

Appendix X

1 In this appendix, I compare the following estimates: (1) maximum likelihood estimate (bino-
2 mial likelihood), (2) posterior distribution (binomial likelihood with a beta prior, complete
3 pooling), (3) posterior distribution (binomial likelihood with a beta prior, hierarchical model
4 for probability of survivorship), (4) posterior distribution (binomial likelihood, hierarchical
5 model for log-odds of probability of survivorship).

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

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

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} .

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

This is a model in which we are completely pooling observations from each site and year.

Another thing we could say about this model is that it is Bayesian but non-hierarchical. This is extends what happens with the maximum likelihood estimates when we sum across all the plots at a site in a given year and calculate the proportion of seedlings that survive to become fruiting plants. One difference between the two approaches is that with the Bayesian model we account for the number of trials and counts; the data from one plot with a single seedling compromises with the prior to give us posterior probability of success (see **Comparison**).

The Bayesian and frequentist estimates converge as the sample size approaches infinity.

Binomial model with a beta prior, partial pooling, parameterization via mean: one population, one year

Next, we'll consider adding pooling to our model. To explain this, we'll focus first on the data from one population in one year. We want a hierarchical model that estimates the probability of survivorship in each plot (θ_i) and simultaneously estimates the mean probability of survivorship in the year (ϕ). The probability of survivorship for each plot i is θ_i . We then assume that the probability of survivorship for the i plots is drawn from a distribution of probabilities defined by the mean probability of survivorship in a given year ϕ and sample size κ . This effectively means that the prior on θ_i is itself a parameter; (i.e. we use hyperpriors rather than directly place priors on the probability of survivorship θ_i). We reparameterize the beta distribution with its mean ϕ and the parameter κ . The mean of

a random variable distributed $\text{beta}(\alpha, \beta)$ is $\phi = \frac{\alpha}{\alpha + \beta}$. With a $\text{beta}(1, 1)$ prior, the parameter $\kappa = \alpha + \beta$ is roughly the sample size plus two.

$$\begin{aligned}
[\mathbf{p}, \boldsymbol{\theta}, \phi, \kappa | \mathbf{y}, \mathbf{n}] &= \prod_{i=1}^N \text{binomial}(n_i, \theta_i) \\
&\times \text{beta}(\theta_i | \phi \kappa, (1 - \phi) \kappa) \\
&\times \text{uniform}(\phi | 0, 1) \text{Pareto}(\kappa | 1.5, 1).
\end{aligned} \tag{7}$$

14 We place a uniform prior on the mean probability of survivorship, ϕ , because it must lie
15 between 0 and 1. We place a bounded, positive prior on κ with the $\text{Pareto}(\alpha, c)$ distribution,
16 which is parameterized by a shape (α) and scale (c) parameter.

17 This parameterization is hierarchical because the estimates for ϕ and κ contribute to
18 estimates for θ_i . [other effects?]

Binomial model with a beta prior, partial pooling, parameterization via mean: one population, multiple years

Next, we expand the scope of our analysis to include data from plots (i) in multiple years (j) of data for a single population $k = 1$. Here, we want a hierarchical model that estimates the probability of survivorship in each plot in each year (θ_{ij}). We want to simultaneously estimate the mean probability of survivorship in each year (ϕ_j) and the mean probability of survivorship in the population. The probability of survivorship for each plot i in a year j is θ_{ij} . We then assume that the probability of survivorship for the i plots in year j is drawn from a distribution of probabilities defined by the mean probability of survivorship in a given year ϕ_j and sample size κ_j . Furthermore, we assume that the mean probability of survivorship in a given year ϕ_j is itself drawn from a distribution of probabilities defined by

the mean probability of survivorship at the population ϕ_0 and the parameter κ_0 .

$$\begin{aligned}
[\mathbf{p}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\kappa}, \phi_0, \kappa_0 | \mathbf{y}, \mathbf{n}] &= \prod_{j=1}^J \prod_{i=1}^N \text{binomial}(n_{ij}, \theta_{ij}) \\
&\times \text{beta}(\theta_{ij} | \phi_j \kappa_j, (1 - \phi_j) \kappa_j) \\
&\times \text{beta}(\phi_j | \phi_0 \kappa_0, (1 - \phi_0) \kappa_0) \text{Pareto}(\kappa_j | 1.5, 1) \\
&\times \text{uniform}(\phi_0 | 0, 1) \text{Pareto}(\kappa_0 | 1.5, 1).
\end{aligned} \tag{8}$$

19 This parameterization is hierarchical because the estimates for ϕ_j and κ_j contribute to esti-
20 mates for θ_{ij} , and the estimates for ϕ_0 and κ_0 contribute to estimates for ϕ_j . [other effects?]

Binomial model with a beta prior, partial pooling, parameterization via mean: multiple populations, multiple years

Finally, we expand the scope of our analysis to include data from plots (i) in multiple years (j) of data for multiple populations k . Here, we want a hierarchical model that estimates the probability of survivorship in each plot in each year at each population (θ_{ijk}). We want to simultaneously estimate the mean probability of survivorship in each population ϕ_k , the mean probability of survivorship in each year ($\phi_j k$). The probability of survivorship for each plot i in a year j at population k is θ_{ijk} . We then assume that the probability of survivorship for the i plots in year j in population k is drawn from a distribution of probabilities defined by the mean probability of survivorship in a given year at a given population ϕ_{jk} and parameter κ_{jk} . Furthermore, we assume that the mean probability of survivorship in a given year in a given population ϕ_{jk} is itself drawn from a distribution of probabilities defined by the mean probability of survivorship for the population $\phi_{0,k}$ and the parameter $\kappa_{0,k}$. The model is similar to the one for one population except the indexing has changed so that we estimate a

mean probability of survivorship for each population.

$$\begin{aligned}
[\mathbf{p}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\kappa}, \phi_0, \kappa_0 | \mathbf{y}, \mathbf{n}] &= \prod_{k=1}^K \prod_{j=1}^J \prod_{i=1}^N \text{binomial}(n_{ijk}, \theta_{ijk}) \\
&\times \text{beta}(\theta_{ijk} | \phi_{jk} \kappa_{jk}, (1 - \phi_{jk}) \kappa_{jk}) \\
&\times \text{beta}(\phi_{jk} | \phi_{0,k} \kappa_{0,k}, (1 - \phi_{0,k}) \kappa_{0,k}) \text{Pareto}(\kappa_j | 1.5, 1) \\
&\times \text{uniform}(\phi_{0,k} | 0, 1) \text{Pareto}(\kappa_{0,k} | 1.5, 1).
\end{aligned} \tag{9}$$

21 This parameterization is hierarchical because the estimates for ϕ_{jk} and κ_{jk} contribute to
22 estimates for θ_{ijk} , and the estimates for $\phi_{0,k}$ and $\kappa_{0,k}$ contribute to estimates for ϕ_{jk} . [other
23 effects?]

Comparison

24 There are a few relevant comparisons. First, we want to explore estimates from a Bayesian
25 model compare to the maximum likelihood estimates. Second, we want to explore the effect
26 of adding hierarchical structure to the parameter estimates in our models. Finally, we want
27 to explore the effect of adding hierarchical structure on how appropriate our models are
28 (compare posterior predictive checks).

29 [need to give details of models here - trace plots, etc.]

30 First, we want to explore estimates from a Bayesian model compare to the maximum
31 likelihood estimates. The comparison I am interested in is the effect of adding a prior,
32 rather than the effect of adding hierarchy to the model. I will compare the maximum likeli-
33 hood estimates (equation (3)) and the non-hierarchical Bayesian model (equation (6)). The
34 first panel in Figure 1 shows that the maximum likelihood estimates are pretty similar to
35 those from the Bayesian model with complete pooling per population and a beta-binomial
36 parameterization. The major differences are where the maximum likelihood estimates ap-

proach 0 or 1. The second panel in Figure 1 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 the 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.

Second, we want to explore the effect of adding hierarchical structure to the parameter estimates in our models. These comparisons will all be among models fit to data for one population ($k = 1$). I am interested in comparing a non-hierarchical model (equation (6)), a hierarchical model with year-level parameters (equation (7)), and a hierarchical model with year-level and population-level parameters (equation (8)). All these comparisons will be made at a randomly selected site.

The relevant figures are:

1 & 2 ; one comparing 1-level H to NH, one comparing 2-level H to NH; posterior distributions are more broad for the H models Figure 2) Figure 3)

3 These posteriors can be summarized by their median and 95% credible intervals. Broader CIs at population level Figure 4)

Figure ?? 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 ?? and ?? compare the posterior distribution for population- and year-combinations with many data points (30 data points in Figure ??) and few data points (1 data point in Figure ??). When there are few data points, the posterior for p_{jk} in the partial pooling model is influenced by the population-level parameter ω

63 (Figure ??).

64 Finally, we want to explore the effect of adding hierarchical structure on how appropriate
65 our models are (compare posterior predictive checks).

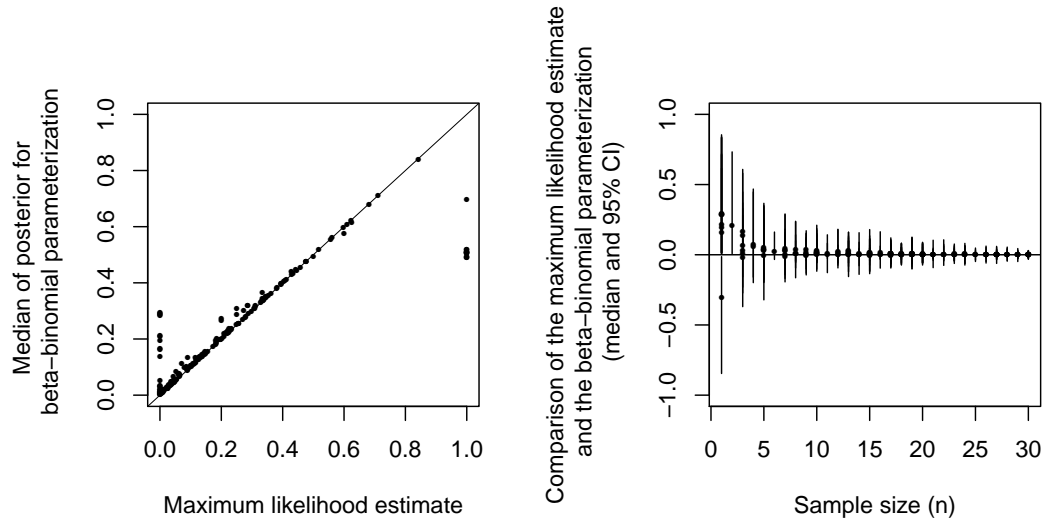


Figure 1: (A) This panel plots the median of the posterior from the beta-binomial with complete pooling per population against the maximum likelihood estimate. (B) This panel compares the full posterior distribution from the beta-binomial parameterization with the maximum likelihood estimate. The plot shows the median of the difference (with 95% CIs) against the sample size in the year-population combination for that estimate.

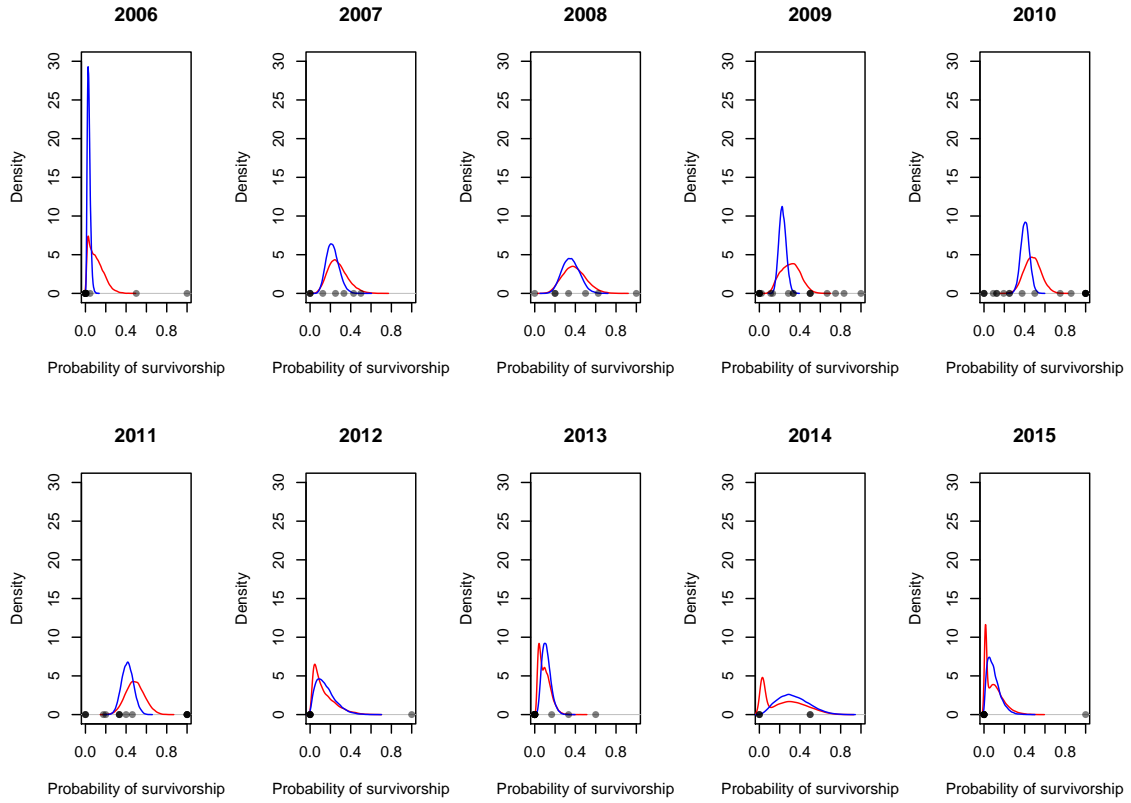


Figure 2: (A) Comparison of and nonhierarchical and year-level parameter hierarchical model.

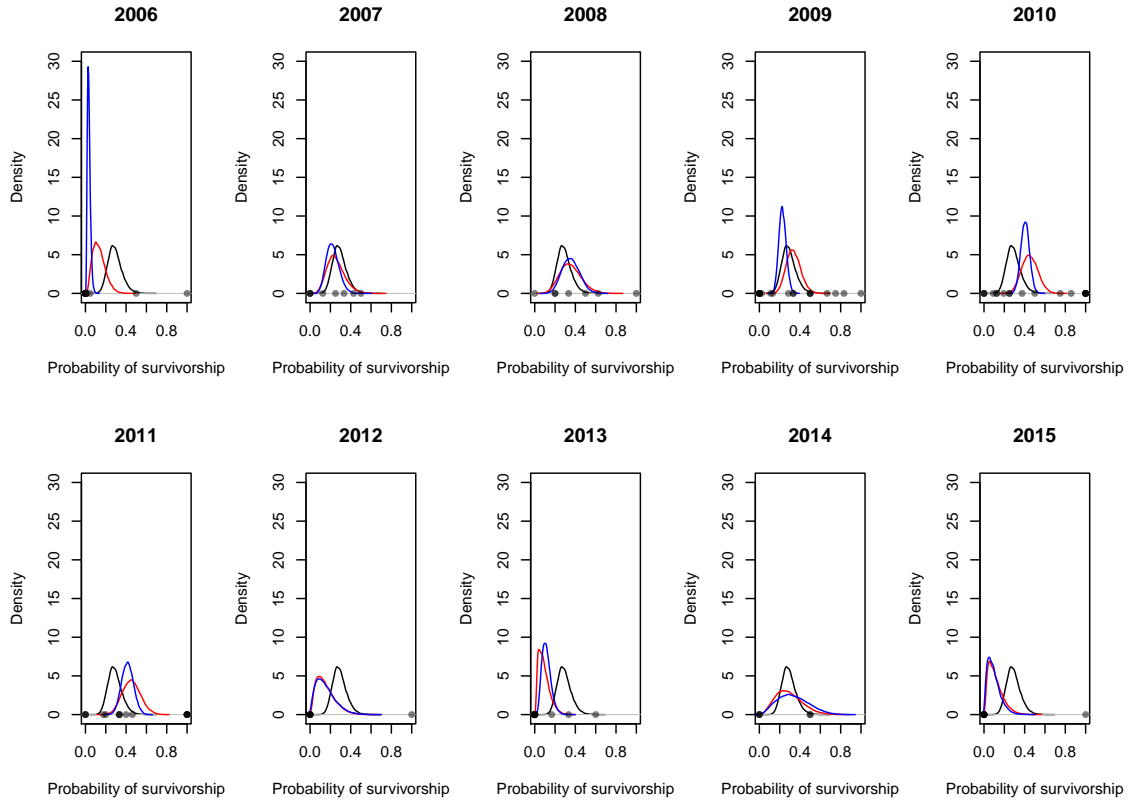


Figure 3: (A) Comparison of and nonhierarchical and year- and population-level parameter hierarchical model.

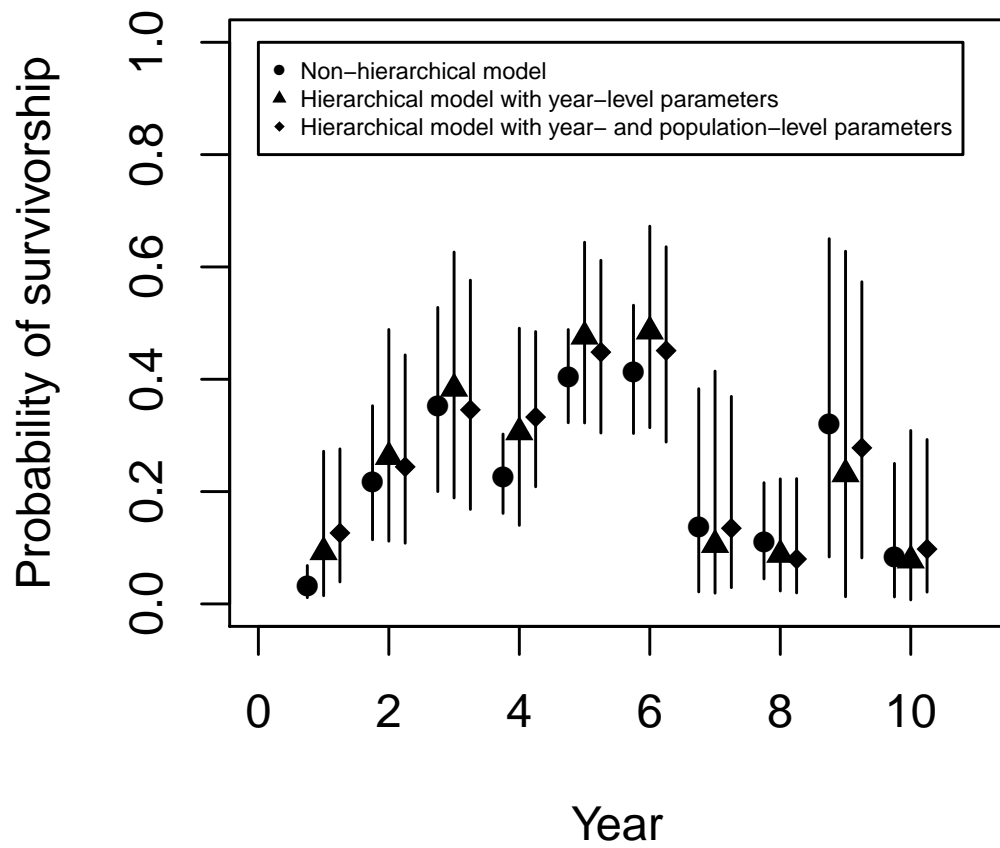


Figure 4: (A) Comparison of median and 95% confidence intervals for models with three levels of structure, fit to the same dataset.