

Maximum likelihood estimate

The following logic underlies how we calculate the maximum likelihood estimate for seedling survival to fruiting. For a single observation, the likelihood that we observe y fruiting plants in a plot if n seedlings were present in the plot can be written as a function of the probability of seedling survival p as $[y|p, n] = \binom{n}{y} p^y (1 - p)^{n-y}$. For a set of N observations, each with a number of seedlings n_i and a number of fruiting plants y_i in the i th observation, then we can write the likelihood as

$$\mathcal{L} = [\mathbf{y}|p, \mathbf{n}] = \prod_{i=1}^N \binom{n_i}{y_i} p^{y_i} (1 - p)^{n_i - y_i}. \quad (1)$$

which is often written as

$$\mathcal{L} = [\mathbf{y}|p, \mathbf{n}] = \prod_{i=1}^N \text{binomial}(n_i, p). \quad (2)$$

We can use the likelihood to obtain a maximum likelihood estimate (by minimizing the negative log-likelihood). The maximum likelihood estimate \hat{p} is the overall proportion of seedlings that survive to become fruiting plants, summing all the observations. The point is that the proportion calculated in the MATLAB code corresponds to the estimate from a maximum likelihood estimate. Specifically, it calculates a maximum likelihood estimate for each population in each year of the dataset. We thus give each site j and year k its own probability of success p_{jk} and obtain the MLEs \hat{p}_{jk} .

$$[\mathbf{y}|\mathbf{p}, \mathbf{n}] = \prod_{j=1}^J \prod_{k=1}^K \prod_{i=1}^N \text{binomial}(n_{ijk}, p_{jk}) \quad (3)$$

Binomial likelihood with a beta prior, complete pooling

We can turn this into a Bayesian model by adding a prior to our model. Because the beta is a conjugate prior for a binomial distribution, we use a beta distribution for the prior. In other words, this choice of prior matches the likelihood in a way that the posterior has the same distribution as the prior (cf. Bolker p 177). A beta distribution with shape parameters $\alpha = \beta = 1$ corresponds to noninformative prior. For a set of N observations, each with a number of seedlings n_i and a number of fruiting plants y_i in the i th observation, we can write the joint posterior as

$$[\mathbf{y}|p, \mathbf{n}] = \prod_{i=1}^N \text{binomial}(n_i, p) \text{beta}(p|1, 1). \quad (4)$$

A single probability p represents the probability of seedling survival to fruiting for the all trials (a model with *complete pooling*). The opposite extreme is a model in which each trial i has its own probability of seedling survival to fruiting p_i (a model with *no pooling*).

$$[\mathbf{y}|\mathbf{p}, \mathbf{n}] = \prod_{i=1}^N \text{binomial}(n_i, p_i) \text{beta}(p_i|1, 1). \quad (5)$$

To compare our site- and year-specific MLEs to estimates from Bayesian models, we give each site j and year k its own probability of success p_{jk} , and place a prior on each p_{jk} .

$$[\mathbf{y}|\mathbf{p}, \mathbf{n}] = \prod_{j=1}^J \prod_{k=1}^K \prod_{i=1}^N \text{binomial}(n_{ijk}, p_{jk}) \text{beta}(p_{jk}|1, 1). \quad (6)$$

- 1 Effectively, this is a model in which we are completely pooling observations from each site
- 2 and year. This extends what happens with the maximum likelihood estimates when we
- 3 sum across all the plots at a site in a given year and calculate the proportion of seedlings
- 4 that survive to become fruiting plants. One difference between the two approaches is that

5 with the Bayesian model we account for the number of trials and counts; the data from one
6 plot with a single seedling compromises with the prior to give us posterior probability of
7 success (see **Comparison**).

Binomial model with a beta prior, partial pooling

Next, we'll consider adding pooling to our model. We do this by putting hyperpriors on the parameters for the beta distribution. Focusing first on one population, we give the population a per-year probability of success p_k . We place a prior on each p_k but rather than directly parameterize the probability of success we use hyperpriors. We'll use the parameterization in Kruschke:

$$\begin{aligned}
 [\mathbf{p}, \omega, \kappa | \mathbf{y}, \mathbf{n}] &= \prod_{k=1}^K \prod_{i=1}^N \text{binomial}(n_{ik}, p_k) \\
 &\times \text{beta}(p_k | \omega(\kappa - 2) + 1, (1 - \omega)(\kappa - 2) + 1) \\
 &\times \text{beta}(\omega | 1, 1) \text{gamma}(\kappa | 0.01, 0.01).
 \end{aligned} \tag{7}$$

8 This parameterization is hierarchical and we can use it to illustrate one of the effects of using
9 this structure in our models. Let's pick a population with years in which there were few data
10 points. For the purposes of illustration, we choose to work with the Lucas Creek East (LCE)
11 population here. We fit the model above to the data from LCE and plot the posteriors for
12 the parameter ω in Figure 2, which corresponds to the mode of the beta distribution. The
13 figure illustrates the concept of *shrinkage*; the posterior distributions of year-level θ_k are
14 pulled towards the population-level mode ω . This effect is particularly evident in years in
15 which there are few data points (number of samples in each year, from 2006-2015: 20, 7,
16 17, 14, 19, 1, 1, 3, 1, 8). In Figure 2, this is shown by the overlap between the posterior
17 distribution for ω and θ_k for years k in which there are few data points. This is particularly

18 true in 2011–2014. One important effect of this is that the [variance between estimated
 19 values θ_k is less than the variance between the individual proportions correct] (paraphrasing
 20 Kruschke)

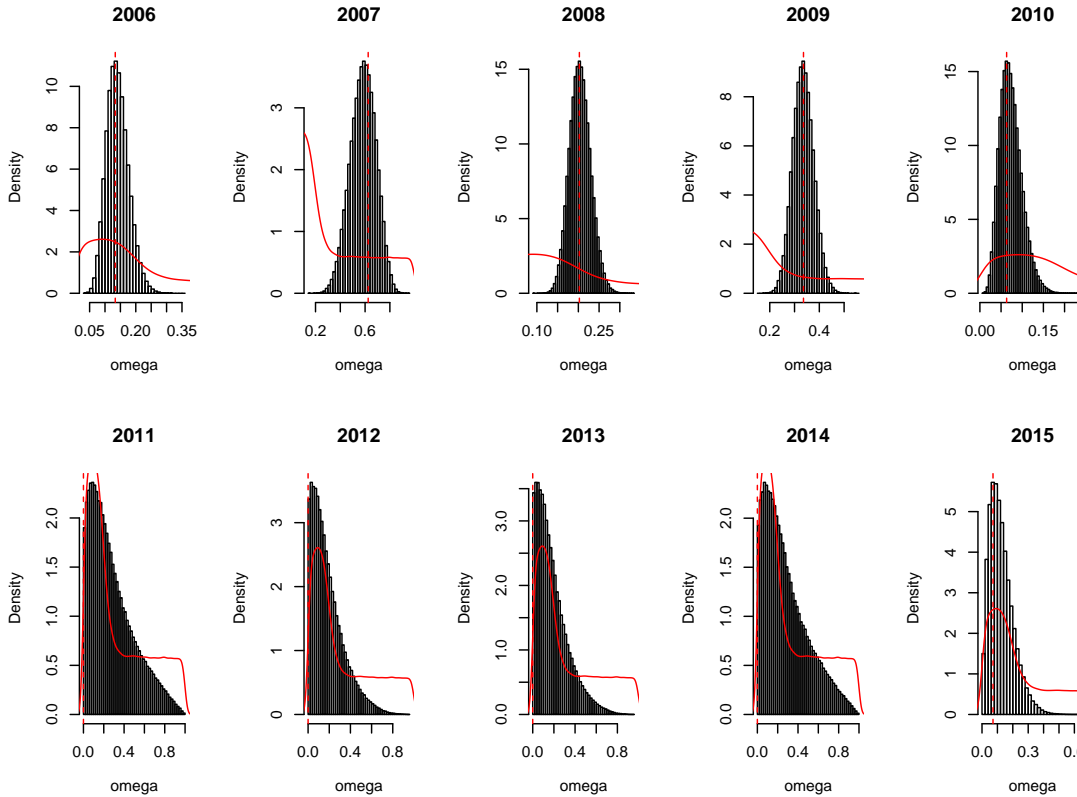


Figure 1: The posterior distribution for θ_k for a model fitted to data from Lucas Creek East. The histogram represents draws from the posterior distribution. The vertical, dotted red line indicates the total proportion of successes (fruiting plants/seedlings) in each year. The solid red line is the density of draws from the posterior of ω .

21 For comparison, we also fit the same model to data from the Black Gulch (BG) pop-
 22 ulation. For BG, most years have a higher number of data points (number of samples in
 23 each year, from 2006-2015: 18, 20, 21, 26, 23, 26, 20, 23, 3, 26). Here, it's clear that most
 24 posterior distributions for year-level θ_k are not greatly influenced by the population-level
 25 mode ω .

To compare our population- and year-specific estimates, we give each population j and

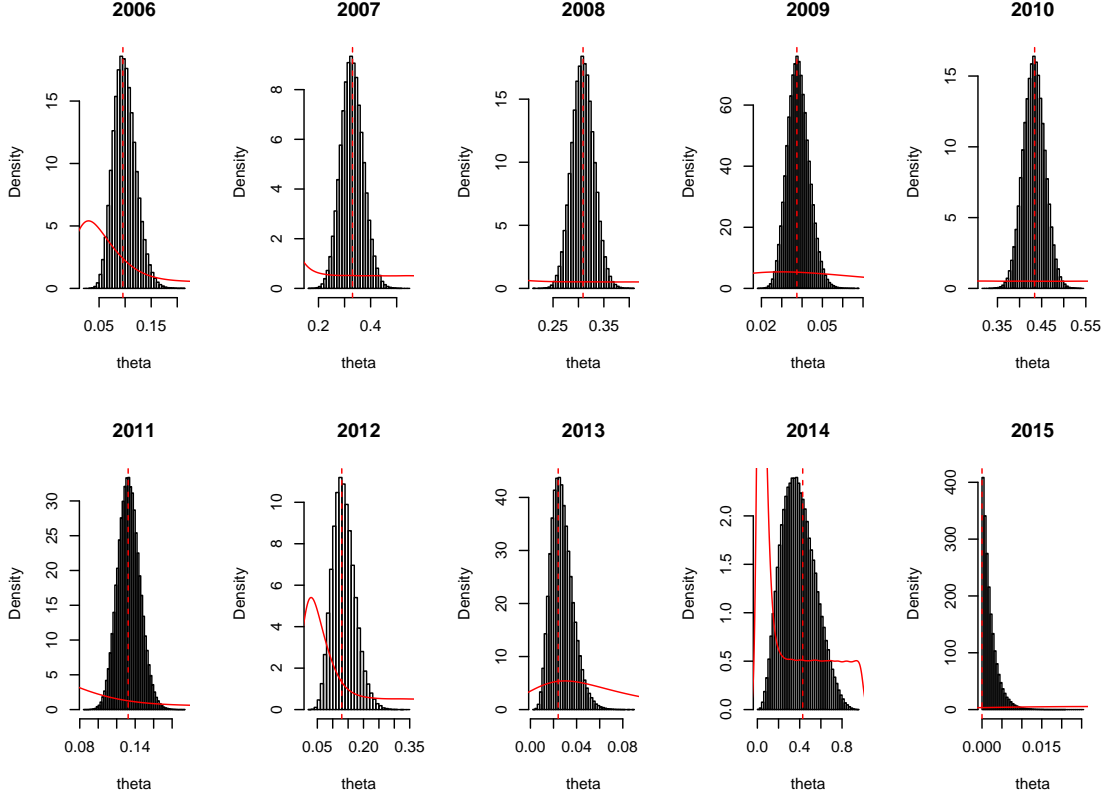


Figure 2: The posterior distribution for θ_k for a model fitted to data from Black Gulch. The histogram represents draws from the posterior distribution. The vertical, dotted red line indicates the total proportion of successes (fruiting plants/seedlings) in each year. The solid red line is the density of draws from the posterior of ω .

year k its own probability of success p_{jk} , place priors ω_j and κ_j on each p_{jk} , and give each prior a set of hyperpriors.

$$\begin{aligned}
 [\mathbf{p}, \omega, \kappa | \mathbf{y}, \mathbf{n}] &= \prod_{k=1}^K \prod_{i=1}^N \text{binomial}(n_{ik}, p_{jk}) \\
 &\times \text{beta}(p_{jk} | \omega_j(\kappa_j - 2) + 1, (1 - \omega_j)(\kappa_j - 2) + 1) \\
 &\times \text{beta}(\omega_j | 1, 1) \text{gamma}(\kappa_j | 0.01, 0.01).
 \end{aligned} \tag{8}$$

Effectively, this is a model in which we are partially pooling observations from each population. The posterior estimates for p_{jk} from this model are similar to one in which ob-

servations from each population are completely pooled. When there are few data points, the population-level parameter ω_j has an effect on the population- and year-level θ_{jk} (see **Comparison**).

Comparison

The first panel in Figure 3 shows that the maximum likelihood estimates are pretty similar to those from the Bayesian model with complete pooling per population and a beta-binomial parameterization. The major differences are where the maximum likelihood estimates approach 0 or 1. The second panel in Figure 3 shows that this is because those estimates come from year-population combinations with a small sample size. For example, the MLE for a year-population combination with one plot and 1 seedling that dies before fruiting would be 0. In a Bayesian model, the prior has a comparatively larger influence on the posterior in situations where there is little data. In this case, the posterior would be a compromise between our one data point and our prior. The estimates converge once we have 5 data points.

Figure 4 shows that, when the data are fit at the population-level only, a model with complete and partial pooling give fairly similar estimates. The difference between the posterior for a model with complete versus partial pooling has a greater range for smaller sample sizes (CIs are larger for smaller sample sizes) and a model with complete pooling returns slightly higher estimates. Figures 5 and 6 compare the posterior distribution for population- and year-combinations with many data points (30 data points in Figure 5) and few data points (1 data point in Figure 6). When there are few data points, the posterior for p_{jk} in the partial pooling model is influenced by the population-level parameter ω (Figure 6).

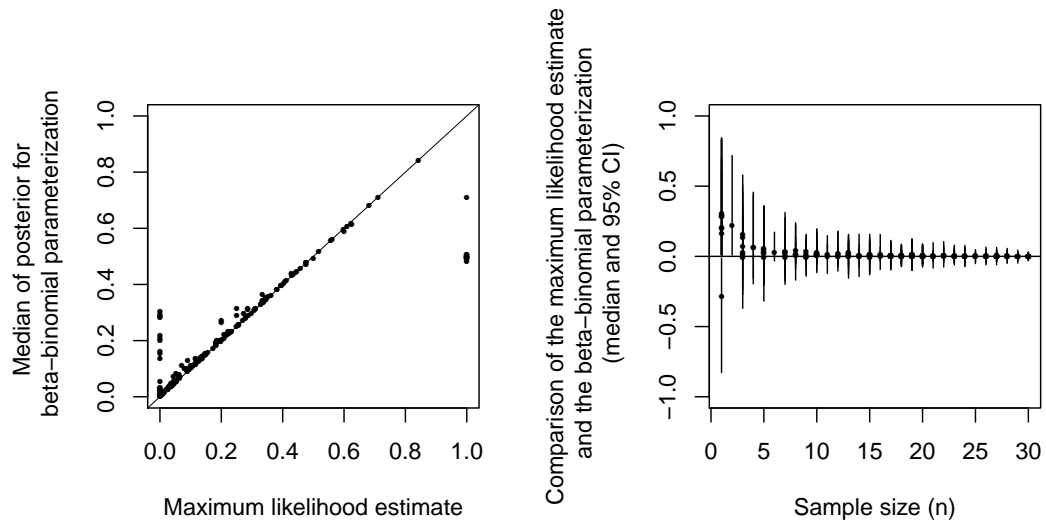


Figure 3: (A) This panel plots the median of the posterior from the beta-binomial with complete pooling per population against the maximum likelihood estimate. (B) This panel compares the full posterior distribution from the beta-binomial parameterization with the maximum likelihood estimate. The plot shows the median of the difference (with 95% CIs) against the sample size in the year-population combination for that estimate.

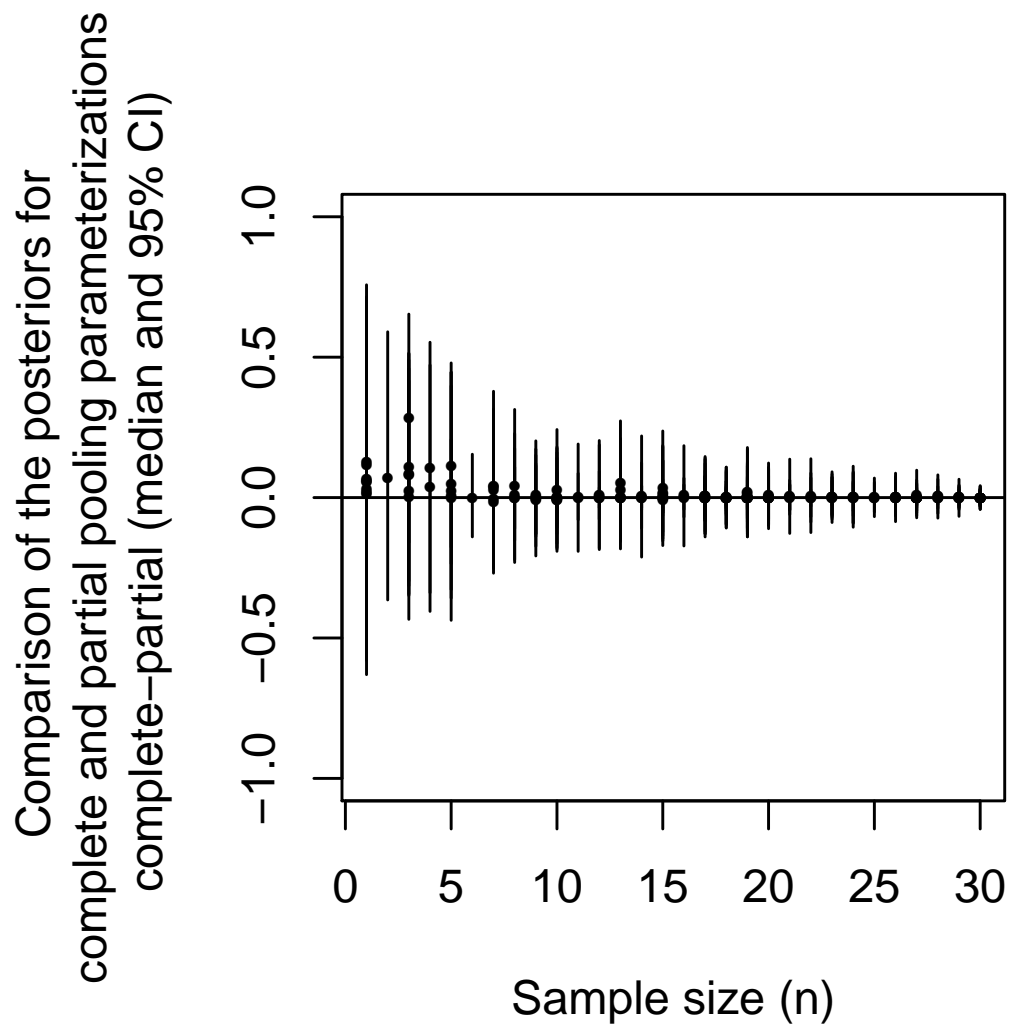


Figure 4: (A) This plot compares the full posterior distribution of the beta-binomial parameterization with complete pooling against the beta-binomial parameterization with partial pooling. The plot shows the median of the difference (with 95% CIs) against the sample size in the year-population combination for that estimate.

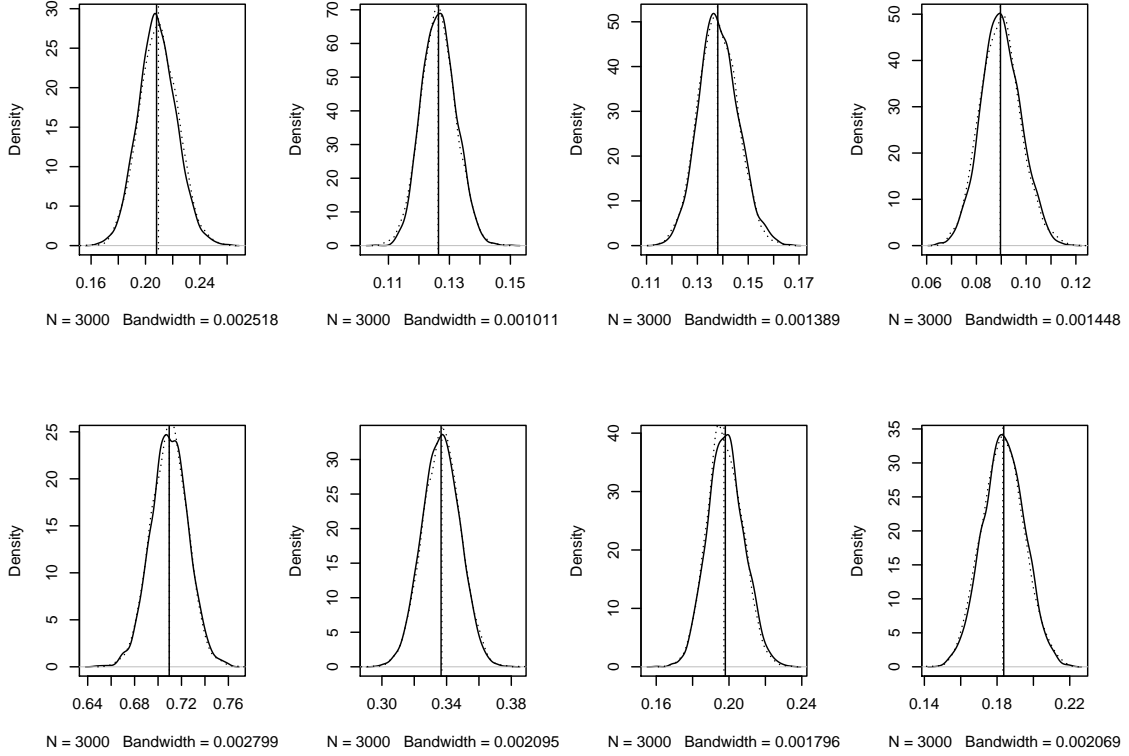


Figure 5: (A) This plot shows the posterior distribution of the beta-binomial parameterization with complete pooling (solid line) against the posterior distribution of the beta-binomial parameterization with partial pooling (dotted line). The medians are given by vertical lines. These are 8 population- and year- combinations that have 30 data points each. These correspond to points on the right side of Figure 4.

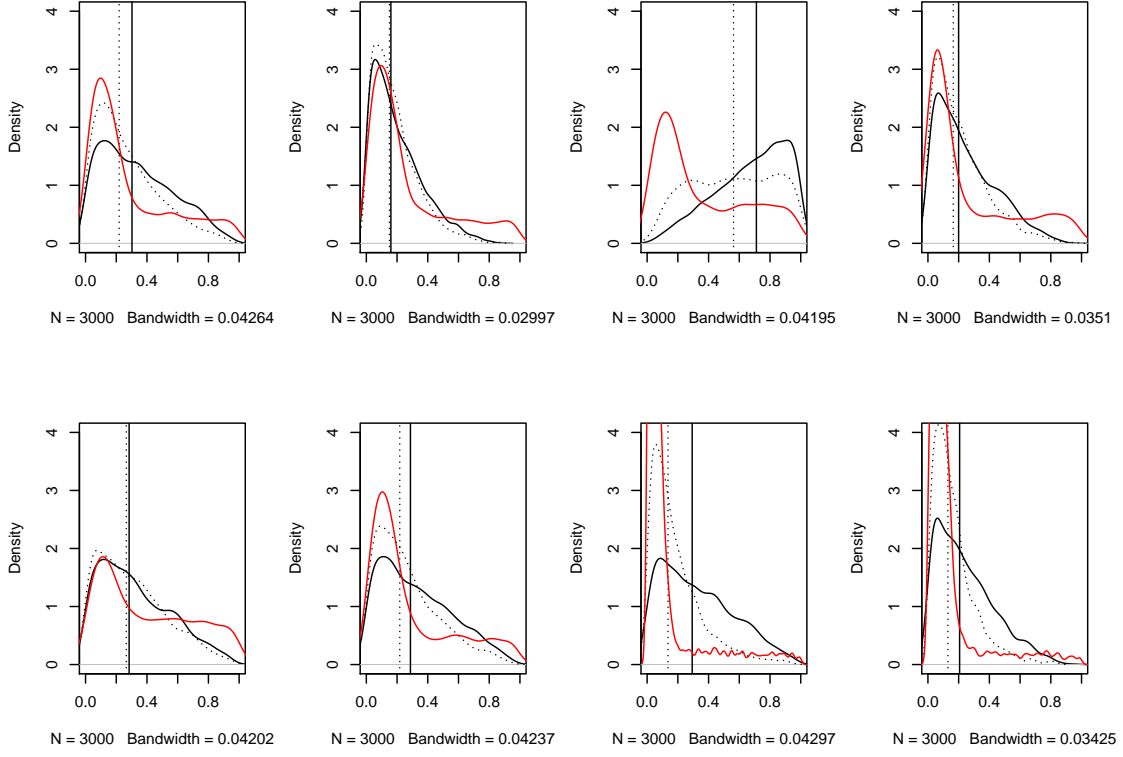


Figure 6: (A) This plot shows the posterior distribution of the beta-binomial parameterization with complete pooling (solid line) against the posterior distribution of the beta-binomial parameterization with partial pooling (dotted line). The medians are given by vertical lines. The red lines are the population-level modes ω . These are 8 population- and year- combinations that have only 1 data point each. These correspond to points on the left side of Figure 4.