Distributions: Beta

Beta distribution

$$B(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

Beta distribution

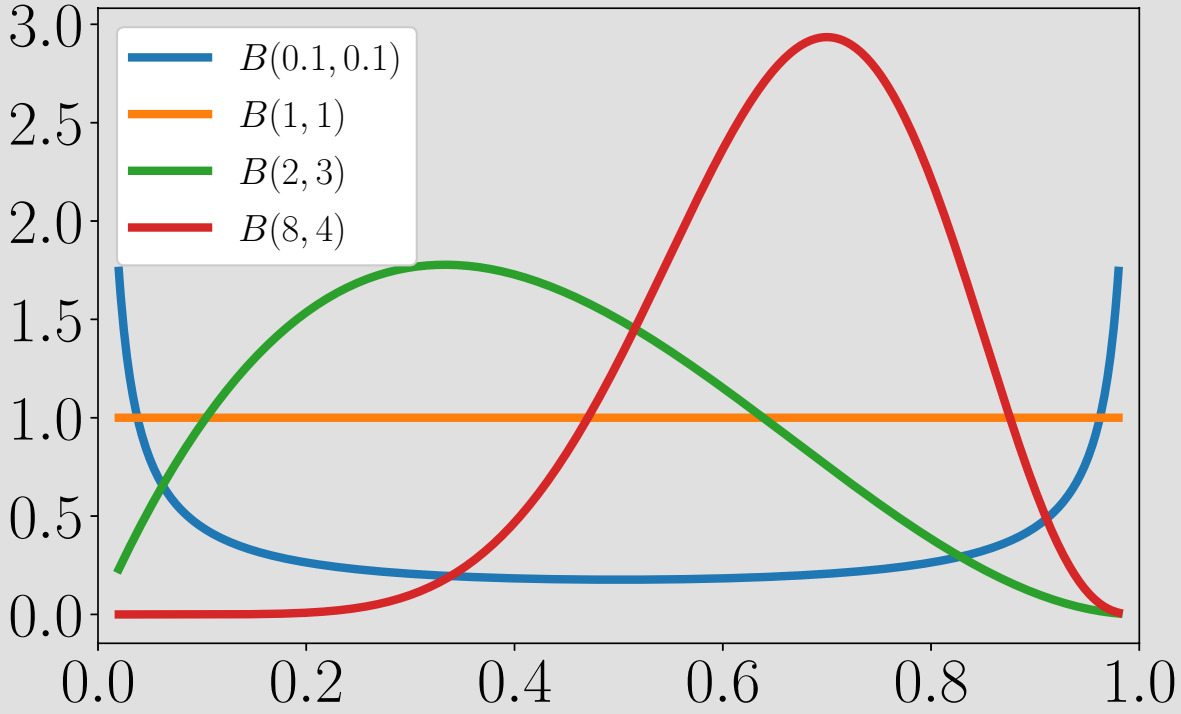
$$B(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

$$x \in [0,1], \quad a,b > 0$$

Beta distribution

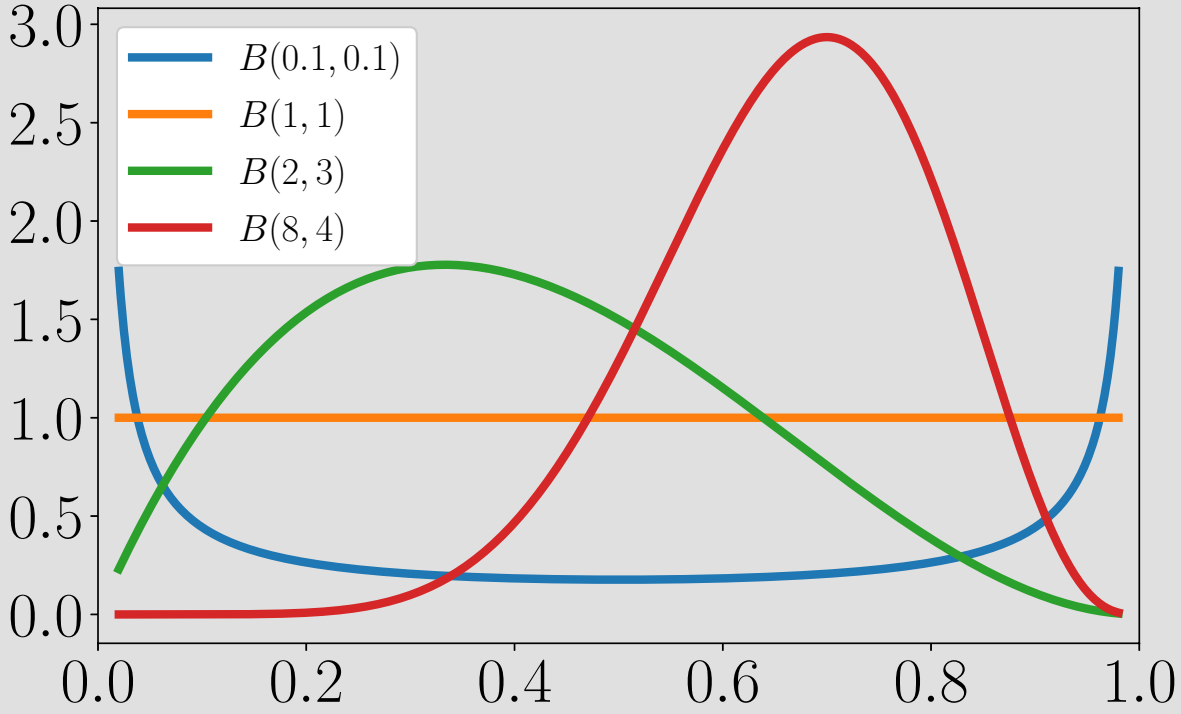
$$B(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

$$x \in [0,1], \quad a,b > 0$$



Beta distribution

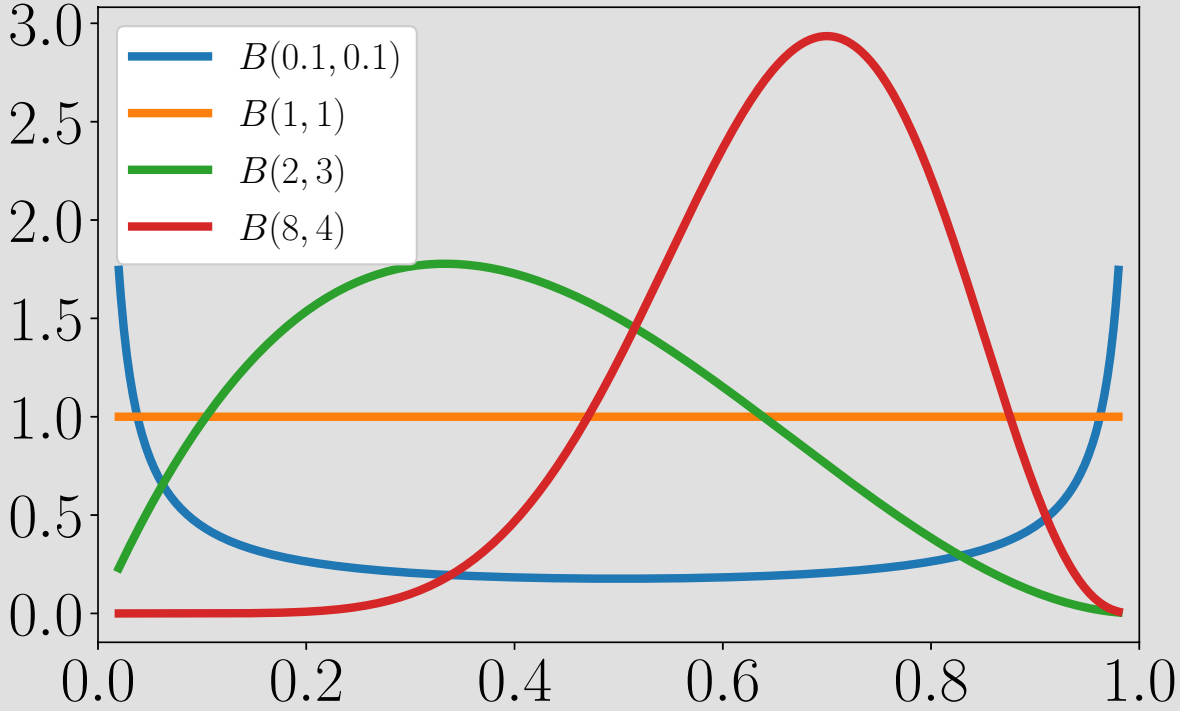
$$B(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$



Beta distribution

$$B(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$



Statistics

$$B(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

$$\mathbb{E}x = \frac{a}{a+b}$$

$$\text{Mode}[x] = \frac{a-1}{a+b-2}$$

$$\text{Var}[x] = \frac{ab}{(a+b)^2(a+b-1)}$$

Example

Movie rank is 0.8 ± 0.1

Example ТЕХНИЧЕСКИЙ СЛАЙД

Movie rank is 0.8 ± 0.1

1 — best movie

0 — Batman & Robin

Example

Movie rank is 0.8 ± 0.1

$$\mathbb{E}x = \frac{a}{a+b} = 0.8$$

$$\mathrm{Var}[x] = \frac{ab}{(a+b)^2(a+b-1)} = 0.1^2$$

Example

Movie rank is 0.8 ± 0.1

$$\mathbb{E}x = \frac{a}{a+b} = 0.8$$

$$\mathrm{Var}[x] = \frac{ab}{(a+b)^2(a+b-1)} = 0.1^2$$

$$\Rightarrow a = 12, b = 3$$

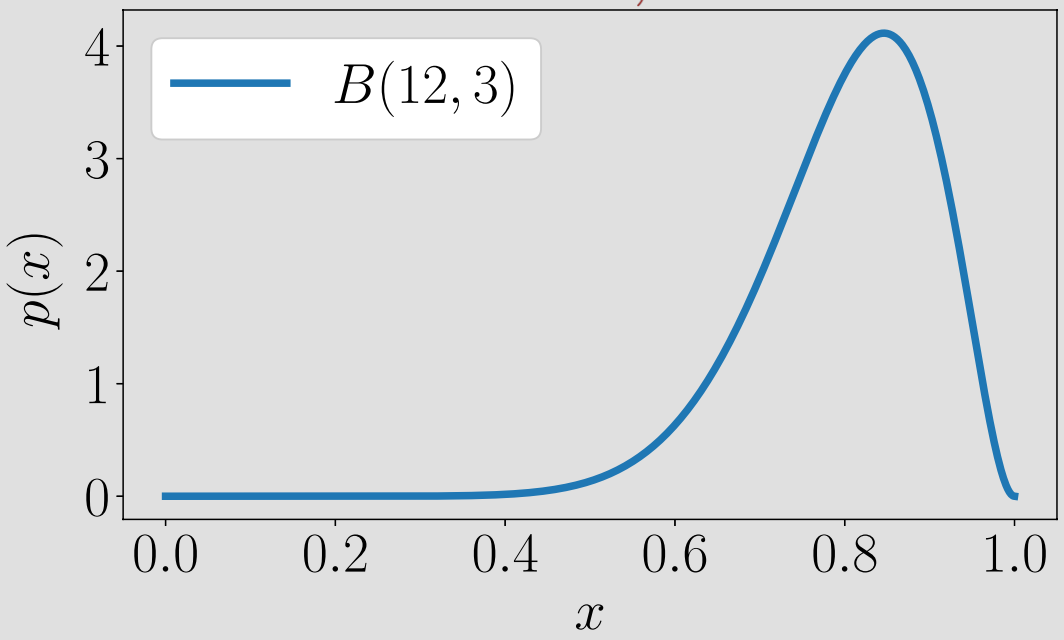
Example

Movie rank is 0.8 ± 0.1

$$\mathbb{E}x = \frac{a}{a+b} = 0.8$$

$$\mathrm{Var}[x] = \frac{ab}{(a+b)^2(a+b-1)} = 0.1^2$$

$$\Rightarrow a = 12, b = 3$$



Example: Bernoulli

Beta prior

$$p(X|\theta) = \theta^{N_1}(1-\theta)^{N_0}$$

Beta prior

$$p(X|\theta) = \theta^{N_1}(1-\theta)^{N_0}$$

$$p(\theta) = B(\theta|a,b) \propto \theta^{a-1}(1-\theta)^{b-1}$$

Beta prior

$$p(X|\theta) = \theta^{N_1}(1-\theta)^{N_0}$$

$$p(\theta) = B(\theta|a,b) \propto \theta^{a-1}(1-\theta)^{b-1}$$

$$p(\theta|X) \propto p(X|\theta)p(\theta)$$

Beta prior

$$p(X|\theta) = \theta^{N_1}(1-\theta)^{N_0}$$

$$p(\theta) = B(\theta|a,b) \propto \theta^{a-1}(1-\theta)^{b-1}$$

$$p(\theta|X) \propto p(X|\theta)p(\theta)$$

$$p(\theta|X) \propto \theta^{N_1}(1-\theta)^{N_0} \cdot \theta^{a-1}(1-\theta)^{b-1}$$

Beta prior

$$p(X|\theta) = \theta^{N_1}(1-\theta)^{N_0}$$

$$p(\theta) = B(\theta|a,b) \propto \theta^{a-1}(1-\theta)^{b-1}$$

$$p(\theta|X) \propto p(X|\theta)p(\theta)$$

$$p(\theta|X) \propto \theta^{N_1}(1-\theta)^{N_0} \cdot \theta^{a-1}(1-\theta)^{b-1}$$

$$p(\theta|X) \propto \theta^{N_1+a-1}(1-\theta)^{N_0+b-1}$$

Beta prior

$$p(X|\theta) = \theta^{N_1}(1-\theta)^{N_0}$$

$$p(\theta) = B(\theta|a,b) \propto \theta^{a-1}(1-\theta)^{b-1}$$

$$p(\theta|X) \propto p(X|\theta)p(\theta)$$

$$p(\theta|X) \propto \theta^{N_1}(1-\theta)^{N_0} \cdot \theta^{a-1}(1-\theta)^{b-1}$$

$$p(\theta|X) \propto \theta^{N_1+a-1}(1-\theta)^{N_0+b-1}$$

$$p(\theta|X) = B(N_1+a, N_0+b)$$

Summary

Summary

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

Summary

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

Summary

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

Summary

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

Pros and cons

Pros:

Pros and cons

Pros:

- Exact posterior

Pros and cons

Pros:

- Exact posterior
- Easy for on-line learning

E.g. $p(\theta|X) = B(N_1 + a, N_0 + b)$

Pros and cons

Pros:

- Exact posterior
- Easy for on-line learning

E.g. $p(\theta|X) = B(N_1 + a, N_0 + b)$

Cons:

Pros and cons

Pros:

- Exact posterior
- Easy for on-line learning

E.g. $p(\theta|X) = B(N_1 + a, N_0 + b)$

Cons:

- Conjugate prior may be inadequate