

# Appendix X

1 I will use this appendix to compare different models for a success/trial dataset to demonstrate  
2 shrinkage due to partial pooling.

...

## Maximum likelihood estimate

3 Below, I'm including a snippet of MATLAB code that I think has been used to calculate  
4 seedling survival to fruiting ( $\sigma$ ) in the past. The code generates estimates for  $\sigma$  for each  
5 population in each year; I do not believe that the code snippet that is used to calculate `sig2`  
6 is used elsewhere in the MATLAB program.

```
7 

---


```

```
8 % ESTIMATE ABOVE-GROUND VITAL RATE:
```

```
9 %   sigma, SURVIVAL GERM>FRUITING
```

```
10
```

```
11 % get data on survival germ>fruiting
```

```
12 dfname=[folder 'Survivorship & Fecundity_06-11vers2.xlsx'];
```

```
13 [numdata,txtdata]=xlsread(dfname);
```

```
14
```

```
15 % VARIABLE
```

```
16 % 1    easting
```

```
17 % 2    northing
```

```
18 % 3    site
```

```
19 % 4    transect
```

```
20 % 5    position
```

```
21 % 6    seedling#_1/06
```

```
22 % 7    flow#_6/06
```

```

23 % 8    fruit#_6/06
24 % 9    seedling#_1/07
25 % 10   flow#_6/07
26 % 11   fruit#_6/07
27 % 12   fruit/pl_6/07
28 % 13   fl>sdl_07
29 % 14   seedling#_1/08
30 % 15   flow#_6/08
31 % 16   fruit#_6.08
32 % 17   fruit/pl_6.08
33 % 18   fl>sdl_08
34 % 19   seedling#_1/09
35 % 20   flow#-5/09
36 % 21   fruit#-6/09
37 % 22   fruit/pl-6.09
38 % 23   fl>sdl_09
39 % 24   seedling#_1/10
40 % 25   fruitpl#-6/10
41 % 26   fruit/pl-6.10
42 % 27   fruitpl#>sdl_10
43 % 28   seedling#_2/11
44 % 29   fruitpl#-6/11
45 % 30   (fruit+fl)/pl-6.11
46 % 31   fruitpl#>sdl_11
47 % 32   Notes
48
49 Site=txtdata(:,3);

```

```

50 Site(1)=[];
51 Sdl=[numdata(:,6) numdata(:,9) numdata(:,14) numdata(:,19) numdata(:,24)
52     numdata(:,28)]; % no. of seedlings
53 Frt=[numdata(:,8) numdata(:,11) numdata(:,16) numdata(:,21) numdata(:,25)
54     numdata(:,29)]; % no. of fruiting plants
55
56 NumSdl=zeros(numpops,numyrs);
57 NumFrt=NumSdl;
58 nsig=NumSdl;
59 sigma=NumSdl;
60 sig2=NumSdl;
61 MeanSdl=NumSdl;
62 MeanFrt=NumSdl;
63 MeanSdlSE=NumSdl;
64 MeanFrtSE=NumSdl;
65 s0sig1=NumSdl; % stores raw estimates of products of 3 vrs from plot data
66 s0sig1_bag=NumSdl;
67 sigest=NumSdl;
68
69 for p=1:numpops
70
71     PopNow=pops(p);
72
73     for y=1:numyrs
74         % only use plots ("positions") with both sdl>0 and frting plt counts
75         I=find(strcmp(Site,PopNow) & Sdl(:,y)>0 & ~isnan(Frt(:,y)));
76         nsig(p,y)=length(I);

```

```

77     x1=sum(Sdl(I,y));
78     NumSdl(p,y)=x1;
79     x2=sum(Frt(I,y));
80     NumFrt(p,y)=x2;
81     sigma(p,y)=min(x2/x1,1); % surv. germ>fruiting not allowed to exceed 1
82
83     % compute sigma using plot means - doesn't require a plot to have both
84         sdl and fruiting plant counts
85     i1=find(strcmp(Site,PopNow) & ~isnan(Sdl(:,y)));
86     i2=find(strcmp(Site,PopNow) & ~isnan(Frt(:,y)));
87     MeanSdl(p,y)=mean(Sdl(i1,y));
88     MeanSdlSE(p,y)=std(Sdl(i1,y))/sqrt(length(i1));
89     MeanFrt(p,y)=mean(Frt(i2,y));
90     MeanFrtSE(p,y)=std(Frt(i2,y))/sqrt(length(i2));
91     sig2(p,y)=MeanFrt(p,y)/MeanSdl(p,y);
92
93 end
94
95 end
96

```

---

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

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

97 We can use the likelihood to obtain a maximum likelihood estimate (by minimizing the  
 98 negative log-likelihood). The maximum likelihood estimate  $\hat{p}$  is the overall proportion of  
 99 seedlings that survive to become fruiting plants, summing all the observations.

## Binomial model with complete pooling and a beta prior

Next, we'll consider adding a prior to our model and writing this as a Bayesian model. We use a beta distribution for the prior, because the beta is a conjugate prior for a binomial distribution; 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

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

This corresponds to a model in which there is *complete pooling*, with a single probability  $p$  representing the probability of seedling survival to fruiting for the all trials. The opposite extreme would be a model with *no pooling*, in which each trial  $i$  has its own probability of seedling survival to fruiting  $p_i$ .

$$[\mathbf{y}|\mathbf{p}, \mathbf{n}] = \prod_{i=1}^N \text{binomial}(n_i, p_i) \text{beta}(p_i|1, 1). \quad (3)$$

This corresponds to a model in which there is *complete pooling*, with a single probability  $p$  representing the probability of seedling survival to fruiting for the all trials. To compare our site- and year-specific MLE fits, it might be best to run the following model:

$$[\mathbf{y}|\mathbf{p}, \mathbf{n}] = \prod_{j=1}^J \prod_{k=1}^K \prod_{i=1}^N \text{binomial}(n_{ijk}, p_{jk}) \text{beta}(p_{jk}|1, 1). \quad (4)$$

Here, we give each site  $j$  and year  $k$  its own probability of success  $p_{jk}$ . Effectively, this is a model in which we are completely pooling observations from each site and year. This is pretty similar to what we are doing 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 do 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. For the most part, though, these estimates are pretty similar (see that most points fall on a 1:1 line).

## binomial model with partial pooling and a beta prior

Next, we'll consider adding pooling to our model. We do this by putting hyperpriors on the parameters for the beta distribution. We are interested in the mean at each site. We'll use the parameterization in Kruschke:

$$\begin{aligned} [\mathbf{p}, \omega, \kappa|\mathbf{y}, \mathbf{n}] &= \prod_{j=1}^J \prod_{i=1}^N \text{binomial}(n_{ij}, p_j) \\ &\times \text{beta}(p_j|\omega(\kappa - 2) + 1, (1 - \omega)(\kappa - 2) + 1) \\ &\times \text{beta}(\omega|1, 1) \text{gamma}(\kappa|.01, .01). \end{aligned} \quad (5)$$

## Multi-level binomial model with partial pooling and a beta prior

Next, we'll consider pooling across all sites and years in our dataset. We do this by putting hyperpriors on the parameters for the beta distribution. We are interested in the mean at each site. We'll use the parameterization in Kruschke:

$$\begin{aligned} [\mathbf{p}, \omega, \kappa | \mathbf{y}, \mathbf{n}] &= \prod_{j=1}^J \prod_{k=1}^K \prod_{i=1}^N \text{binomial}(n_{ijk}, p_{jk}) \\ &\times \text{beta}(p_{jk} | \omega(\kappa - 2) + 1, (1 - \omega)(\kappa - 2) + 1) \\ &\times \text{beta}(\omega | 1, 1) \text{gamma}(\kappa | .01, .01). \end{aligned} \tag{6}$$

## Binomial model with partial pooling and a log-odds parameterization

109 Below, I'm including a snippet of MATLAB code that I think has been used to calculate  
110 seedling survival to fruiting ( $\sigma$ ) in the past. The code generates estimates for  $\sigma$  for each  
111 population in each year; I do not believe that the code snippet that is used to calculate `sig2`  
112 is used elsewhere in the MATLAB program.

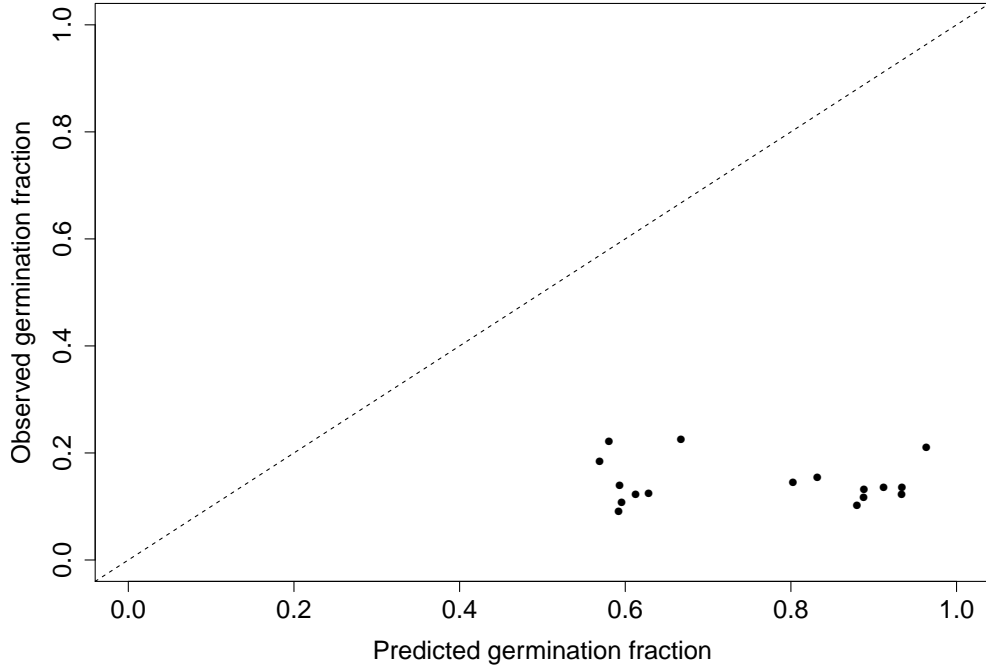


Figure 1: Observed germination probability plotted against the optimal germination probability predicted by a density-independent model. For each population, the observed germination probability is the obtained from the model for seed bank vital rates. Each point is the population-specific median of the posterior of  $g_1$  for a model fit to data from seed bag experiments from 2006–2009. Data was pooled across years. The dotted line indicates a 1:1 relationship between observations and predictions. Values below the line indicate that the model predicts higher germination probabilities than observed; values above the line would indicate that the model predicts lower germination probabilities than observed.