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

♡ Let θ 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 θ given the data y.

♡ Given θ, 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 θ 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 θ 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 θ is an average of the prior mean $(\frac{\alpha}{\alpha+\beta})$ and the sample mean $(\frac{y}{n})$:

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

♡ The posterior variance of θ is

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

♡ The prior density is equivalent to having received $\alpha-1$ previous yes's and $\beta-1$ previous no's. As y and $n-y$ grow large compared to α and β, 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.

♡ Suppose there be 5 subgroups such that the prior distribution is a joint distribution of 5 distinct priors:

$$p(\theta) = p(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5).$$

♡ Where the θs come from the Beta distributions:

	θ_1	θ_2	θ_3	θ_4	θ_5
α	6	11	16	18	20
β	16	11	6	4	2

♡ Each subgroup represents expected responses from $n = \alpha + \beta - 2 = 20$ people.

♡ The joint distributions can be computed by taking 1000 samples from each of the subgroup prior distributions.

♡ How will the posterior distributions update if you receive the following data? Each column represents responses associated with a given subgroup. Each row represents 20 responses to invitations received over time.

$$y = \begin{bmatrix} 0.26 & 0.53 & 0.74 & 0.84 & 0.95 \\ 0.35 & 0.35 & 0.75 & 0.85 & 1.00 \\ 0.45 & 0.50 & 0.85 & 0.90 & 1.00 \\ 0.15 & 0.55 & 0.70 & 0.80 & 1.00 \\ 0.40 & 0.45 & 0.95 & 0.90 & 0.90 \end{bmatrix}.$$