

## RSVPs or how I learned how many people will attend your wedding

♡ Let  $\theta$  be the true proportion of people that will attend your wedding out of  $N$  total invitations that you send.

♡ Let  $y$  be the yes's out of  $n$  RSVPs you have received to date.

♡ We want to know the posterior density of  $\theta$  given the data  $y$ .

♡ Given  $\theta$ , the likelihood of getting  $y$  yes's and  $n - y$  no's is

$$p(y|\theta) \propto \theta^y(1 - \theta)^{n-y}.$$

♡ A conjugate prior for  $\theta$  is

$$p(\theta) \propto \theta^{\alpha-1}(1 - \theta)^{\beta-1}, \text{ where } \theta \sim \text{Beta}(\alpha, \beta).$$

♡ Using Bayes' rule, the posterior density for  $\theta$  is

$$\begin{aligned} p(\theta|y) &\propto p(y|\theta)p(\theta) \\ &= \theta^{y+\alpha-1}(1 - \theta)^{n-y+\beta-1} \\ &= \text{Beta}(\alpha + y, \beta + n - y). \end{aligned}$$

♡ The posterior mean of  $\theta$  is an average of the prior mean ( $\frac{\alpha}{\alpha+\beta}$ ) and the sample mean ( $\frac{y}{n}$ ):

$$E(\theta|y) = \frac{\alpha + y}{\alpha + \beta + n}.$$

♡ The posterior variance of  $\theta$  is

$$\text{var}(\theta|y) = \frac{(\alpha + y)(\beta + n - y)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)} = \frac{E(\theta|y)[1 - E(\theta|y)]}{\alpha + \beta + n + 1}.$$

♡ The prior density is equivalent to having received  $\alpha$  previous yes's and  $\beta$  previous no's. As  $y$  and  $n - y$  grow large compared to  $\alpha$  and  $\beta$ , the posterior mean approaches the sample mean and the posterior variance goes to 0 at rate  $\frac{1}{n}$ .

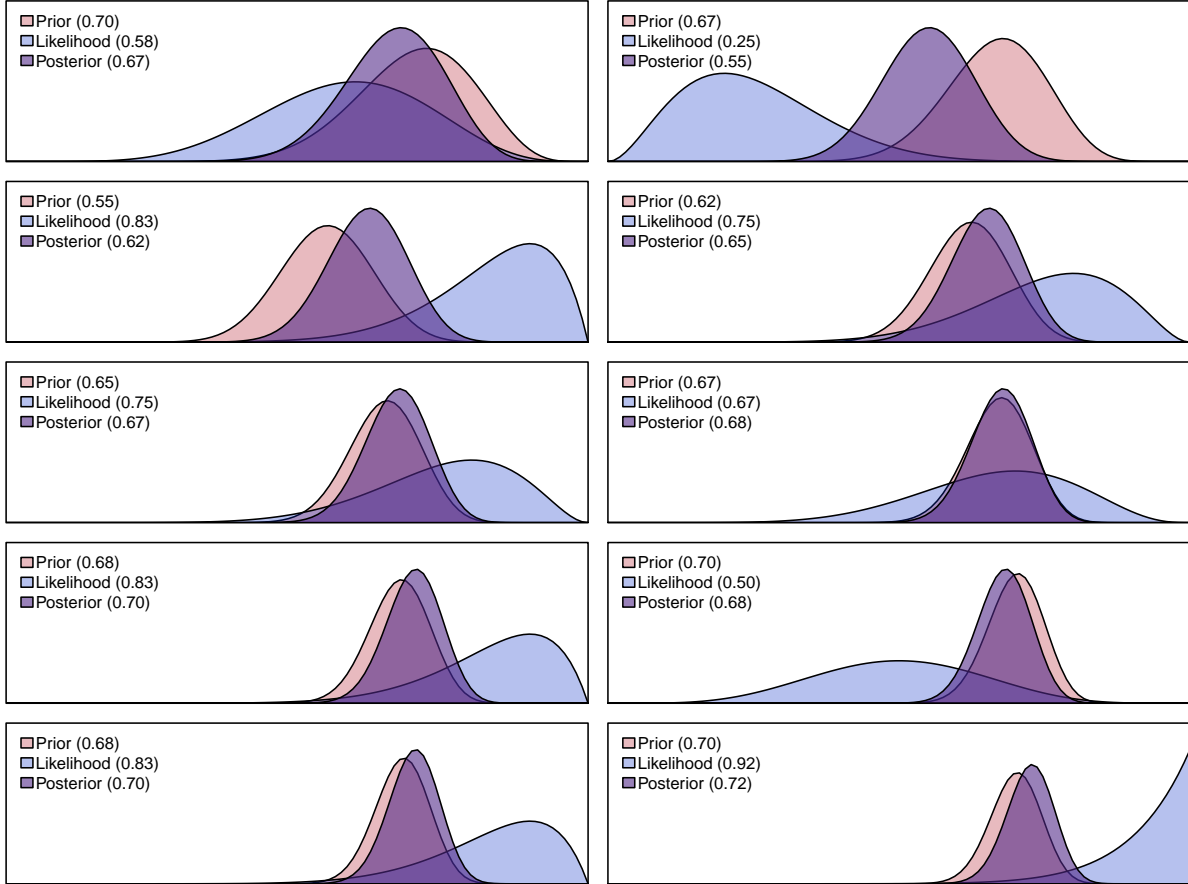
♡ After each response we can update the prior with the previous posterior.

♡ Below is an example of how the posterior will update as you receive responses.

♡ Suppose you believe that  $\theta = 0.7$ , but you want your prior to be relatively weak compared to the data, such that the mean of the posterior after receiving all of the RSVPs will be close to  $\frac{y}{N}$ . (Maybe your prior can account for people that respond yes but do not attend the wedding, or forget to respond but still attend). For example, if  $N = 100$ , let  $\theta \sim \text{Beta}(\alpha = 14, \beta = 6)$  such that the prior is the same weight as 20 invitations and  $E(\theta) = 0.7$ .

♡ If you receive responses in groups of 10, where each row of  $y$  is the number of yes's out of the 10 responses, here is how your posterior will update:

$$y_i = [6, 2, 9, 8, 8, 7, 9, 5, 9, 10]. \quad \frac{y}{N} = 0.73.$$



$$E(\theta|y_i) = [0.67, 0.55, 0.62, 0.65, 0.67, 0.68, 0.70, 0.68, 0.70, 0.72].$$

$$E(\theta|y) = 0.72. \text{ var}(\theta|y) = 9e - 04. \text{ 95\% confidence interval: } [0.64, 0.80].$$