

Bernoulli model with four Beta priors

Likelihood, priors, posteriors, prior predictives and Bayes factors

Hedibert Freitas Lopes

Inspire

Disclaimer: these slides (derivations, figures, computations and $\LaTeX$ code) were prepared with the assistance of Claude Opus 5.5 (Anthropic), an AI model, from the prompt reproduced on the next slides. Content reviewed by the author, who is the sole responsible for the document.

The prompt given to the AI (1/2)

Let $x_1, \dots, x_n$ be a sample of zeros and ones.

Assume that the data comes from a Bernoulli model, ie.

$x_1, \dots, x_n$ are iid Bernoulli(θ), where $\Pr(x_i=1)=\theta$, for θ in $(0,1)$.

Let us compare the four posterior distributions of θ based on four priors and keeping the likelihood (the model) the same.

Prior 1: $\theta \sim \text{Beta}(a_1, b_1)$, $a_1=b_1=1 \Rightarrow U(0,1)$

Prior 2: $\theta \sim \text{Beta}(a_2, b_2)$, $a_2=1, b_2=5$

Prior 3: $\theta \sim \text{Beta}(a_3, b_3)$, $a_3=2, b_3=10$

Prior 4: $\theta \sim \text{Beta}(a_4, b_4)$, $a_4=10, b_4=2$

We know that $y=x_1+x_2+\dots+x_n=9$ and $n=27$.

The prompt given to the AI (2/2)

- 1) Create one plot with the likelihood (normalized), prior and posterior for each prior, separately
- 2) Create one plot with only the 4 posteriors and a vertical line at $y=9/27=1/3$ (the MLE)
- 3) Show the derivation of the prior predictive $p(x_1, \dots, x_n | \text{Prior } i)$ for $i=1,2,3,4$.
- 4) When $y=9$ and $n=27$, compute Bayes factors and posterior model probabilities based on a uniform prior for the 4 models (4 priors).

Observation: Obviously, the four priors are $\text{Beta}(a,b)$ priors for different values of a and b . Therefore 3) and 4) can be derived generically and subsequently apply for the specific prior specifications.

Create slides in the latex/beamer format.

Why it works: model, data and priors stated precisely; numbered deliverables; a hint on structure (derive generically, then specialize); output format named.

Setup

- Data: $x_1, \dots, x_n \in \{0, 1\}$, with $x_i \mid \theta \stackrel{iid}{\sim} \text{Bernoulli}(\theta)$, $\theta \in (0, 1)$.
- Sufficient statistic: $y = \sum_{i=1}^n x_i$. Observed: $y = 9$, $n = 27$, so $\hat{\theta}_{MLE} = y/n = 1/3$.
- Likelihood:

$$L(\theta) = p(x_1, \dots, x_n \mid \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^y (1 - \theta)^{n-y}.$$

- Four priors, same likelihood:

	a	b	$E(\theta)$	comment
Prior 1	1	1	0.500	$U(0, 1)$
Prior 2	1	5	0.167	mass near 0
Prior 3	2	10	0.167	same mean, more concentrated
Prior 4	10	2	0.833	mass near 1

Conjugacy: Beta prior $\Rightarrow$ Beta posterior

With $\theta \sim \text{Beta}(a, b)$, $p(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a, b)}$, where $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$:

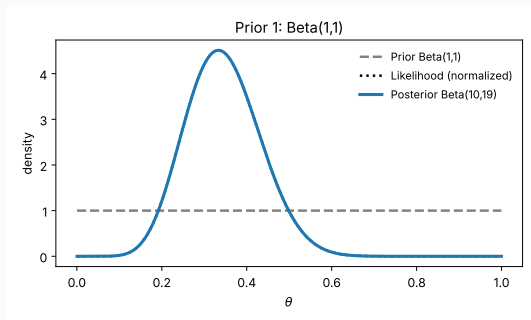
$$p(\theta \mid x_{1:n}) \propto \theta^y (1-\theta)^{n-y} \theta^{a-1} (1-\theta)^{b-1} = \theta^{a+y-1} (1-\theta)^{b+n-y-1},$$

so

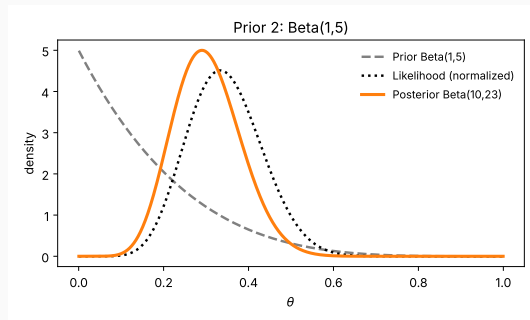
$$\theta \mid x_{1:n} \sim \text{Beta}(a+y, b+n-y).$$

- Normalized likelihood: $L(\theta) / \int_0^1 L(\theta) d\theta$ is the $\text{Beta}(y+1, n-y+1) = \text{Beta}(10, 19)$ density (= posterior under Prior 1).
- Posterior mean: $E(\theta \mid x_{1:n}) = \frac{a+y}{a+b+n} = w \frac{a}{a+b} + (1-w) \frac{y}{n}$, with $w = \frac{a+b}{a+b+n}$.

1) Prior, likelihood and posterior: Priors 1 and 2

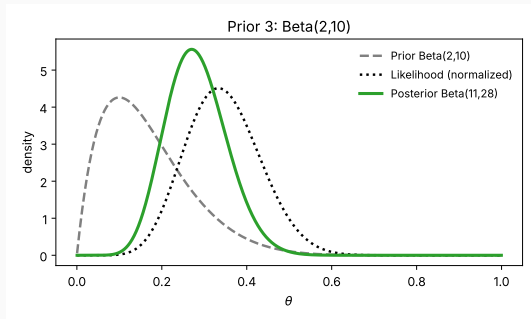


$\text{Beta}(1, 1) \rightarrow \text{Beta}(10, 19)$

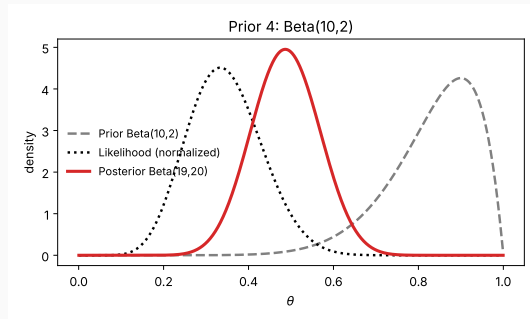


$\text{Beta}(1, 5) \rightarrow \text{Beta}(10, 23)$

1) Prior, likelihood and posterior: Priors 3 and 4

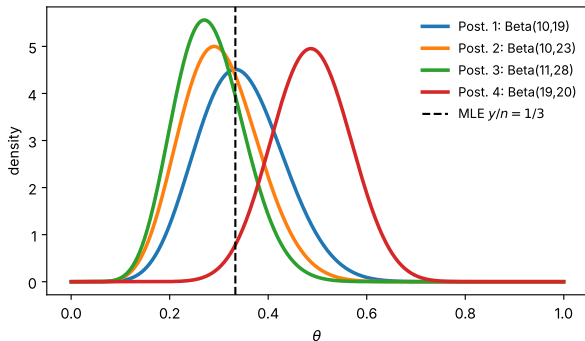


Beta(2, 10) $\rightarrow$ Beta(11, 28)



Beta(10, 2) $\rightarrow$ Beta(19, 20)

2) The four posteriors and the MLE



	Mean	SD	95% CI
P1	0.345	0.087	(0.186, 0.524)
P2	0.303	0.079	(0.161, 0.467)
P3	0.282	0.071	(0.154, 0.431)
P4	0.487	0.079	(0.334, 0.642)

Priors 2–3 pull θ below $1/3$; Prior 4 pulls it well above. With $n = 27$, Prior 4 (worth $a + b = 12$ “pseudo-observations”) still dominates a large share of the posterior.

3) Prior predictive: generic derivation

For $\theta \sim \text{Beta}(a, b)$,

$$\begin{aligned} p(x_1, \dots, x_n \mid a, b) &= \int_0^1 p(x_{1:n} \mid \theta) p(\theta) d\theta = \int_0^1 \theta^y (1 - \theta)^{n-y} \frac{\theta^{a-1} (1 - \theta)^{b-1}}{B(a, b)} d\theta \\ &= \frac{1}{B(a, b)} \int_0^1 \theta^{a+y-1} (1 - \theta)^{b+n-y-1} d\theta = \boxed{\frac{B(a+y, b+n-y)}{B(a, b)}} \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+y)\Gamma(b+n-y)}{\Gamma(a+b+n)}. \end{aligned}$$

- Depends on the data only through (y, n) .
- Equivalently: (prior $\times$ likelihood)/posterior, evaluated at any θ .
- For y itself (Beta-Binomial): $p(y \mid a, b) = \binom{n}{y} \frac{B(a+y, b+n-y)}{B(a, b)}$; the $\binom{n}{y}$ cancels in Bayes factors.

3) Prior predictive for each prior ($y = 9, n = 27$)

$$\text{Prior 1: } p_1 = \frac{B(10, 19)}{B(1, 1)} = \frac{9! 18!}{28!} = 7.620 \times 10^{-9}$$

$$\text{Prior 2: } p_2 = \frac{B(10, 23)}{B(1, 5)} = 5 \frac{9! 22!}{32!} = 7.751 \times 10^{-9}$$

$$\text{Prior 3: } p_3 = \frac{B(11, 28)}{B(2, 10)} = 110 \frac{10! 27!}{38!} = 8.310 \times 10^{-9}$$

$$\text{Prior 4: } p_4 = \frac{B(19, 20)}{B(10, 2)} = 110 \frac{18! 19!}{38!} = 1.638 \times 10^{-10}$$

(using $B(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}$ for integers, so $1/B(1, 1) = 1$, $1/B(1, 5) = 5$, $1/B(2, 10) = 1/B(10, 2) = 110$.)

Log scale: $\log p_i = -18.693, -18.676, -18.606, -22.532$.

4) Bayes factors and posterior model probabilities: generic

Model M_i : Bernoulli likelihood + Beta(a_i, b_i) prior.

$$BF_{ij} = \frac{p(x_{1:n} \mid M_i)}{p(x_{1:n} \mid M_j)} = \frac{B(a_i + y, b_i + n - y)}{B(a_j + y, b_j + n - y)} \cdot \frac{B(a_j, b_j)}{B(a_i, b_i)}.$$

With prior model probabilities $\Pr(M_i)$,

$$\Pr(M_i \mid x_{1:n}) = \frac{\Pr(M_i) p(x_{1:n} \mid M_i)}{\sum_{k=1}^4 \Pr(M_k) p(x_{1:n} \mid M_k)} \stackrel{\Pr(M_i)=1/4}{=} \frac{p_i}{\sum_{k=1}^4 p_k} = \left(\sum_{k=1}^4 BF_{ki} \right)^{-1}.$$

4) Results for $y = 9$, $n = 27$

Bayes factors BF_{ij} (row i vs column j)

	M_1	M_2	M_3	M_4
M_1	1	0.983	0.917	46.52
M_2	1.017	1	0.933	47.32
M_3	1.091	1.072	1	50.74
M_4	0.021	0.021	0.020	1

Posterior model probabilities

(uniform $\Pr(M_i) = 1/4$)

Model	$\Pr(M_i \mid x_{1:n})$
M_1 : Beta(1, 1)	0.320
M_2 : Beta(1, 5)	0.325
M_3 : Beta(2, 10)	0.349
M_4 : Beta(10, 2)	0.007

- Priors 1–3 are essentially indistinguishable ($BF \approx 1$); Prior 3 is marginally preferred.
- Prior 4 placed most mass near $\theta \approx 0.83$, far from $y/n = 1/3$: data favor every other model by a factor of ≈ 50 .

Appendix: R code

```
y <- 9; n <- 27
a <- c(1, 1, 2, 10); b <- c(1, 5, 10, 2)
th <- seq(0.001, 0.999, length = 1000)
lik <- dbeta(th, y + 1, n - y + 1)                # normalized likelihood
par(mfrow = c(2, 2))
for (i in 1:4) {
  post <- dbeta(th, a[i] + y, b[i] + n - y)
  plot(th, post, type = "l", lwd = 2, ylim = c(0, max(post, lik, dbeta(th, a[i], b[i]))),
        xlab = expression(theta), ylab = "density", main = paste0("Beta(", a[i], ", ", b[i], ")"))
  lines(th, dbeta(th, a[i], b[i]), lty = 2); lines(th, lik, lty = 3)
}
par(mfrow = c(1, 1))
matplot(th, sapply(1:4, function(i) dbeta(th, a[i] + y, b[i] + n - y)), type = "l",
        lty = 1, lwd = 2, xlab = expression(theta), ylab = "posterior")
abline(v = y / n, lty = 2)
logp <- lbeta(a + y, b + n - y) - lbeta(a, b)      # log prior predictive
BF <- exp(outer(logp, logp, "-"))                  # BF[i,j] = p_i / p_j
pmp <- exp(logp - max(logp)); pmp <- pmp / sum(pmp) # uniform model prior
round(BF, 3); round(pmp, 4)
```