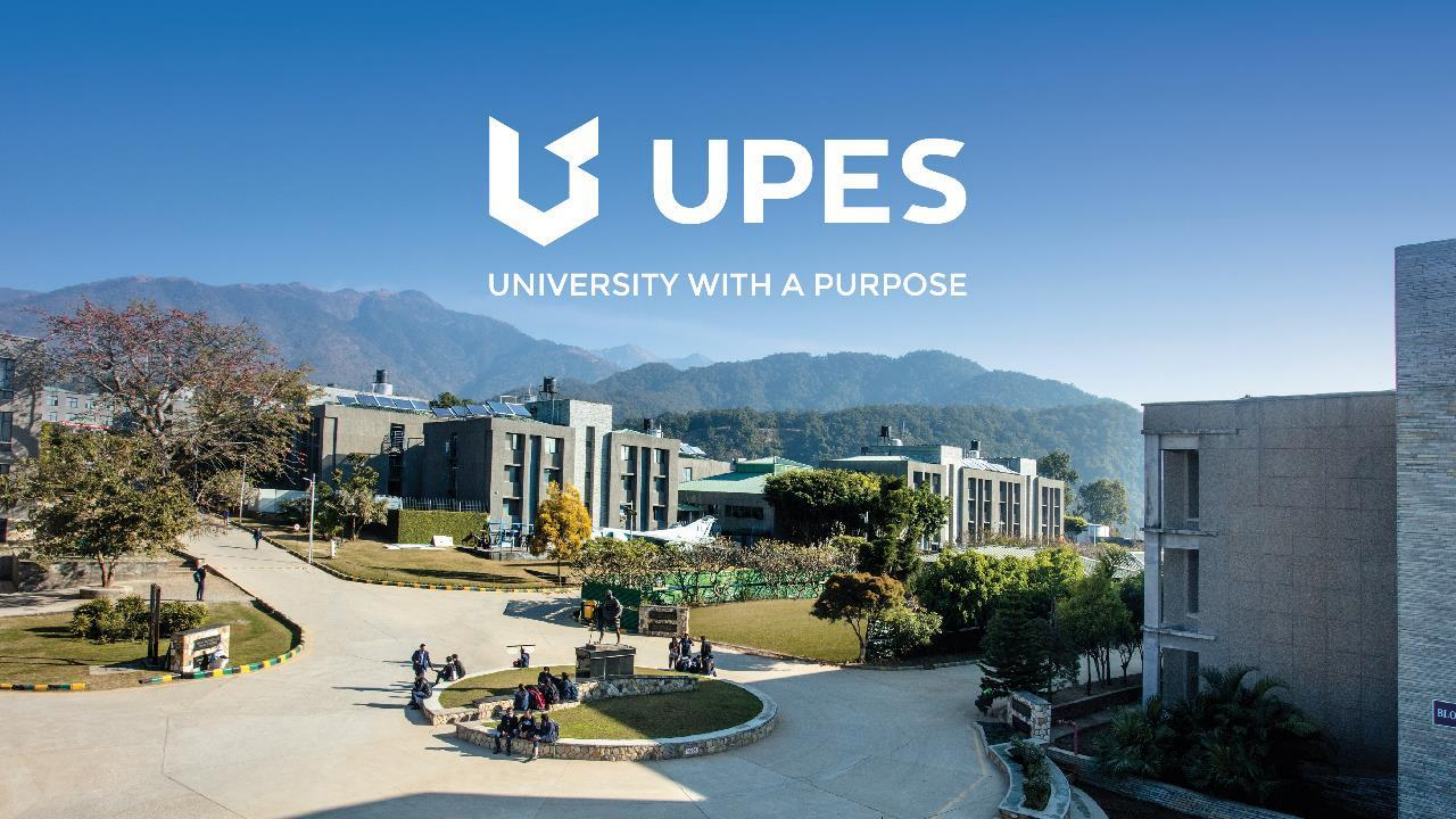
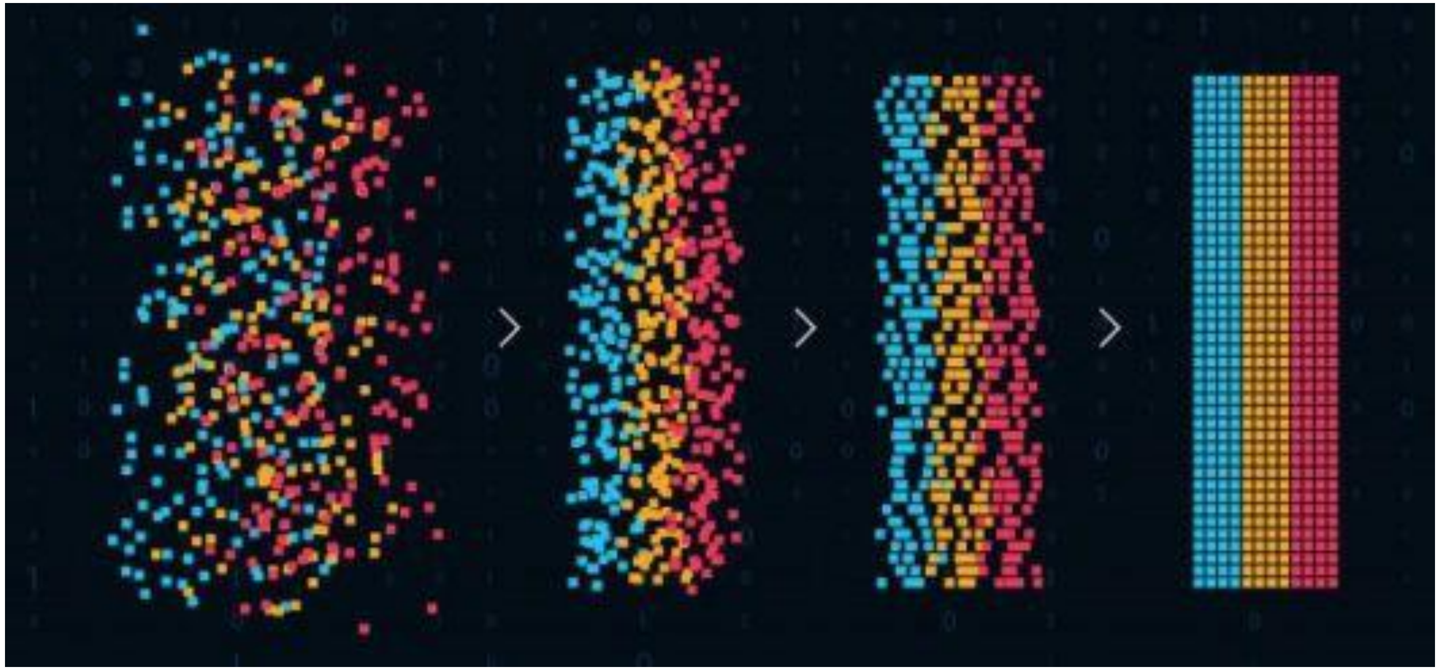

UPES
UNIVERSITY WITH A PURPOSE

Pattern and Anomaly Detection



Source: Edureka

B. Tech., CSE + AI/ML

Dr Gopal Singh Phartiyal

16/09/2021

Binary variables

$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

$$\operatorname{var}[f] = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

For discrete random binary variable

Coin flipping: heads =1, tails = 0

$$p(x = 1|\mu) = \mu$$

- Probability distribution over x

$$\operatorname{Bern}(x|\mu) = \mu^x (1-\mu)^{1-x}$$

Bernoulli Distribution

- Distribution statistics

$$\mathbb{E}[x] = \mu$$

$$\operatorname{var}[x] = \mu(1-\mu)$$

Bernoulli Distribution

- Likelihood function for Bernoulli
- Given: $\mathcal{D} = \{x_1, \ldots, x_N\}$, m heads (1), $N - m$ tails (0)

$$p(\mathcal{D}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}$$

- This can be interpreted as likelihood function
- In order to find μ that maximizes the likelihood

$$\ln p(\mathcal{D}|\mu) = \sum_{n=1}^{N} \ln p(x_n|\mu) = \sum_{n=1}^{N} \{x_n \ln \mu + (1 - x_n) \ln(1-\mu)\}$$

The probability of landing heads is given, in this maximum likelihood framework, by the fraction of observations of heads in the data set.

$$\mu_{\mathrm{ML}} = \frac{1}{N} \sum_{n=1}^{N} x_n = \frac{m}{N}$$

Simple mean

$$\mu_{\mathrm{ML}} = \frac{m}{N}$$

Summaries

- We have seen that the joint probability of two independent events is given by the product of the marginal probabilities for each event separately.

$$p(\mathbf{x}|\mu, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}\left(x_n|\mu, \sigma^2\right).$$

- In terms of parameters, this function is called likelihood function

Summaries

- One common criterion for determining the parameters in a probability distribution using an observed data set is to find the parameter values that maximize the likelihood function.

- This might seem like a strange criterion because, from our foregoing discussion of probability theory, it would seem more natural to maximize the probability of the parameters given the data, not the probability of the data given the parameters..

Challenges with Bernoulli and maximizing likelihood

- The probability of landing heads is given, in this maximum likelihood framework, by the fraction of observations of heads in the data set.

- Example: $$\mathcal{D} = \{1, 1, 1\} \rightarrow \mu_{\mathrm{ML}} = \frac{3}{3} = 1$$

- Prediction: *all* future tosses will land heads up

Overfitting to D

- Possible solution: Prior distribution over parameter

Binary Variables and Binomial Distribution

- N coin flips:

$$p(m \text{ heads}|N, \mu)$$

- Binomial Distribution

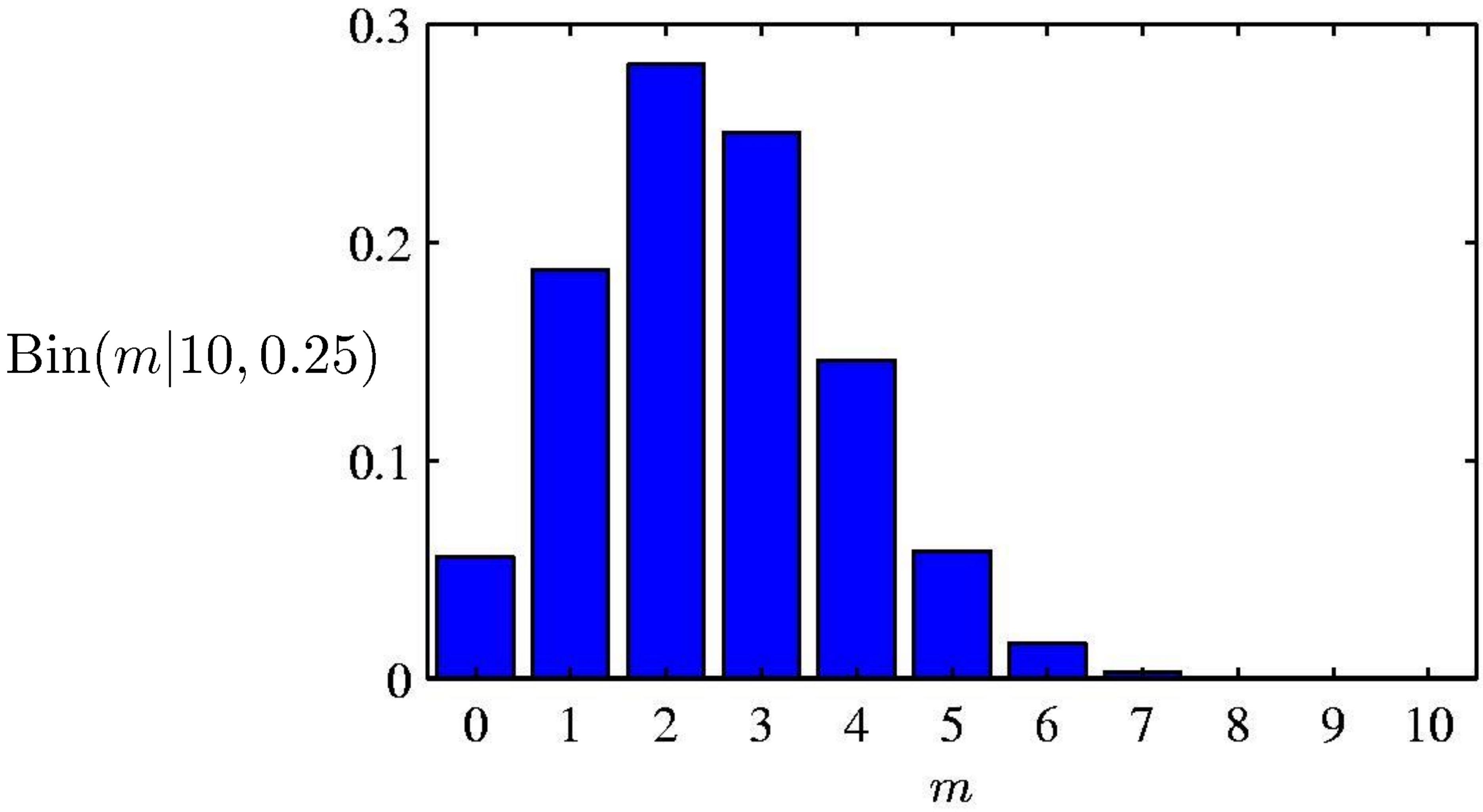
$$\mathrm{Bin}(m|N, \mu) = \binom{N}{m} \mu^m (1-\mu)^{N-m}$$

$$\binom{N}{m} \equiv \frac{N!}{(N-m)!m!}$$

$$\mathbb{E}[m] \equiv \sum_{m=0}^{N} m\mathrm{Bin}(m|N, \mu) = N\mu$$

$$\mathrm{var}[m] \equiv \sum_{m=0}^{N} (m - \mathbb{E}[m])^2 \, \mathrm{Bin}(m|N, \mu) = N\mu(1-\mu)$$

Example: Binary variables and Binomial Distribution



Next time: Multinomial Distribution

Thank You

UPES

UNIVERSITY WITH A PURPOSE

Binary Variables and Beta Distribution

- Distribution over

$$\mu \in [0, 1]$$

$$\begin{aligned}
\mathrm{Beta}(\mu|a, b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1} \\
\mathbb{E}[\mu] &= \frac{a}{a+b} \\
\mathrm{var}[\mu] &= \frac{ab}{(a+b)^2(a+b+1)}
\end{aligned}$$

$$\Gamma(x) \equiv \int_0^\infty u^{x-1} e^{-u}\,\mathrm{d}u.$$

Beta function

$$
\begin{aligned}
p(\mu|a_0, b_0, \mathcal{D}) &\propto p(\mathcal{D}|\mu)p(\mu|a_0, b_0) \\
&= \left(\prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}\right) \mathrm{Beta}(\mu|a_0, b_0) \\
&\propto \mu^{m+a_0-1}(1-\mu)^{(N-m)+b_0-1} \\
&\propto \mathrm{Beta}(\mu|a_N, b_N)
\end{aligned}
$$

$$
a_N = a_0 + m \qquad b_N = b_0 + (N-m)
$$

The Beta distribution provides the *conjugate* prior for the Bernoulli distribution.