Posterior Predictive for Bernoulli Trials with Beta Prior

Derivation and Numerical Example

Hedibert Freitas Lopes
Insper Instituto de Ensino e Pesquisa

May 21, 2026

Contents

1 Setup and General Fact

Let $y_i \mid \theta \overset{iid}{\sim} \text{Bernoulli}(\theta)$ with prior $\theta \sim \text{Beta}(\alpha_0, \beta_0)$. Partition the data into past and future blocks:

$$\boldsymbol{y}_0 = (y_1, \ldots, y_{n_0}), \qquad \boldsymbol{y}_1 = (y_{n_0+1}, \ldots, y_{n_0+n_1}),$$

with sufficient statistics $s_0 = \sum_{i=1}^{n_0} y_i$ and $s_1 = \sum_{i=n_0+1}^{n_0+n_1} y_i$.

General fact. For any joint density,

$$p(y_1 \mid y_0) = \frac{\int p(y_0, y_1, \beta)\, d\beta}{p(y_0)} = \int p(y_1 \mid \beta, y_0)\, p(\beta \mid y_0)\, d\beta.$$

"Condition then marginalise" and "marginalise then condition" are the same Fubini/Tonelli calculation.

2 Route 1: Marginalise θ First, Then Condition on $\boldsymbol{y}_0$

Because $\boldsymbol{y}_0 \perp \boldsymbol{y}_1 \mid \theta$:

$$\begin{aligned} p(s_0, s_1) &= \binom{n_0}{s_0}\binom{n_1}{s_1} \frac{B(\alpha_0 + s_0 + s_1,\ \beta_0 + n_0 + n_1 - s_0 - s_1)}{B(\alpha_0, \beta_0)}, \\ p(s_0) &= \binom{n_0}{s_0} \frac{B(\alpha_0 + s_0,\ \beta_0 + n_0 - s_0)}{B(\alpha_0, \beta_0)}. \end{aligned}$$

Dividing:

$$p(s_1 \mid s_0) = \binom{n_1}{s_1} \frac{B(\alpha_0 + s_0 + s_1,\ \beta_0 + n_0 + n_1 - s_0 - s_1)}{B(\alpha_0 + s_0,\ \beta_0 + n_0 - s_0)}.$$

3 Route 2: Condition on $\boldsymbol{y}_0$ First (Posterior), Then Marginalise

Posterior from $\boldsymbol{y}_0$:

$$\theta \mid \boldsymbol{y}_0 \sim \text{Beta}(\alpha_1, \beta_1), \qquad \alpha_1 = \alpha_0 + s_0, \quad \beta_1 = \beta_0 + n_0 - s_0.$$

Predictive by marginalising θ:

$$\begin{aligned} p(s_1 \mid \boldsymbol{y}_0) &= \binom{n_1}{s_1} \int_0^1 \theta^{s_1}(1-\theta)^{n_1 - s_1} \frac{\theta^{\alpha_1 - 1}(1-\theta)^{\beta_1 - 1}}{B(\alpha_1, \beta_1)}\, d\theta \\ &= \binom{n_1}{s_1} \frac{B(\alpha_1 + s_1,\ \beta_1 + n_1 - s_1)}{B(\alpha_1, \beta_1)}. \end{aligned}$$

4 Conclusion

$$\boxed{s_1 \mid s_0 \ \sim\ \text{Beta-Binomial}(n_1,\ \alpha_0 + s_0,\ \beta_0 + n_0 - s_0)}$$

Both routes yield the same distribution, confirming the Fubini/Tonelli identity.

Special Case: $n_1 = 1$, $\alpha_0 = \beta_0 = 1$

With a Uniform$(0, 1)$ prior, the posterior predictive for a single new trial is:

$$P(y_{n_0+1} = 1 \mid s_0) = \frac{1 + s_0}{2 + n_0},$$

Laplace's rule of succession. Before any data: $P(y_1 = 1) = \frac{1}{2}$.

5 Numerical Example

5.1 Parameters

$$n_0 = 10, \quad s_0 = 3, \quad \alpha_0 = \beta_0 = 1.5, \quad n_1 = 5.$$

Updated hyperparameters:

$$\alpha_1 = \alpha_0 + s_0 = 4.5, \qquad \beta_1 = \beta_0 + n_0 - s_0 = 8.5.$$

Quantity	Formula	Value
Prior	$\text{Beta}(\alpha_0, \beta_0)$	Beta(1.5, 1.5)
Prior mean	$\alpha_0/(\alpha_0 + \beta_0)$	0.500
Likelihood mode (MLE)	s_0/n_0	0.300
Posterior	$\text{Beta}(\alpha_1, \beta_1)$	Beta(4.5, 8.5)
Posterior mean	$\alpha_1/(\alpha_1 + \beta_1)$	0.346
Posterior mode	$(\alpha_1 - 1)/(\alpha_1 + \beta_1 - 2)$	0.318
Post. pred. mean	$n_1\,\alpha_1/(\alpha_1 + \beta_1)$	1.731
Post. pred. variance	—	1.455

5.2 Prior, Likelihood and Posterior for θ

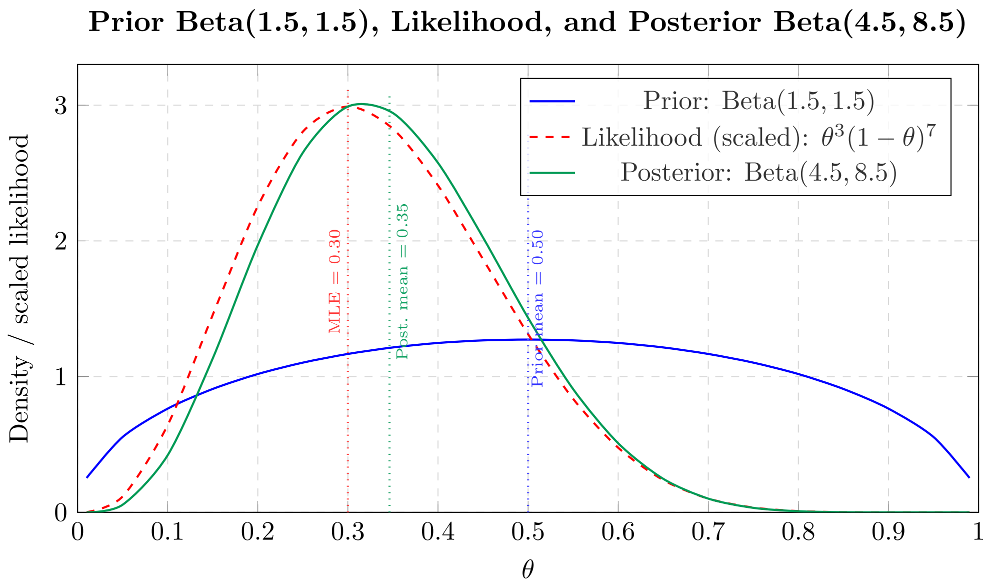

Figure 1: Prior Beta(1.5, 1.5) (blue solid), scaled likelihood $\theta^3(1-\theta)^7$ (red dashed), and posterior Beta(4.5, 8.5) (green solid) as functions of θ. Dotted verticals mark the prior mean, MLE, and posterior mean. The posterior is pulled toward the prior relative to the MLE, reflecting moderate prior strength vs. $n_0 = 10$ observations.

5.3 Prior Predictive and Posterior Predictive for S_1

Prior predictive $S_1 \sim \text{Beta-Binomial}(5, 1.5, 1.5)$ (symmetric, before seeing any data):

$p_{\text{pr}}(0) = 0.129, \quad p_{\text{pr}}(1) = 0.176, \quad p_{\text{pr}}(2) = 0.195, \quad p_{\text{pr}}(3) = 0.195, \quad p_{\text{pr}}(4) = 0.176, \quad p_{\text{pr}}(5) = 0.129.$

Posterior predictive $S_1 \mid s_0 \sim \text{Beta-Binomial}(5, 4.5, 8.5)$ (after observing $s_0 = 3$ successes in $n_0 = 10$ trials):

$p_{\text{po}}(0) = 0.164, \quad p_{\text{po}}(1) = 0.295, \quad p_{\text{po}}(2) = 0.283, \quad p_{\text{po}}(3) = 0.175, \quad p_{\text{po}}(4) = 0.069, \quad p_{\text{po}}(5) = 0.014.$

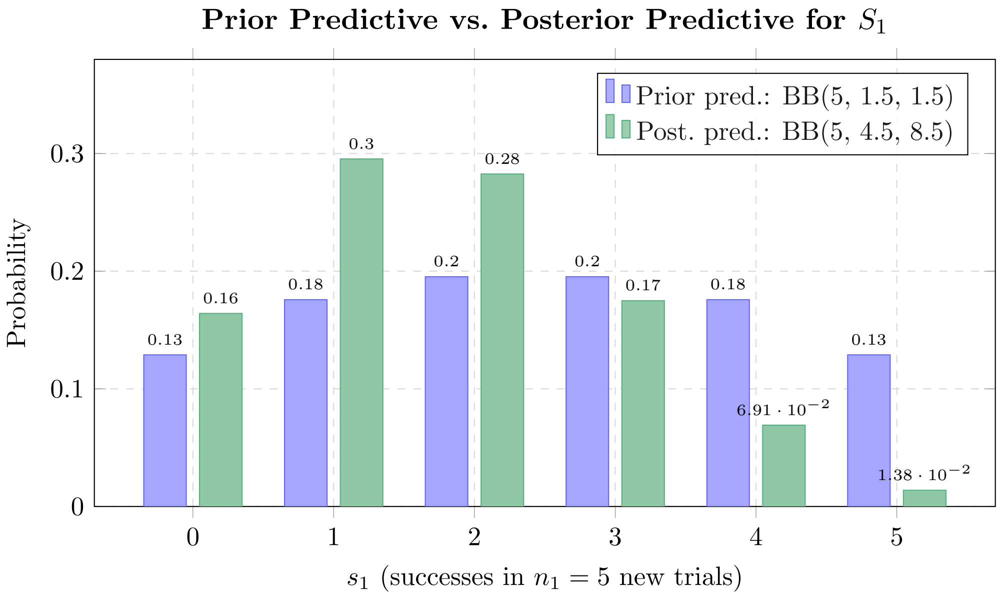

Figure 2: Prior predictive Beta-Binomial$(5, 1.5, 1.5)$ (blue) is symmetric around $s_1 = 2.5$, reflecting prior ignorance. The posterior predictive Beta-Binomial$(5, 4.5, 8.5)$ (green) is right-skewed toward low counts, consistent with the observed low success rate ($s_0 = 3$ out of $n_0 = 10$, posterior mean $\hat{\theta} \approx 0.346$). Posterior predictive mean $= n_1\hat{\theta} \approx 1.73$.