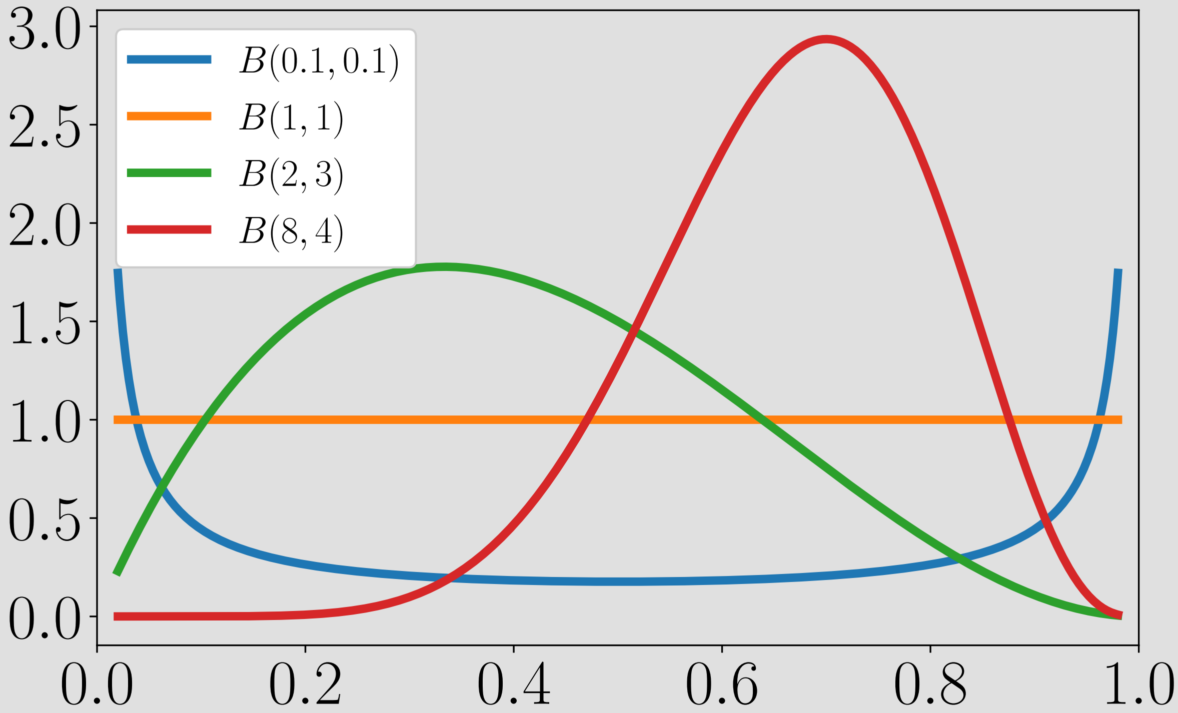
# Beta distribution

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

this is a normalization constant, that is expressed in terms of the gamma function (factorial!)

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

Recall that Gamma(n) = (n-1)!



# Statistics

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

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

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

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

# Example ТЕХНИЧЕСКИЙ СЛАЙД

Suppose we have a website that ranks movies from 0 to 1.

Movie rank is $0.8 \pm 0.1$


1 — best movie

0 — Batman & Robin

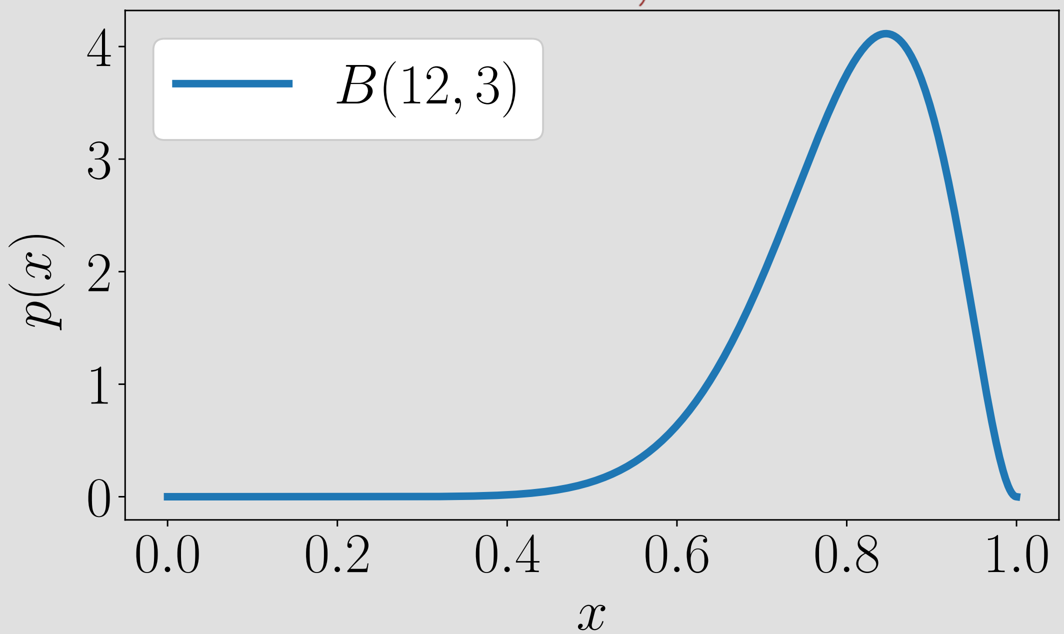
# Example

Movie rank is 0.8 ± 0.1

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

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

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



# Example: Bernoulli

We study the bernoulli variables because the beta distribution is conjugate to the bernoulli likelihood.

# Beta prior

Recall that the likelihood is the probability that this theta paramaeter fits the values X.
For each X that is 1, we use probability theta.

We have datapoints that can either be 0 or 1s.

N1 is the number of 1's appeared in X. N0 = the number of 0's appeared in X.

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

<- This here is the bernoulli likelihood.

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

Now, try to prove that p(theta | X) is proportional to the likelihood x prior.

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

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

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

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

It is ! the posterior is the beta distribution again!

# Summary

# Summary

Essentially, we want to compute the posterior.

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

# Summary

But, in order to do this, we have to compute all the possible values of X and their probabilities P(X).
This is really infeasible.

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

# Summary

So if we have a p(theta) (prior) that is conjugate to the likelihood, then the posterior will be proportional to the prior, and we can skip p(X) since this will just be a 'constant'.

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

# 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

Why inadequate?