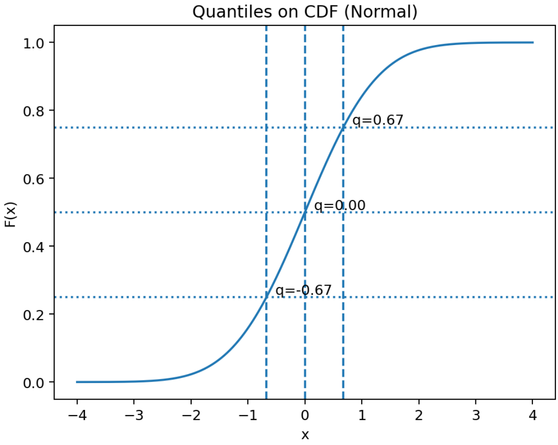
Let X have CDF

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0, \end{cases} \quad \text{with } \lambda = 1.$$

Compute the probability $P(1 \leq X \leq 2)$ using $F_X(b) - F_X(a)$.

Quantiles and Percentiles

- The p-quantile q_p solves $F_X(q_p) = p$ (e.g., median $= q_{0.5}$).
- **Percentiles:** $p = 0.25, 0.5, 0.75$ give quartiles..



How to Find a Quantile (General Steps)

To find the p-quantile q_p for a random variable X:

1. Start with the CDF $F_X(x)$ of X.
2. Set $F_X(q_p) = p$ for the desired percentile p (e.g., $p = 0.25$ for the first quartile).
3. Solve the equation for q_p.
4. Interpret: q_p is the cutoff so that p proportion of values lie below it.

Quartiles (Worked Example)

Example: Suppose a random variable X has CDF

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \dfrac{x}{100}, & 0 \le x \le 100, \\ 1, & x > 100. \end{cases}$$

Then

$$Q_1 = 25, \quad Q_2 = 50, \quad Q_3 = 75.$$

Quartiles Practice

Suppose X has CDF

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-x}, & x \geq 0. \end{cases}$$

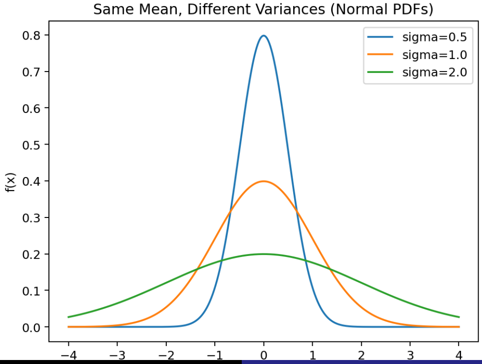
1. Find the first quartile Q_1: solve $F_X(Q_1) = 0.25$.
2. Find the median Q_2: solve $F_X(Q_2) = 0.50$.
3. Find the third quartile Q_3: solve $F_X(Q_3) = 0.75$.

Mean and Variance (Continuous)

For a continuous X with pdf f:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\,dx, \quad \mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx.$$

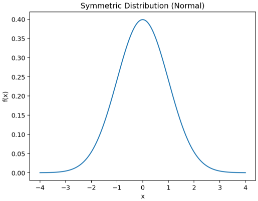
Intuition: Center and spread of the distribution (we'll compute these via software).



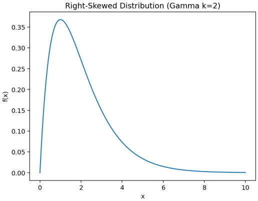
Shape: Symmetry and Skewness (Light Intro)

- **Symmetric:** left/right mirror around the mean (e.g., Normal).
- **Right-skewed:** long right tail (e.g., Exponential, Lognormal).
- **Left-skewed:** long left tail.

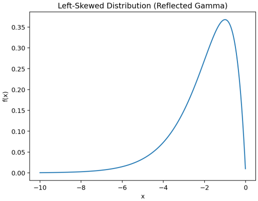
(Optional curiosity) Pearson's moment skewness $\gamma_1 = \frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3}$.



Symmetric (Normal)



Right-skew (Gamma)



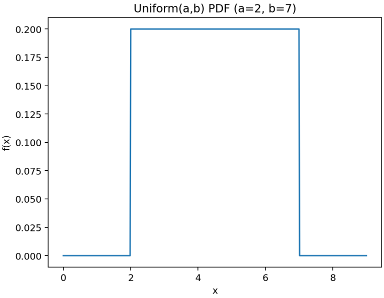
Left-skew (Reflected)

Uniform Distribution (Baseline)

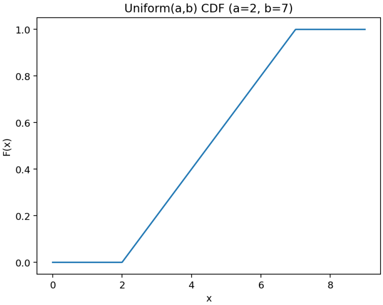
If $X \sim \mathrm{Unif}(a, b)$:

$$f(x) = \frac{1}{b-a} \text{ for } x \in [a, b], \quad 0 \text{ otherwise.}$$

$$\mathbb{E}[X] = \frac{a+b}{2}, \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$



Uniform PDF



Uniform CDF

Uniform Example (Worked)

Suppose dinner time in a household is equally likely to occur anytime between 5:00 pm and 7:00 pm. Let $X =$ the dinner time. Then $X \sim \text{Unif}(5, 7)$.

- Mean: $\mathbb{E}[X] = \frac{5+7}{2} = 6 \;\Rightarrow\;$ average dinner time $=$ 6:00 pm.
- Variance: $\mathrm{Var}(X) = \frac{(7-5)^2}{12} = \frac{4}{12} = 0.33$.
- Probability: $P(5.5 \leq X \leq 6.5) = \frac{1}{2}(6.5 - 5.5) = 0.5$. That is, there is a 50% chance dinner is between 5:30 pm and 6:30 pm.
- 30th percentile: $q_{0.30} = 5 + 0.30 \cdot 2 = 5.6 \;\Rightarrow\;$ 5:36 pm.

You Try It: Uniform Practice

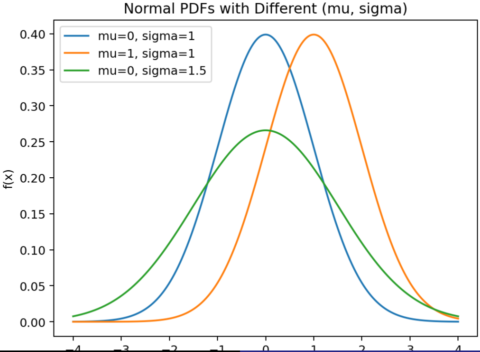
Suppose a bus arrives uniformly at random between 8:00 am and 8:30 am. Let X = minutes past 8:00. Then $X \sim \text{Unif}(0, 30)$.

1. What is the mean and variance of X?
2. What is $P(10 \leq X \leq 20)$ (between 8:10 and 8:20)?
3. What is the 90th percentile of X (the time by which 90% of buses have arrived)?

Normal Distribution

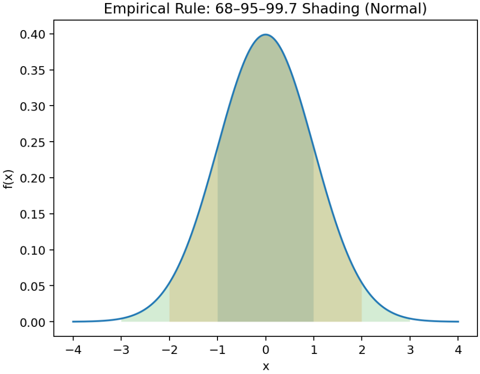
$$X \sim \mathcal{N}(\mu, \sigma), \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right).$$

- Symmetric, fully described by μ and σ.
- Common for aggregated effects; central in inference.



Standardization & Empirical Rule

- **Z-score:** $Z = \dfrac{X - \mu}{\sigma}$ converts X to standard normal.
- **Empirical rule (approx):** 68% in $\mu \pm \sigma$, 95% in $\mu \pm 2\sigma$, 99.7% in $\mu \pm 3\sigma$.



Using the 68–95–99.7 Rule (Worked Example)

Suppose exam scores are normally distributed, $X \sim \mathcal{N}(75, 10)$.

- About 68% of scores are within 1 σ: between 65 and 85.
- About 95% are within 2 σ: between 55 and 95.
- About 99.7% are within 3 σ: between 45 and 105.

Example: Estimate $P(65 \leq X \leq 85)$.

$$P(65 \leq X \leq 85) \approx 0.68.$$

Work out the following

- $P(75 \leq X \leq 95)$.
- $P(85 \leq X \leq 95)$.
- $P(55 \leq X \leq 85)$

Empirical Rule Practice

Suppose heights of a plant species are $X \sim \mathcal{N}(50, 5)$ cm.

1. Estimate $P(45 \leq X \leq 55)$.
2. Estimate $P(40 \leq X \leq 50)$.
3. Estimate $P(35 \leq X \leq 65)$.

The Normal Table

- Standard Normal: $Z \sim \mathcal{N}(0,1)$.
- The table gives $P(Z \leq z)$ for values of z.
- For any Normal $X \sim \mathcal{N}(\mu, \sigma)$, standardize:

$$Z = \frac{X - \mu}{\sigma}.$$

Example: Suppose $X \sim \mathcal{N}(100, 15)$. Find $P(X \leq 115)$.

$$Z = \frac{115 - 100}{15} = 1.$$

From the table: $P(Z \leq 1.00) \approx 0.8413$.

Finding Percentiles Using the Normal Table

- Percentiles are quantiles: find q_p such that $P(X \leq q_p) = p$.
- Standardize: solve for z_p with $P(Z \leq z_p) = p$ using the Normal table.
- Convert back: $q_p = \mu + \sigma z_p$.

Example: Suppose $X \sim \mathcal{N}(70, 8)$ (exam scores). Find the 90th percentile.

$$\text{Step 1: } P(Z \leq z) = 0.90 \;\Rightarrow\; z \approx 1.28.$$

$$\text{Step 2: } q_{0.90} = 70 + 8(1.28) \approx 80.2.$$

Interpretation: 90% of scores are below about 80 points.

Normal Table Practice

Suppose IQ scores are approximately $X \sim \mathcal{N}(100, 15)$.

1. What is the probability that a randomly chosen person has an IQ above 120?
2. What IQ score marks the top 10% (the 90th percentile)?

Hints:

- Standardize: $Z = (X - \mu)/\sigma$.
- Use the Normal table to find $P(Z \leq z)$.
- For (2), look up z such that $P(Z \leq z) = 0.95$.

Normal Example (Worked)

Suppose adult male heights follow $X \sim \mathcal{N}(\mu = 70 \text{ in}, \ \sigma = 3 \text{ in})$.

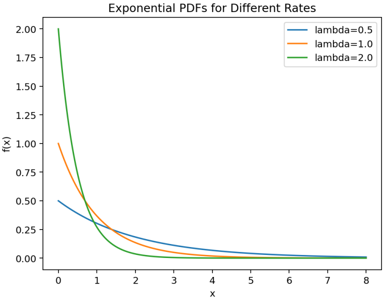
- Mean: $\mathbb{E}[X] = 70$ in.
- Variance: $\mathrm{Var}(X) = 9$.
- Probability: $P(67 \leq X \leq 73) = P(-1 \leq Z \leq 1) \approx 0.6826$. So about 68% of men are between 5'7" and 6'1".
- 30th percentile: $z_{0.30} \approx -0.52$. $q_{0.30} = 70 + 3(-0.52) \approx 68.4$ in.

Exponential Distribution (Waiting Times)

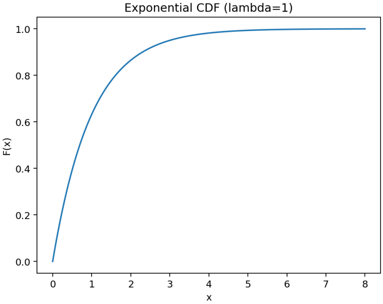
If $X \sim \mathrm{Exp}(\lambda)$ with rate $\lambda > 0$:

$$f(x) = \lambda e^{-\lambda x} \ (x \geq 0), \quad F(x) = 1 - e^{-\lambda x}, \quad \mathbb{E}[X] = \frac{1}{\lambda}, \ \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Memoryless: $P(X > s + t \mid X > s) = P(X > t)$.



Exponential PDFs (various λ)



Exponential CDF ($\lambda = 1$)

Exponential Example (Worked)

Suppose the waiting time between buses at a station is $X \sim \text{Exp}(\lambda = 1/10)$, i.e. average wait = 10 minutes.

- Mean: $\mathbb{E}[X] = 10$ minutes.
- Variance: $\mathrm{Var}(X) = 100$.
- Probability: $P(5 \leq X \leq 15) = F(15) - F(5) = (1 - e^{-1.5}) - (1 - e^{-0.5}) \approx 0.38$.
- 30th percentile: Solve $F(q_{0.30}) = 0.30$, $q_{0.30} = -10 \ln(0.70) \approx 3.6$ minutes.

Exponential Practice

Suppose the time until your next text message is $X \sim \text{Exp}(\lambda = 2)$ minutes (average wait $= 0.5$ minutes).

1. What is the mean and variance?
2. Compute $P(0.5 \leq X \leq 1.5)$.
3. Find the 80th percentile.

(Check after you try: Mean=0.5 min, Var=0.25, Probability ≈ 0.38, *80th percentile* ≈ 0.80 *min.)*