

Lab: Priors & The Power of Fake Data

1. Posterior Credible Intervals via Free Throws

Recall the beta binomial model:

$$\begin{cases} W \sim \text{Binomial}(W+L, p) & \longrightarrow \text{likelihood} \\ p \sim \text{Beta}(\alpha, \beta) & \longrightarrow \text{prior} \end{cases}$$

By Bayes Rule, the posterior distribution of $p|W, L$ is **proportional to** $PCP|W, L) \propto p^{W+\alpha-1} (1-p)^{L+\beta-1}$

so is $p|W, L \sim \text{Beta}(\alpha+W, \beta+L)$ [the symbol \propto means "proportional to"]

Which reflects our Bayesian beliefs about p in a distribution.

α acts like a synthetic number of successful trials and β acts like a synthetic number of unsuccessful trials.

Since we now have a **full distribution** (a Beta distribution) encoding our beliefs

(and uncertainty) about p given the data W, L ,
we can create a Bayesian posterior
credible interval CI to summarize our
uncertainty in p :

Find $CI = [A, B]$ so that

$$\begin{aligned} &P(A \leq p | W, L \leq B) \\ &= P(A \leq \text{Beta}(\alpha + W, \beta + L) \leq B) \\ &= 0.95. \end{aligned}$$

- Using the free throw data, compare the credible interval to Agresti, Wald, and Boot intervals.
- You will need to find suitable prior hyperparameters α, β (e.g. those that minimize out-of-sample loss).
- Note: Bayesian posterior credible intervals are NOT frequentist confidence intervals, even if they turn out similar! Credible intervals are the result of a full probability model.