

# Story: Beta-Binomial Conjugacy

Problem:

$n$  is Finite!

$n$  tosses:

$k$  of  $n$  tosses are landing heads.

Estimation of  $p$ :  $\hat{p} = \frac{k}{n}$

$n=5, k=1, \hat{p}=\frac{1}{5}$

$n=100, k=2, \hat{p}=\frac{2}{100}$

$n=5, k=5, \hat{p}=1$

$n=10^7, k=10^7, \hat{p}=1$

$n=3, k=3, \hat{p}=1$

We have a coin that lands Heads with probability  $p$ , but we don't know what  $p$  is. Our goal is to infer the value of  $p$  after observing the outcomes of  $n$  tosses of the coin. The larger that  $n$  is, the more accurately we should be able to estimate  $p$ .

$n=2, k=2, \hat{p}=1$

# Bayesian Inference

- Treats all unknown quantities as random variables.
- In the Bayesian approach, we would treat the unknown probability  $p$  as a random variable and give  $p$  a distribution.
- This is called a **prior distribution**, and it reflects our uncertainty about the true value of  $p$  before observing the coin tosses.
- After the experiment is performed and the data are gathered, the prior distribution is updated using Bayes' rule; this yields the **posterior distribution**, which reflects our new beliefs about  $p$ .

# Story: Beta-Binomial Conjugacy

①  $p$ : unknown, model it as a r.v.  $\in [0, 1]$

Beta distribution.  $p \sim \text{Beta}(a, b)$   hyperparameters.

$X$ : # of heads in  $n$  tosses of coin.

$X | p=p \sim \text{Bin}(n, p)$  data model  
generative model.

②  $f(p)$ : prior distribution PDF of  $p$ .

experiments:  $X=k$ .

$f(p|X=k)$ : posterior distribution PDF of  $p$ . (after observation of  $k$  heads out of  $n$  tosses)

# Story: Beta-Binomial Conjugacy

$$f(p|X=k) = \frac{\text{Prob.}(X=k|p) \underline{f(p)}}{\text{Prob.}(X=k)} = \frac{\binom{n}{k} \underbrace{p^k}_{\sim} \underbrace{(1-p)^{n-k}}_{\sim} \cdot \underbrace{\frac{1}{\text{Beta}(a,b)}}_{\sim} \underbrace{p^{a-1}}_{\sim} \underbrace{(1-p)^{b-1}}_{\sim}}{\text{Prob.}(X=k) \dots}$$

$$\text{Prob.}(X=k) = \int_0^1 \text{Prob.}(X=k|p) f(p) dp = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} f(p) dp$$

③  $f(p|X=k)$  is a function of  $p$ . [every item that does not depend on  $p$  is a constant]

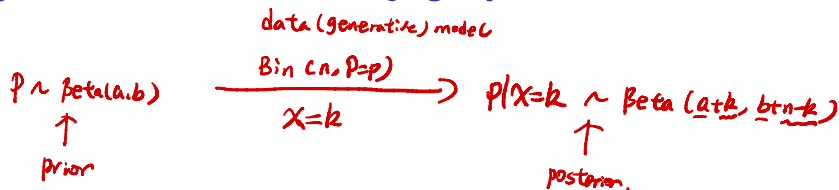
$$f(p|X=k) \propto \frac{p^{k+a-1} \cdot (1-p)^{n-k+b-1}}{( = p^{k+a-1} (1-p)^{n-k+b-1} \cdot c )}$$

$\text{Beta}(k+a, n-k+b)$

$$\begin{aligned} Y &\sim \text{Beta}(a, b) \\ \text{PDF of } Y &\propto p^{a-1} (1-p)^{b-1} \end{aligned}$$

$$\Rightarrow p|X=k \sim \text{Beta}(a+k, b+n-k)$$

## Story: Beta-Binomial Conjugacy



# Story: Beta-Binomial Conjugacy

- Furthermore, notice the very simple formula for updating the distribution of  $p$ .
- We just add the number of observed successes,  $k$ , to the first parameter of the Beta distribution.
- We also add the number of observed failures,  $n - k$ , to the second parameter of the Beta distribution.
- So  $a$  and  $b$  have a concrete interpretation in this context:
  - ▶  $a$  as the number of prior successes in earlier experiments
  - ▶  $b$  as the number of prior failures in earlier experiments
  - ▶  $a, b$ : pseudo counts

# Mean vs. Bayesian Average

$$Y \sim \text{Beta}(a, b), E(Y) = \frac{a}{a+b}$$

$$P|X=k \sim \text{Beta}(a+k, b+n-k)$$

$$E(P|X=k) = \frac{a+k}{a+k+b+n-k} = \frac{a+k}{a+b+n}$$

- Infer the value of  $p$  (probability of coin lands heads)
- Observed  $k$  heads out of  $n$  tosses of the coin
- Mean:  $\frac{k}{n}$
- Bayesian Average:  $E(p|X=k) = \frac{a+k}{a+b+n}$
- Suppose the prior distribution is  $\text{Unif}(0,1)$ :  $a=1, b=1$
- Bayesian Average:  $\frac{k+1}{n+2}$
- When  $k=n$ , we have:  $1$  (mean) vs.  $\frac{n+1}{n+2}$  (Bayesian average)

$$\begin{array}{l} n \rightarrow \infty \\ k=n \end{array} \quad \frac{n+1}{n+2} \rightarrow 1 \quad \checkmark$$

# Story: Beta-Binomial Conjugacy

If we have a Beta prior distribution on  $p$  and data that are conditionally Binomial given  $p$ , then when going from prior to posterior, we don't leave the family of Beta distributions. We say that **the Beta is the conjugate prior of the Binomial.**