Royal Institute of Technology

BETA & BETA-BINOMIAL

BINOMIAL - IID BERNOULLI

- Likelihood
  - N_1 heads and N_0 tails
  - p(head) = θ
  - sequence

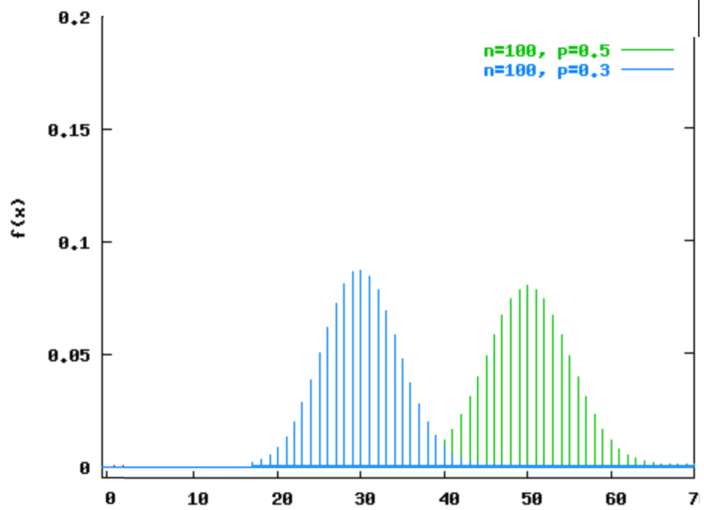
$$p(D) = \theta^{N_1}(1-\theta)^{N_0}$$

  - counts

$$p(D) = \binom{N_1+N_0}{N_1}\theta^{N_1}(1-\theta)^{N_0}$$

  - MLE

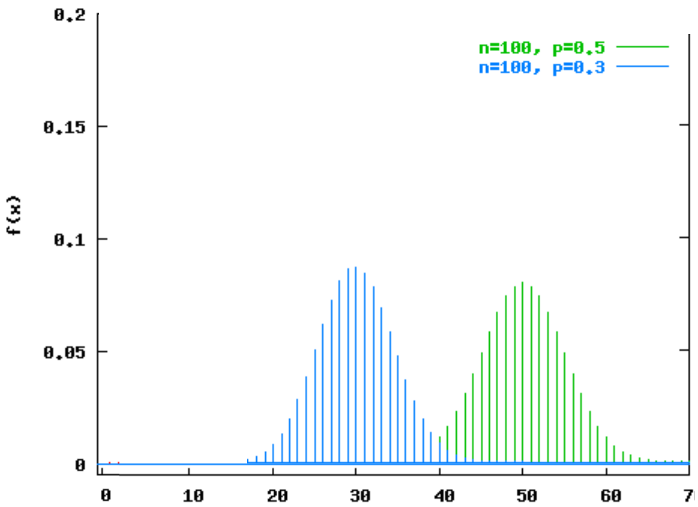
$$\theta = \frac{N_1}{N_1+N_0}$$



SUFFICIENT STATISTICS

- Maximum Likelihood (ML) estimates maximize the probability of the data
  - here $\max_\theta p(N_1, N_0|\theta)$
- The pair N_1, N_0 is a sufficient statistic for our coin model
  - i.e., given those ML estimate or any other follows

PRIOR FOR CATEGORICAL AND BERNOULLI - FIRST BETA-BINOMIAL



- ★ Assumption: we don't know θ
- ★ We need a prior over the outcome probabilities
  - ★ N_1 heads and N_0 tails
  - ★ p(head) denoted θ

$$p(D) \propto \theta^{N_1}(1-\theta)^{N_0}$$

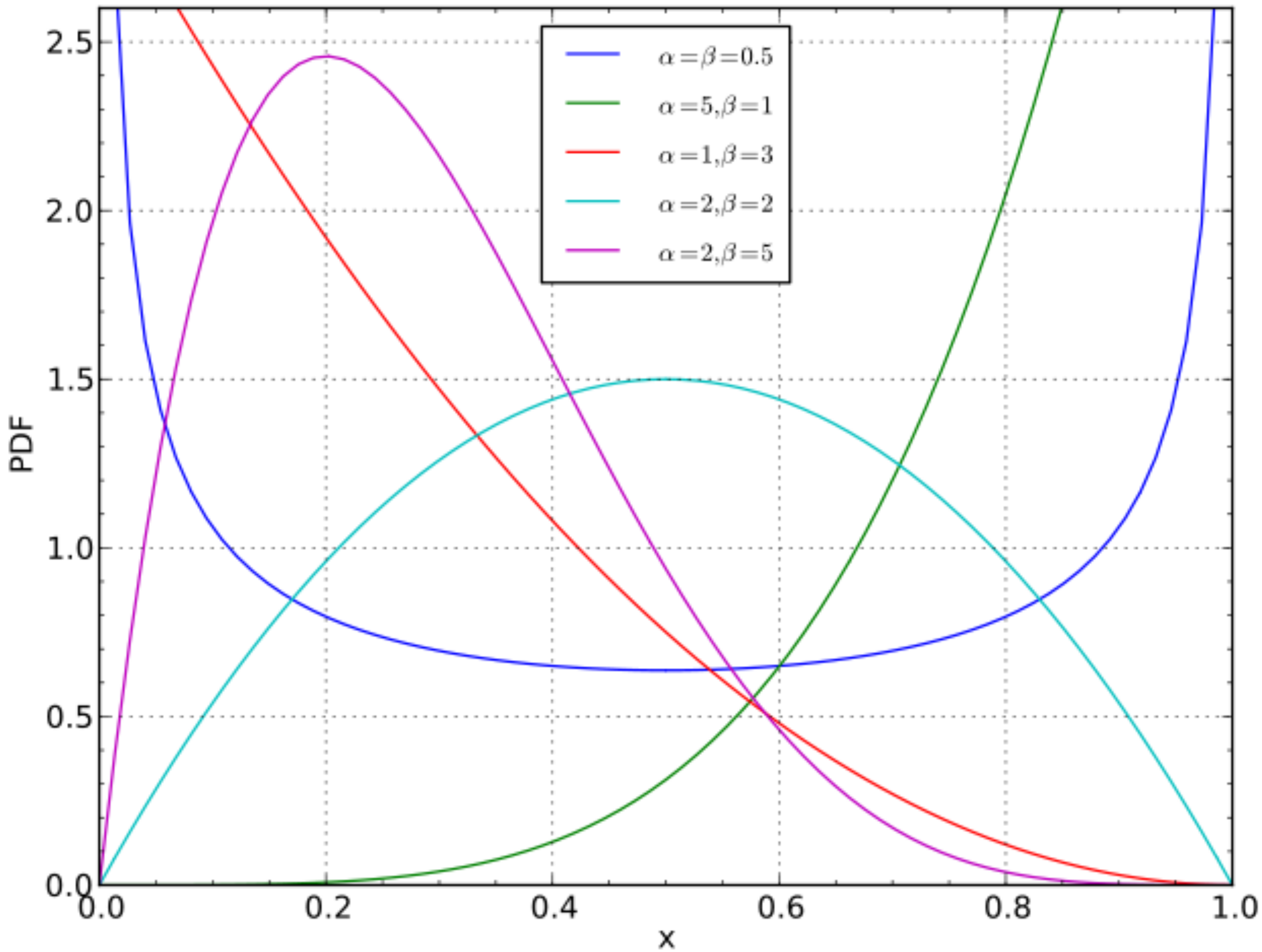
BETA DISTRIBUTION

- PDF $\mathrm{Beta}(\theta \mid a, b) = \dfrac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1}$
  - where
  $$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$
  - and
  $$\Gamma(a) = (a-1)\Gamma(a-1)$$
- In particular, for integer n
$$\Gamma(n) = (n-1)!$$
- Expectation, mode, and variance

$$\frac{a}{a+b} \qquad \frac{a-1}{a+b-2} \text{ for a,b>1} \qquad \frac{ab}{(a+b)^2(a+b+1)}$$

HYPER-PARAMETERS

- Parameters a and b for prior called
  - hyperparameters
  - pseudocounts
- Prior's effective sample size is $a + b$
- Prior that gives posterior of the same sort is called conjugate



BETA DISTRIBUTION

BETA FOR SOME PARAMETER CHOICES

BETA FOR SOME PARAMETER CHOICES

BETA-BINOMIAL

- Beta distribution up to a constant $p(\theta|a,b) \propto \theta^{a-1}(1-\theta)^{b-1}$

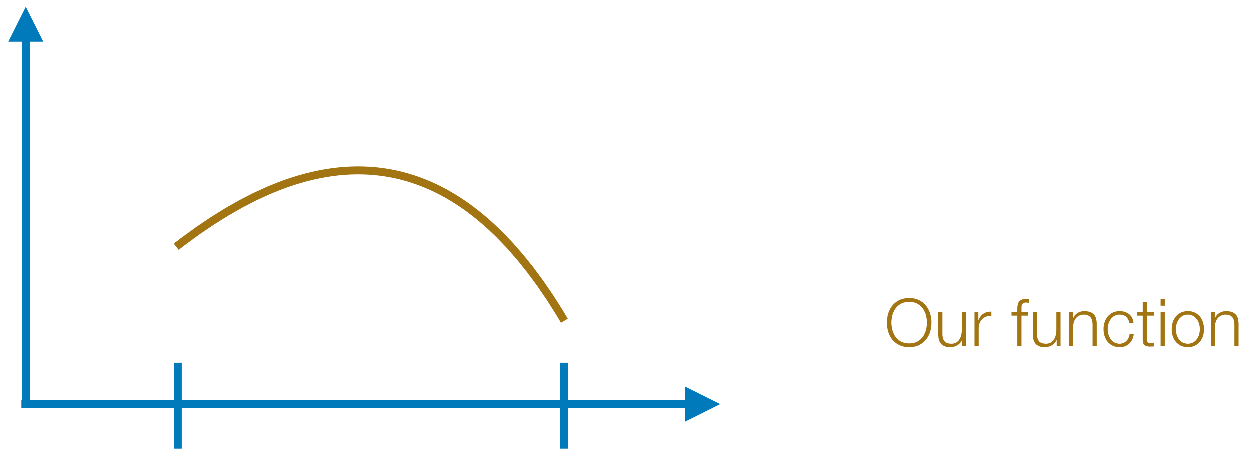
  means there is a constant c s/t $p(\theta|a,b) = c\theta^{a-1}(1-\theta)^{b-1}$

- Posterior $p(\theta|D) \propto \mathrm{Beta}(\theta|N_1 + a, N_0 + b)$

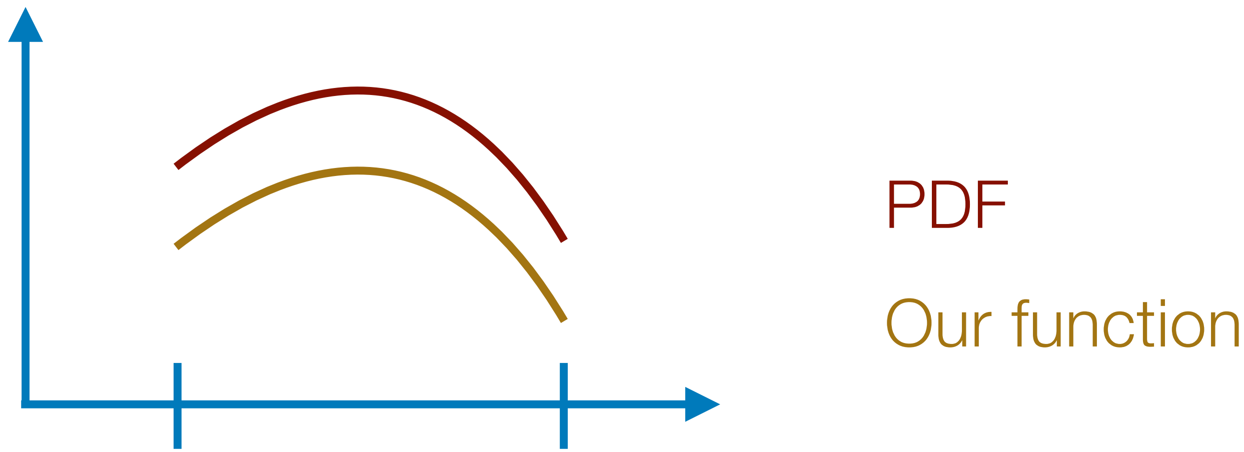
- Beta is a *conjugate* prior for Binomial

- Maximum posterior (MAP), posterior mean, posterior variance

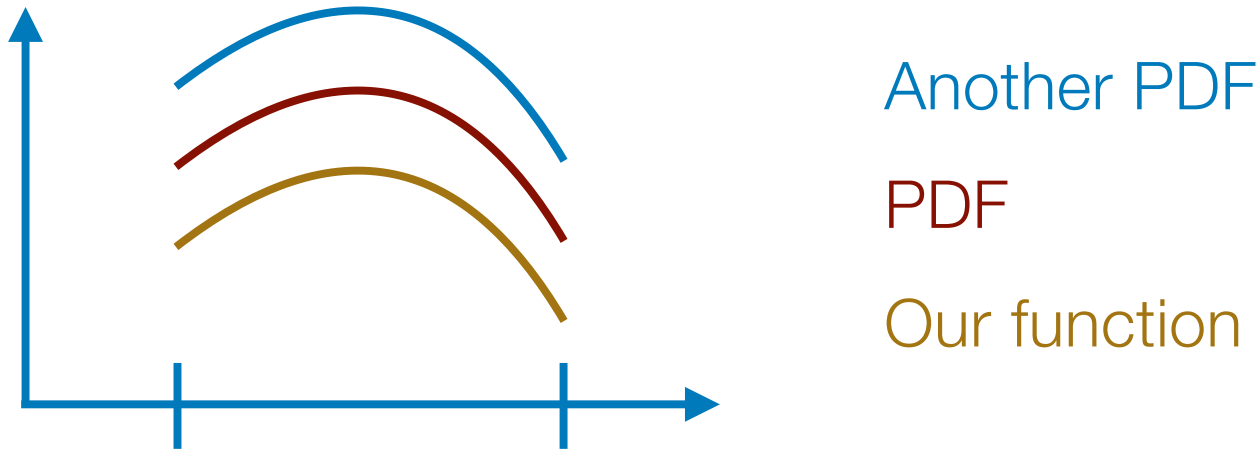
IDENTIFYING A DISTRIBUTION



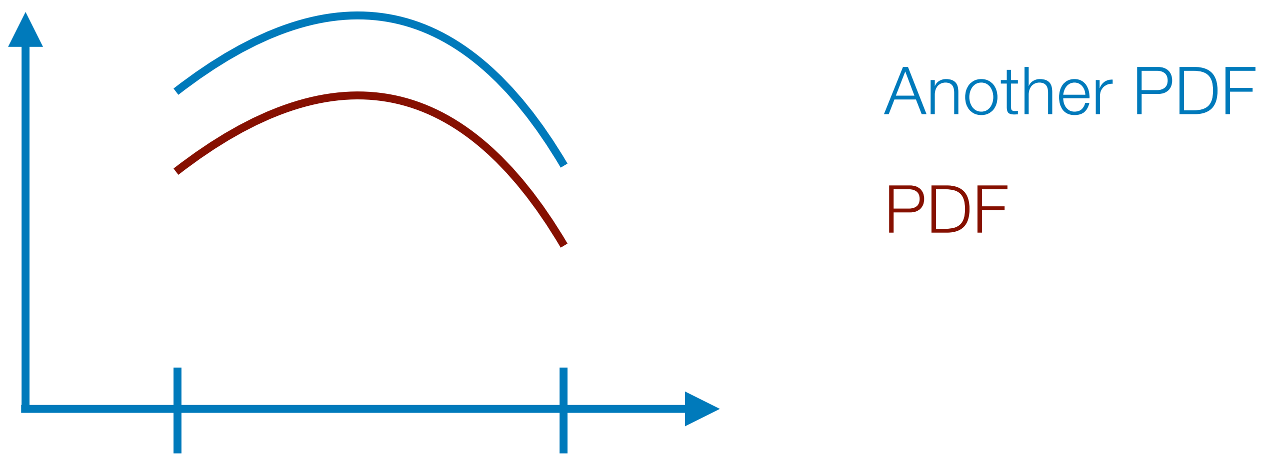
OUR PDF



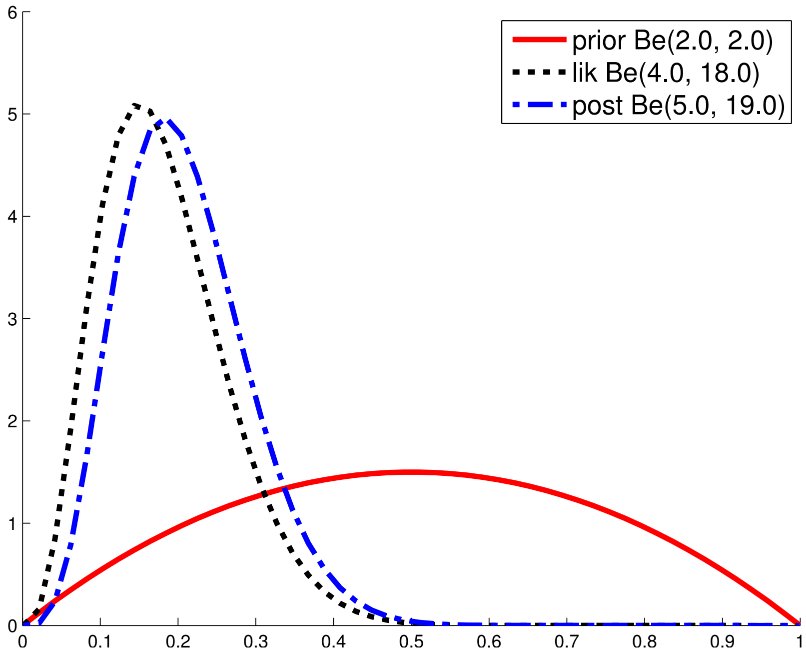
ANOTHER PDF



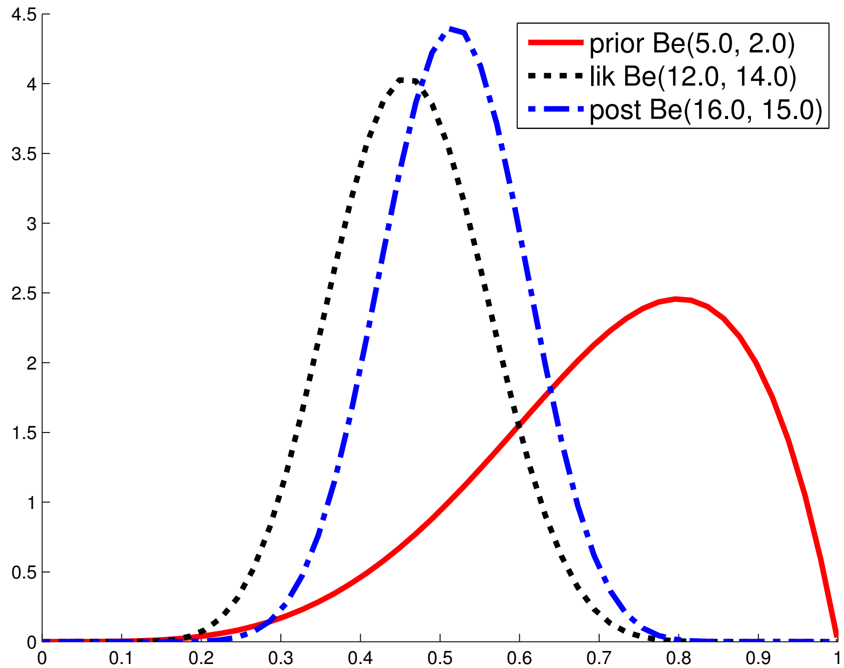
BOTH INTEGRALS CANNOT BE 1



BETA-BINOMIAL: PRIOR, LIKELIHOOD, POSTERIOR



N_1=3, N_0= 17



N_1=11, N_0= 13

As before: N_1 heads and N_0 tails

Uniform prior, i.e., a=b=1, gives

$$p(x = 1 \mid D) = \frac{N_1 + 1}{N_0 + N_1 + 2}$$

Black Swan “paradox”

LAPLACE’S RULE OF SUCCESSION