Random Variables, PMFs, PDFs

Contents

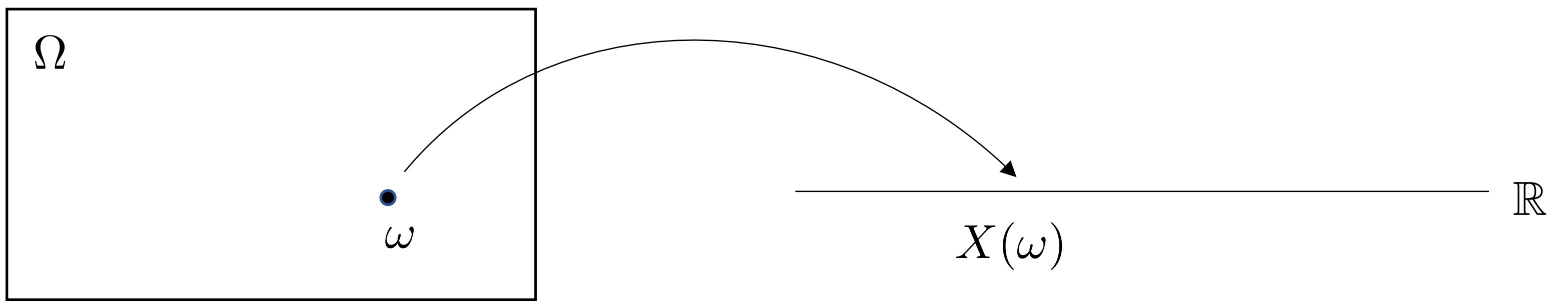
- random variables
- discrete random variables
- probability mass functions
- cumulative distribution functions
- probability density functions

Random variables

- probability space consists of $(\Omega, \mathcal{F}, \mathbb{P})$
- A is a set of outcomes, i.e, a subset of Ω
  - $\mathbb{P}(A)$ is the probability that the outcome of a random experiment is in the set A
  - $\mathbb{P}(A)$ is the probability that the event A *occurs* or is *true*

- a random variable X is a real valued function $X(\omega)$ over the sample space Ω

$$X : \Omega \to \mathbb{R}$$



- use upper case letters for random variables: $X, Y, Z, \ldots$
- lower case letters for the values the random variable assumes, i.e, the event $X = x$

Discrete Random Variables and pmfs

- definition: a random variable is a *discrete random variable* if

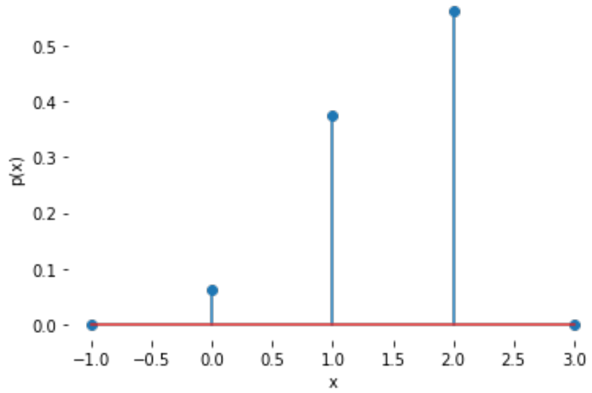
$$\mathbb{P}(X \in \mathcal{X}) = 1 \quad \text{for some countable set } \mathcal{X} \subset \mathbb{R}$$

- i.e, if you can enumerate the possible values assumed by X, then it's a discrete R.V.
- a discrete random variable is completely specified by its probability mass function (pmf) $p(x)$

$$p(x) = \mathbb{P}(X = x) \qquad \text{for all } x \in X$$

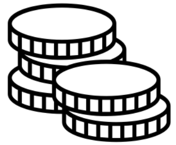
properties of pmfs:

1. $p(x) \geq 0$
2. $\sum_{x \in \mathcal{X}} p(x) = 1$



- notation: $X \sim p(x)$ indicates that the random variable X has pmf $p(X)$
- a random variable that is not discrete *might* be *continuous* (more on this later)
- we'll say pmf when we are being specific, and *distribution* more generically

Bernoulli RVs, Logistic regression



bias

- Bernoulli: $Y \sim \text{Bern}(p)$ for $0 \leq p \leq 1$

$$y = \begin{cases} 1 & \text{with probability } p \\ -1 & \text{with probability } 1 - p \end{cases}$$

- **logistic regression** is a **discriminative classifier** that models the posterior probability as a Bernoulli R.V.

$$\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1\ldots}$$

bias

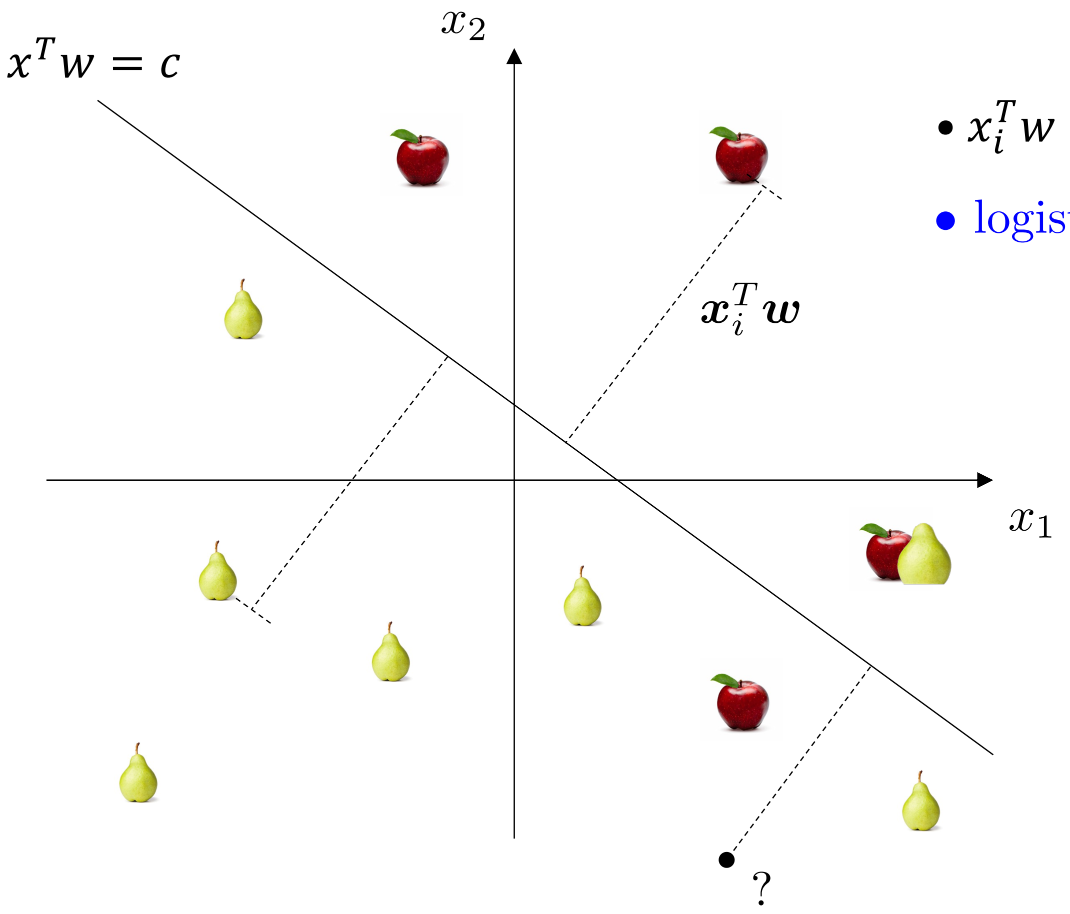
$$y_i = \begin{cases} 1 & \text{with probability } p_i \\ -1 & \text{with probability } 1 - p_i \end{cases} \qquad p_i = \frac{1}{1 + e^{-\boldsymbol{x}_i^T \boldsymbol{w}}}$$

$0 \leq p_i \leq 1$

example: logistic regression

- **logistic regression** is a **discriminative classifier** that models the posterior probability as a Bernoulli R.V.

$$y_i = \begin{cases} 1 & \text{with probability } p_i \\ -1 & \text{with probability } 1 - p_i \end{cases} \qquad p_i = \frac{1}{1 + e^{-\boldsymbol{x}_i^T \boldsymbol{w}}}$$



- $x_i^T w$ is related to the signed distance from the decision boundary
- logistic function takes $\boldsymbol{x}_i^T \boldsymbol{w}$ and squishes it to make it a probability

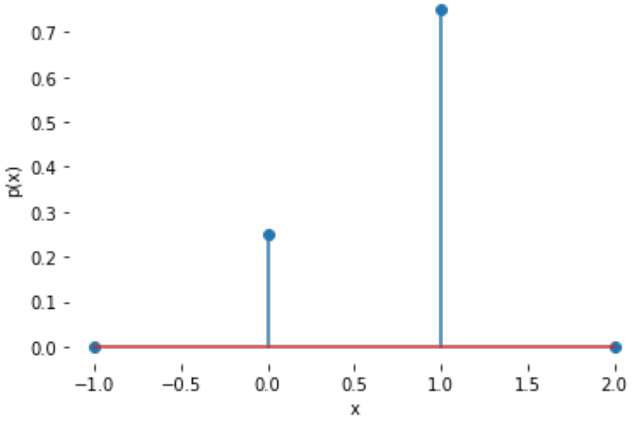
- probability that ? is a pear/apple depends distance from boundary

- training is finding a $\boldsymbol{w}$ that does a good job, i.e., maximum likelihood estimation for $\boldsymbol{w}$

Famous Discrete Random Variables

- Bernoulli: $X \sim \text{Bern}(p)$ for $0 \leq p \leq 1$

$$p(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

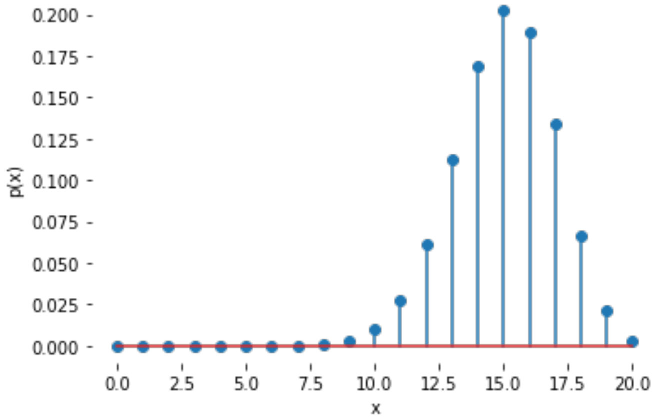


- Binomial: $X \sim \text{B}(n, p)$ for integer $n > 0$, $0 \leq p \leq 1$

$$p(x) = \binom{n}{x} p^x (1 - p)^{(n-x)} \quad \text{for } x = 0, \ldots, n$$

represents the number of heads in n flips

$(n = 20, p = .75)$

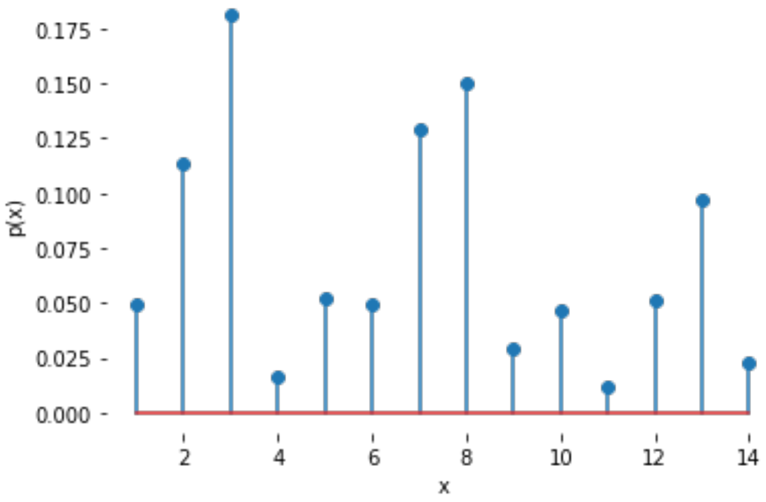


Famous Discrete Random Variables

- categorical : $X \sim p(x)$, for $|\mathcal{X}| < \infty$

$$p(x) = \begin{cases} p_{\text{cat}} & \text{if } x = \text{cat} \\ p_{\text{dog}} & \text{if } x = \text{dog} \\ \vdots & \\ p_{\text{fish}} & \text{if } x = \text{fish} \\ 0 & \text{otherwise} \end{cases}$$

$$p(x) = \begin{cases} p_1 & \text{if } x = 1 \\ p_2 & \text{if } x = 2 \\ \vdots & \\ p_k & \text{if } x = k \\ 0 & \text{otherwise} \end{cases}$$



- multinomial : $X \sim p(\boldsymbol{x}) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x^k}$ with $\sum_i x_i = n$

represents counts of outcomes of each category from n i.i.d categorical random variables

Famous Discrete Random Variables

- Geometric: $X \sim \text{Geom}(p)$ for $0 \leq p \leq 1$

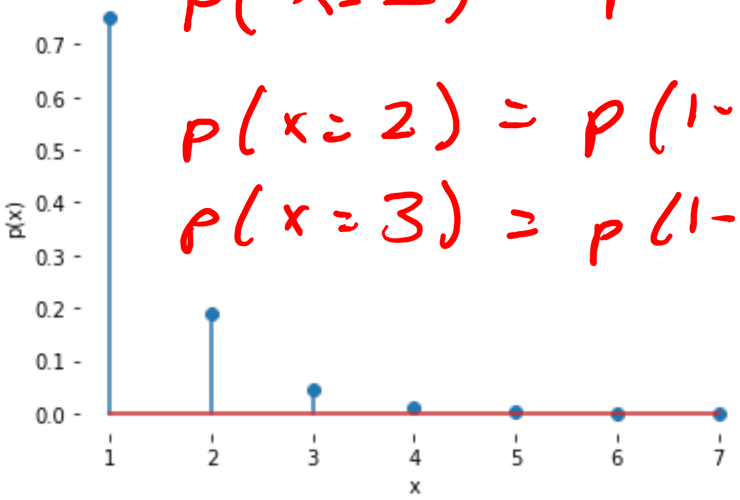
$$p(x) = p(1-p)^{x-1} \text{ for } x = 1, 2, \ldots$$

represents the number of coin flips until a heads shows up

$p(X=1) = p$

$p(x=2) = p(1-p)$

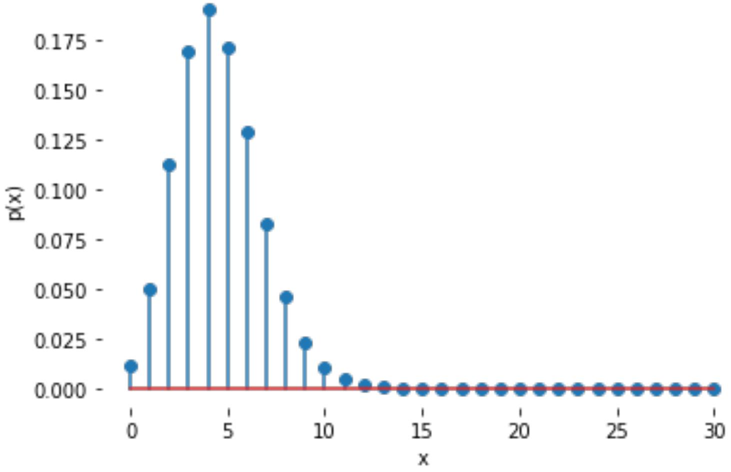
$p(x=3) = p(1-p)^2$



- Poisson: $X \sim \text{Poisson}(\lambda)$ for $\lambda > 0$

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \text{ for } x = 0, 1, 2, \ldots$$

represents the number of events in a time fixed time interval



Famous Discrete Random Variables

- Zipf's law:

$$p_m(x) = C_m/x^m, \quad x = 1, 2, 3, \ldots \text{ and } m > 1.$$

CDFs and PDFs

Cumulative Distribution Functions

- probability space consists of $(\Omega, \mathcal{F}, \mathbb{P})$
- a random variable X is a mapping $X : \Omega \to \mathbb{R}$
- ways to characterize $\mathbb{P}(X \in A)$:

    i) specify $\mathbb{P}(X \in A)$ for any Borel set $A \subset \mathbb{R}$

    - Borel set $A \subset \mathbb{R} :=$ any set generated by countable unions, intersections and compliments of intervals

    ii) specify $\mathbb{P}(X \in (a, b])$ for all intervals

    iii) specify the Cumulative Density Function (CDF) of X

$$F(x) = \mathbb{P}(X \leq x)$$

- example: $\mathbb{P}(a < X \leq b)$?

$$= \mathbb{P}(X \leq b, X > a)$$

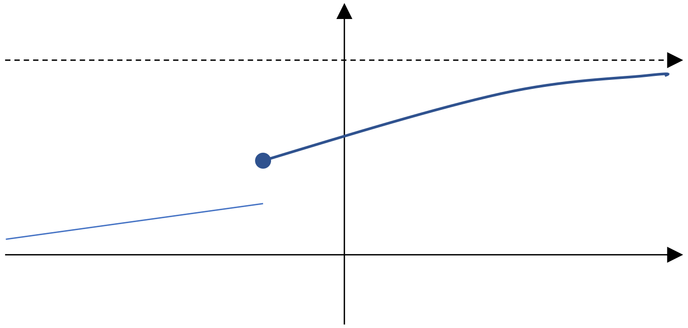
$$= \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a)$$

$$= F(b) - F(a)$$

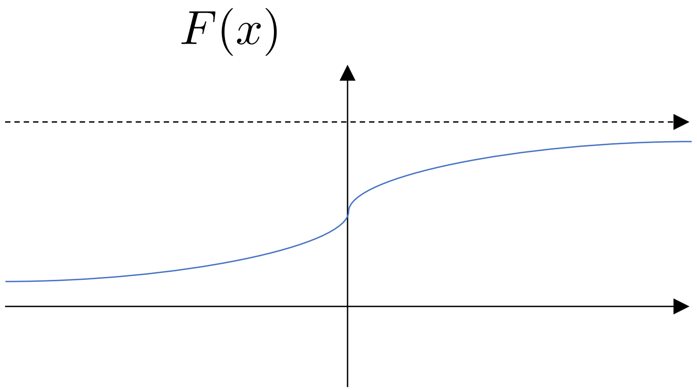
Cumulative Distribution Functions

Cumulative Density Function (CDF) of X

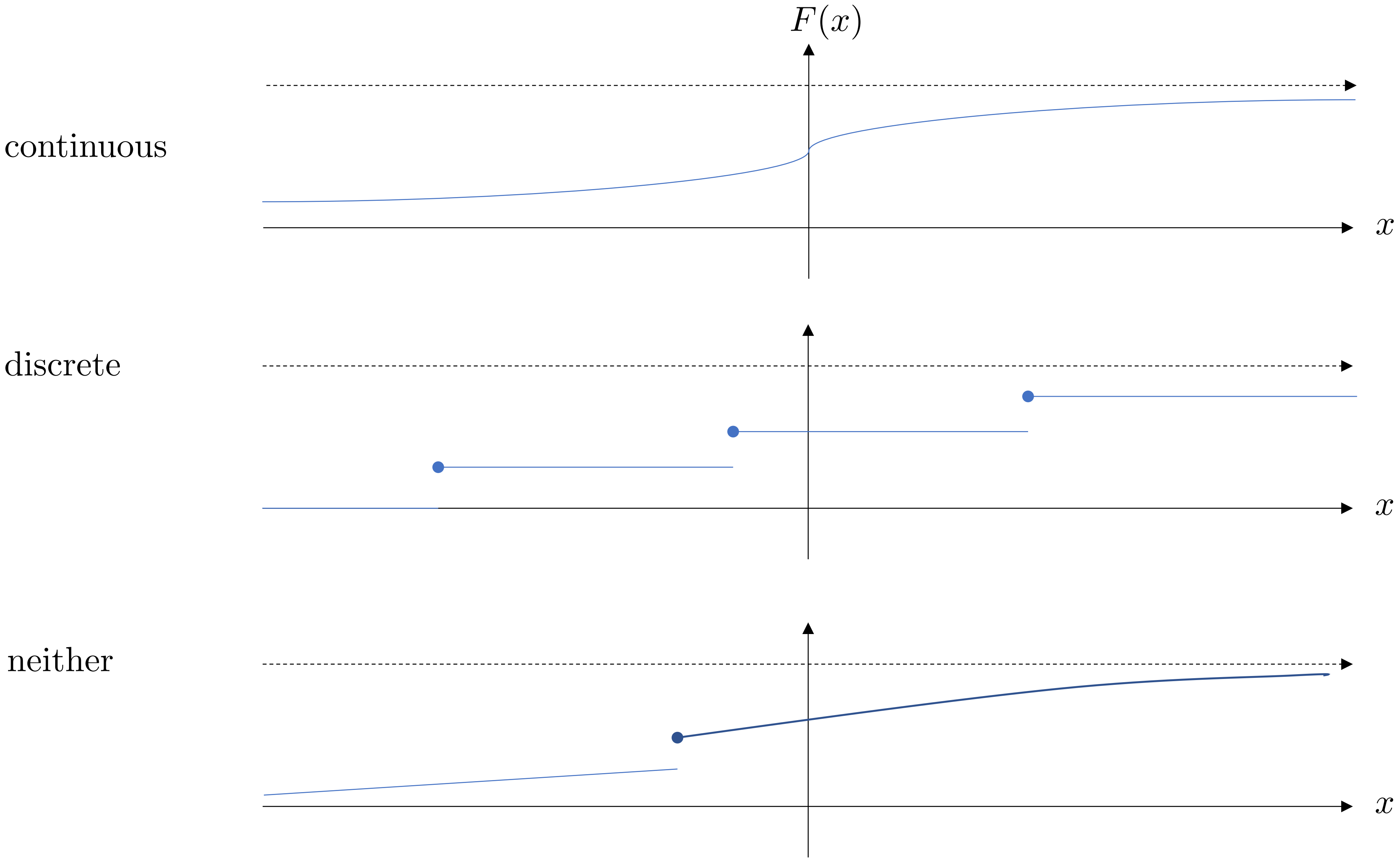
$$F(x) = \mathbb{P}(X \leq x)$$



- properties:
  - $F(x) \geq 0$, monotonically nondecreasing
  - $\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$
  - $F(x)$ is right continuous: $\lim_{x \to a^+} F(x) = F(a)$
  - $\mathbb{P}(X = a) = F(a) - \lim_{x \to a^-} F(x)$
  - a random variable is continuous if $F(x)$ is a continuous function



Continuous Random Variables and Probability Density Functions



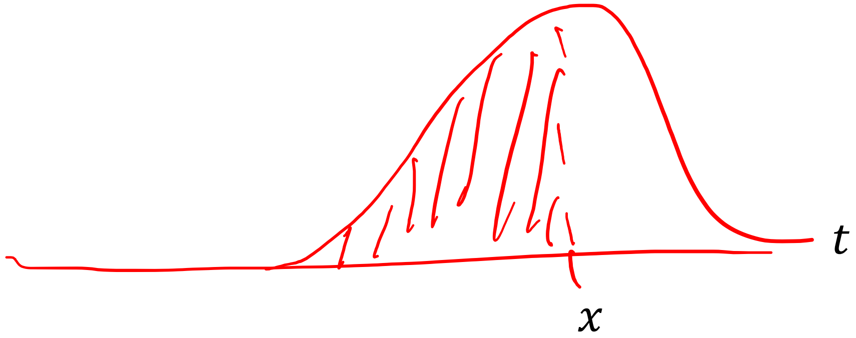
Probability Density Functions

- A **continuous random variable** is a ***real-valued*** RV for which there is a function f such that for all Borel sets A,

$$\mathbb{P}(X \in A) = \int_A f(t)\, dt.$$

- Specializing to $A = (-\infty, x]$ shows that

$$F(x) := \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\, dt.$$



- If f is continuous at x, then

$$F'(x) = \frac{d}{dx} F(x) = f(x).$$

Probability Density Functions

- Saying X is real valued means $\mathbb{P}(X \in \mathbb{R}) = 1$. Thus, taking $A = \mathbb{R}$ shows that

$$1 = \mathbb{P}(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f(t)\, dt.$$

- Densities must be nonnegative; i.e., $f(x) \geq 0$ (except on sets of zero length; e.g., a discrete set of points).

- If f is is allowed to contain Dirac impulses $\delta(x)$, then we sometimes say that f is a ***generalized density***. Such densities can accommodate discrete RVs and RVs with both “continuous” and “discrete” parts.

- A density function f is not a probability; it can be greater than one or less than one!

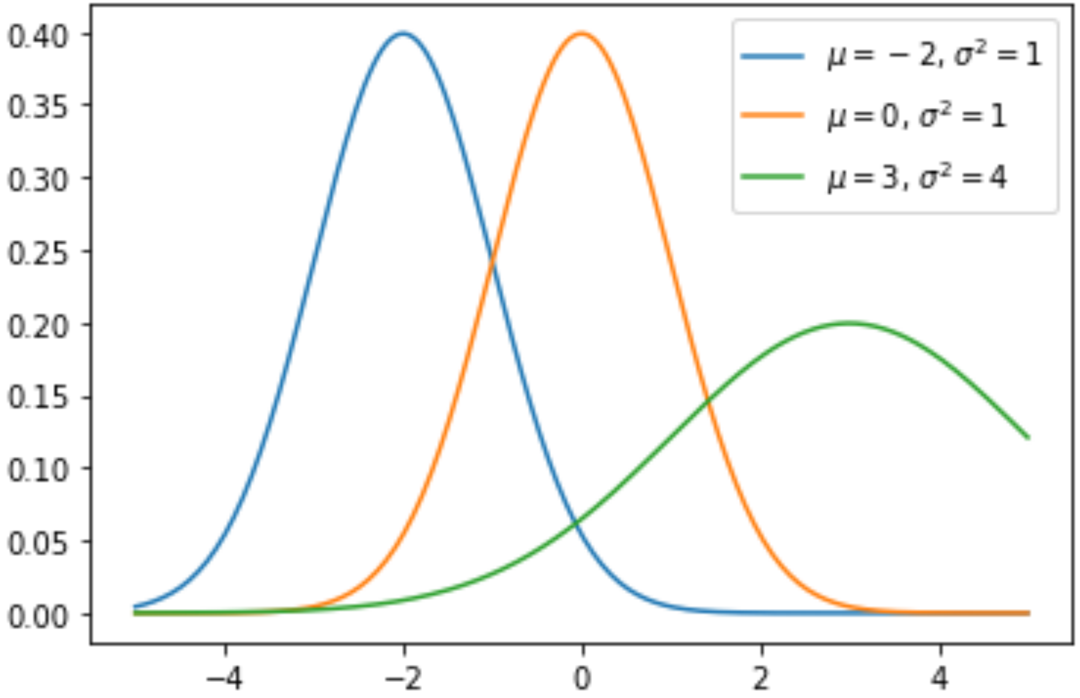
- We write $X \sim f$ or $X \sim F$

Probability Density Functions – Normal or Gaussian

- The standard normal density, denoted by $N(0,1)$, is

$$\varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

- The normal density with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$, is



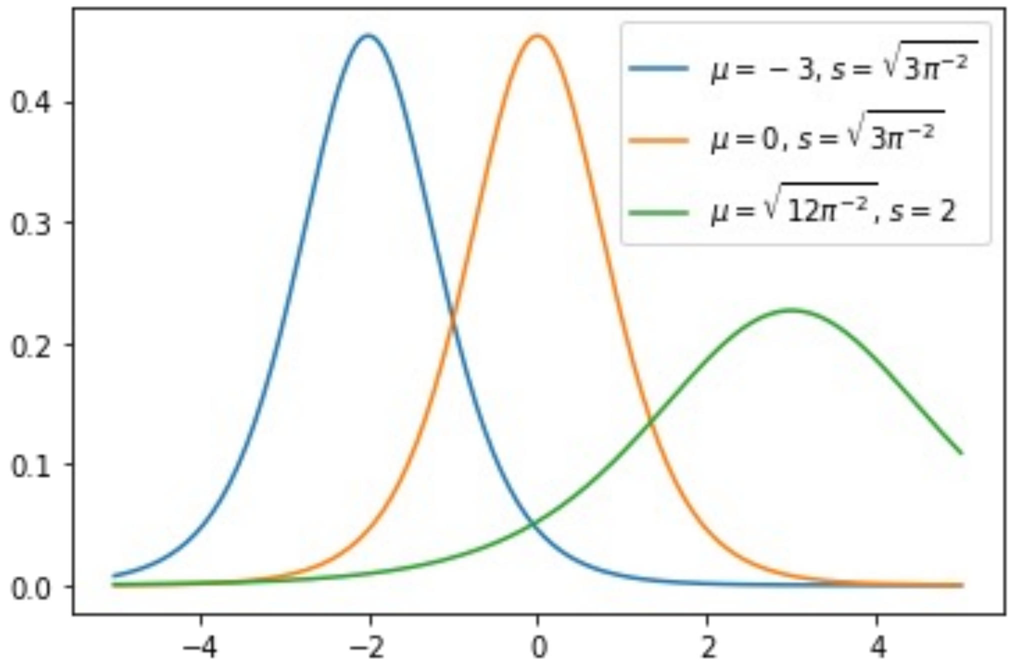
$$\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{e^{-[(x-\mu)^2/\sigma^2]/2}}{\sqrt{2\pi}\ \sigma}.$$

Probability Density Functions – Logistic

- The logistic density and cdf are

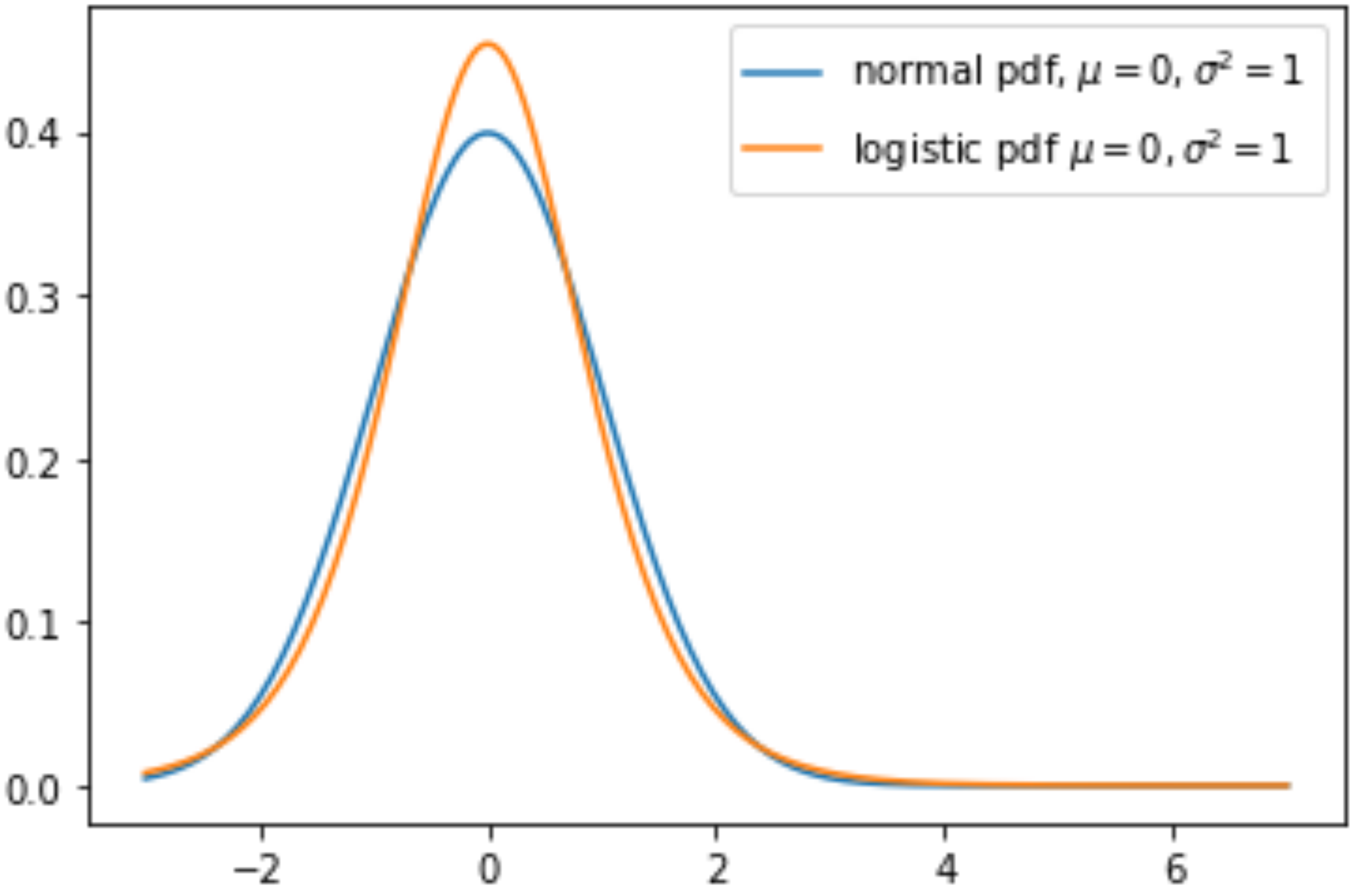
$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{e^{x}}{(1+e^{x})^2} \qquad F(x) = \frac{1}{1+e^{-x}}$$

- The logistic density with ***location parameter*** μ and ***scale parameter*** s is



$$\frac{1}{s} f\left(\frac{x-\mu}{s}\right) = \frac{e^{-(x-\mu)/s}}{s[1+e^{-(x-\mu)/s}]^2}.$$

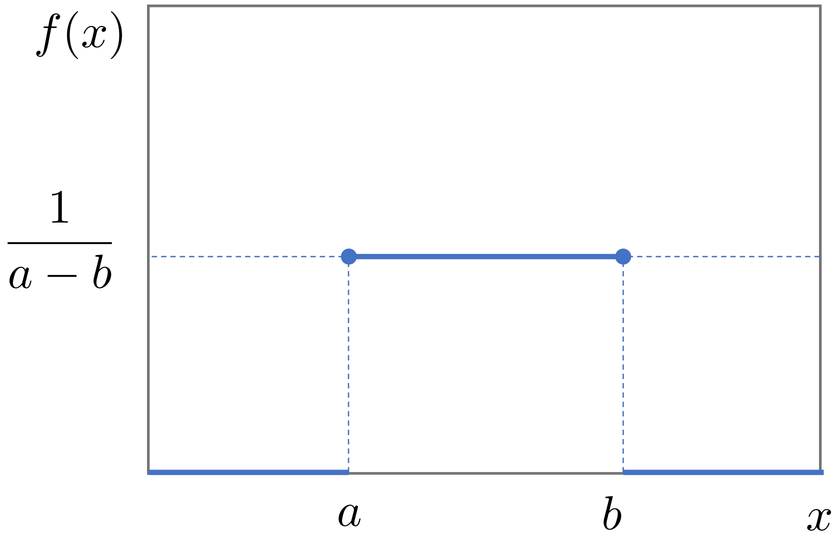
Probability Density Functions – Normal & Logistic Comparison



Common Continuous Random Variables

- uniform: $X \sim \mathrm{U}[a, b]$ for $b > a$

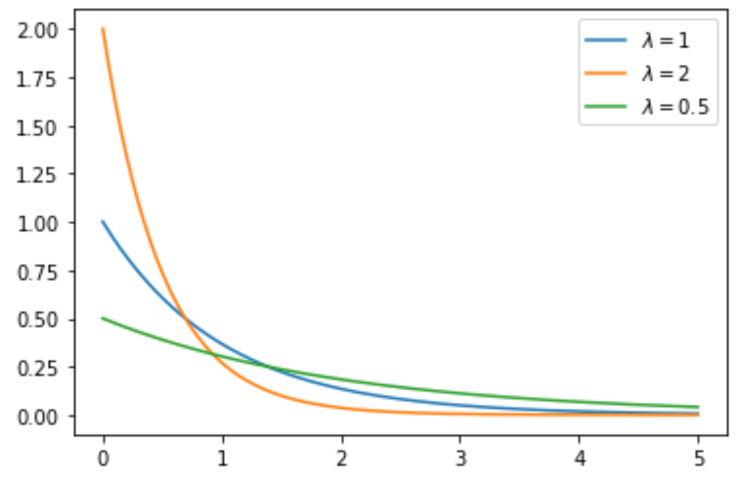
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$



Common Continuous Random Variables

- exponential: $X \sim \text{Exp}(\lambda)$ for $\lambda > 0$

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



- Laplace: $X \sim \text{Lap}(\mu, b)$ for $b > 0$

$$f(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$