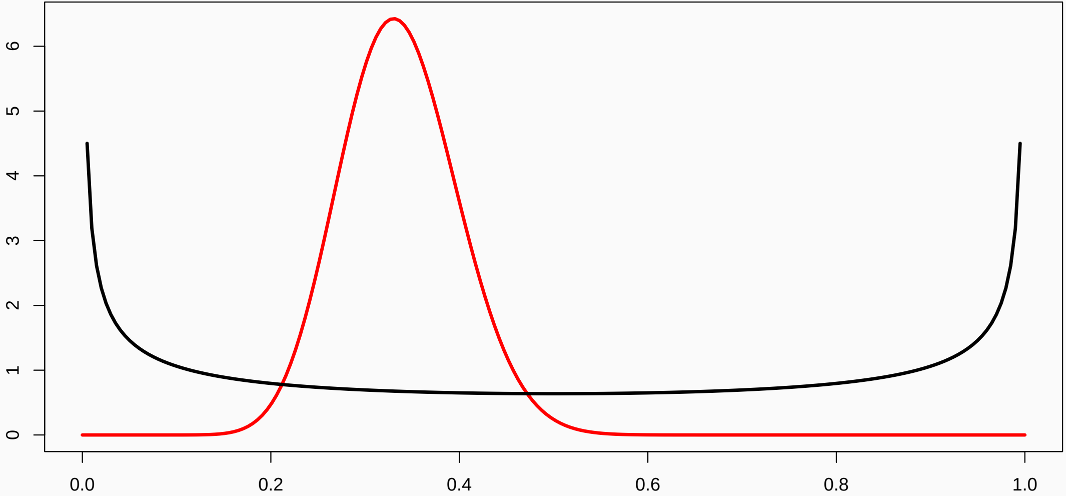
Introduction to Bayesian statistics with R
4. Priors

Olivier Gimenez

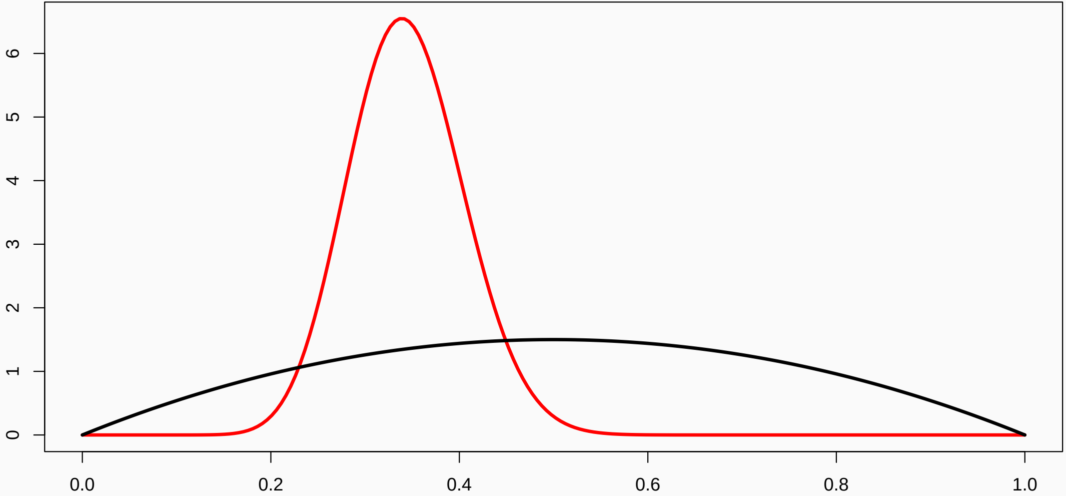
last updated: 2025-03-10

Influence of the prior: Back to our guiding example

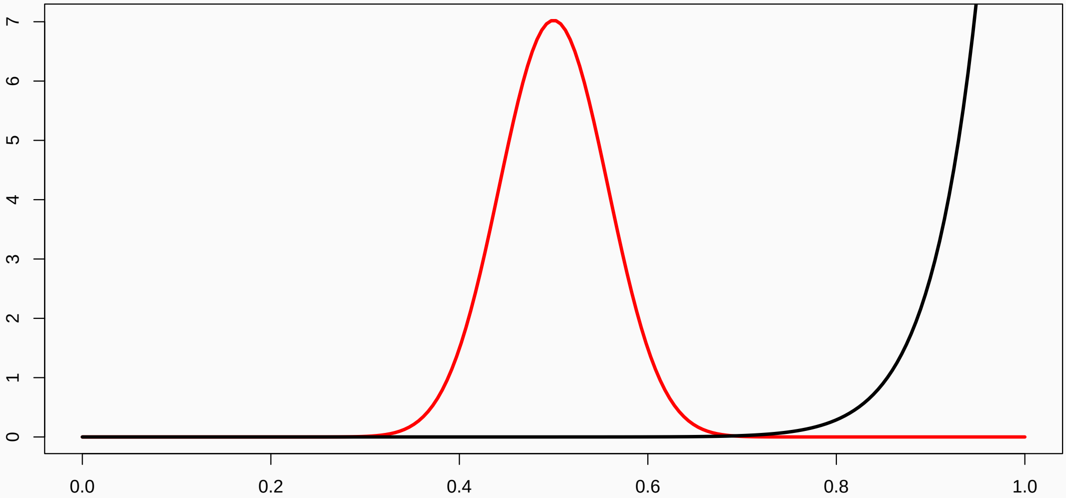
Prior $Beta(0.5, 0.5)$ and posterior survival $Beta(19.5, 38.5)$



Prior $Beta(2, 2)$ and posterior survival $Beta(21, 40)$



Prior $Beta(20, 1)$ and posterior survival $Beta(39, 49)$



The role of the prior

- In biological applications, the prior is a convenient means of incorporating expert opinion or information from previous or related studies that would otherwise need to be ignored. We'll get back to that.

The role of the prior

- In biological applications, the prior is a convenient means of incorporating expert opinion or information from previous or related studies that would otherwise need to be ignored. We'll get back to that.
- With sparse data, the role of the prior can be to enable inference on key parameters that would otherwise be impossible.

The role of the prior

- In biological applications, the prior is a convenient means of incorporating expert opinion or information from previous or related studies that would otherwise need to be ignored. We'll get back to that.
- With sparse data, the role of the prior can be to enable inference on key parameters that would otherwise be impossible.
- With sufficiently large and informative datasets the prior typically has little effect on the results.

The role of the prior

- In biological applications, the prior is a convenient means of incorporating expert opinion or information from previous or related studies that would otherwise need to be ignored. We'll get back to that.
- With sparse data, the role of the prior can be to enable inference on key parameters that would otherwise be impossible.
- With sufficiently large and informative datasets the prior typically has little effect on the results.
- This can be assessed with a sensitivity analysis.

Informative priors vs. no information

- Informative priors aim to reflect information available to the analyst that is gained independently of the data being studied.

Informative priors vs. no information

- Informative priors aim to reflect information available to the analyst that is gained independently of the data being studied.
- In the absence of any prior information on one or more model parameters we wish to ensure that this lack of knowledge is properly reflected in the prior.

How to incorporate prior information?

Estimating survival using capture-recapture data

- A bird might captured, missed and recaptured; this is coded 101.

Estimating survival using capture-recapture data

- A bird might captured, missed and recaptured; this is coded 101.
- Simplest model relies on constant survival ϕ and detection p probabilities.

Estimating survival using capture-recapture data

- A bird might captured, missed and recaptured; this is coded 101.
- Simplest model relies on constant survival ϕ and detection p probabilities.
- Likelihood for that particular bird:

$$\Pr(101) = \phi(1 - p)\phi p$$

Estimating survival using capture-recapture data

- A bird might captured, missed and recaptured; this is coded 101.
- Simplest model relies on constant survival ϕ and detection p probabilities.
- Likelihood for that particular bird:

$$\Pr(101) = \phi(1 - p)\phi p$$

- We assume a vague prior:

$$\phi_{prior} \sim \text{Beta}(1, 1) = \text{Uniform}(0, 1)$$

European dippers in Eastern France (1981-1987)

How to incorporate prior information?

- If no information, mean posterior survival is $\phi_{posterior} = 0.56$ with credible interval $[0.51, 0.61]$.

How to incorporate prior information?

- If no information, mean posterior survival is $\phi_{posterior} = 0.56$ with credible interval $[0.51, 0.61]$.

- Using information on body mass and annual survival of 27 European passerines, we can predict survival of European dippers using only body mass.

How to incorporate prior information?

- If no information, mean posterior survival is $\phi_{posterior} = 0.56$ with credible interval $[0.51, 0.61]$.
- Using information on body mass and annual survival of 27 European passerines, we can predict survival of European dippers using only body mass.
- For dippers, body mass is 59.8g, therefore $\phi = 0.57$ with sd $= 0.073$.

How to incorporate prior information?

- If no information, mean posterior survival is $\phi_{posterior} = 0.56$ with credible interval $[0.51, 0.61]$.
- Using information on body mass and annual survival of 27 European passerines, we can predict survival of European dippers using only body mass.
- For dippers, body mass is 59.8g, therefore $\phi = 0.57$ with sd $= 0.073$.
- Assuming an informative prior $\phi_{prior} \sim \text{Normal}(0.57, 0.073^2)$.

How to incorporate prior information?

- If no information, mean posterior survival is $\phi_{posterior} = 0.56$ with credible interval $[0.51, 0.61]$.
- Using information on body mass and annual survival of 27 European passerines, we can predict survival of European dippers using only body mass.
- For dippers, body mass is 59.8g, therefore $\phi = 0.57$ with sd $= 0.073$.
- Assuming an informative prior $\phi_{prior} \sim \text{Normal}(0.57, 0.073^2)$.
- Mean posterior $\phi_{posterior} = 0.56$ with credible interval $[0.52, 0.60]$.

How to incorporate prior information?

- If no information, mean posterior survival is $\phi_{posterior} = 0.56$ with credible interval $[0.51, 0.61]$.
- Using information on body mass and annual survival of 27 European passerines, we can predict survival of European dippers using only body mass.
- For dippers, body mass is 59.8g, therefore $\phi = 0.57$ with sd $= 0.073$.
- Assuming an informative prior $\phi_{prior} \sim \text{Normal}(0.57, 0.073^2)$.
- Mean posterior $\phi_{posterior} = 0.56$ with credible interval $[0.52, 0.60]$.
- No increase of precision in posterior inference.

How to incorporate prior information?

- Now if you had only the three first years of data, what would have happened?

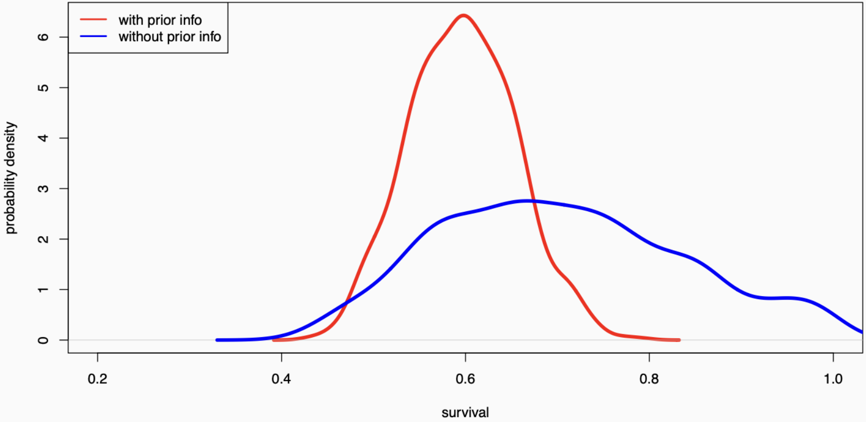
How to incorporate prior information?

- Now if you had only the three first years of data, what would have happened?
- Width of credible interval is 0.47 (vague prior) vs. 0.30 (informative prior).

How to incorporate prior information?

- Now if you had only the three first years of data, what would have happened?
- Width of credible interval is 0.47 (vague prior) vs. 0.30 (informative prior).
- Huge increase of precision in posterior inference (40% gain)!

Compare vague vs. informative prior



Prior elicitation via moment matching

Remember the Beta distribution

- Recall that the Beta distribution is a continuous distribution with values between 0 and 1. Useful for modelling survival or detection probabilities.

Remember the Beta distribution

- Recall that the Beta distribution is a continuous distribution with values between 0 and 1. Useful for modelling survival or detection probabilities.
- If $X \sim Beta(a, b)$, then the first and second moments of X are:

$$\mu = \mathrm{E}(X) = \frac{a}{a+b}$$

$$\sigma^2 = \mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Moment matching

- In the capture-recapture example, we know a priori that the mean of the probability we're interested in is $\mu = 0.57$ and its variance is $\sigma^2 = 0.073^2$.

Moment matching

- In the capture-recapture example, we know a priori that the mean of the probability we're interested in is $\mu = 0.57$ and its variance is $\sigma^2 = 0.073^2$.
- Parameters μ and σ^2 are seen as the moments of a $Beta(a, b)$ distribution.

Moment matching

- In the capture-recapture example, we know a priori that the mean of the probability we're interested in is $\mu = 0.57$ and its variance is $\sigma^2 = 0.073^2$.
- Parameters μ and σ^2 are seen as the moments of a *Beta*(a, b) distribution.
- Now we look for values of a and b that match the observed moments of the Beta distribution (μ and σ^2).

Moment matching

- In the capture-recapture example, we know a priori that the mean of the probability we're interested in is $\mu = 0.57$ and its variance is $\sigma^2 = 0.073^2$.
- Parameters μ and σ^2 are seen as the moments of a *Beta*(a, b) distribution.
- Now we look for values of a and b that match the observed moments of the Beta distribution (μ and σ^2).
- We need another set of equations:

$$a = \left(\frac{1-\mu}{\sigma^2} - \frac{1}{\mu}\right)\mu^2$$

$$b = a\left(\frac{1}{\mu} - 1\right)$$

- For our model, that means:

```r
(a <- ( (1 - 0.57)/(0.073*0.073) - (1/0.57) )*0.57^2)
#> [1] 25.64636
(b <- a * ( (1/0.57) - 1))
#> [1] 19.34726
```

- For our model, that means:

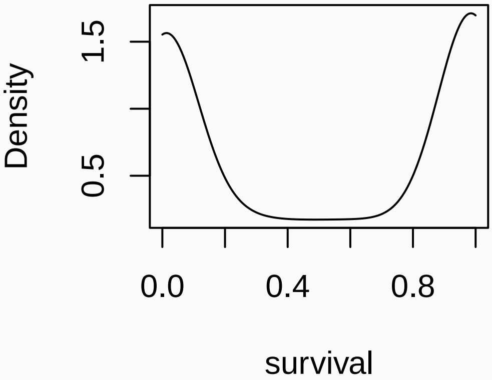
```
(a <- ( (1 - 0.57)/(0.073*0.073) - (1/0.57) )*0.57^2)
#> [1] 25.64636
(b <- a * ( (1/0.57) - 1))
#> [1] 19.34726
```

- Now use $\phi_{prior} \sim \text{Beta}(a = 25.6, b = 19.3)$ instead of $\phi_{prior} \sim \text{Normal}(0.57, 0.073^2)$

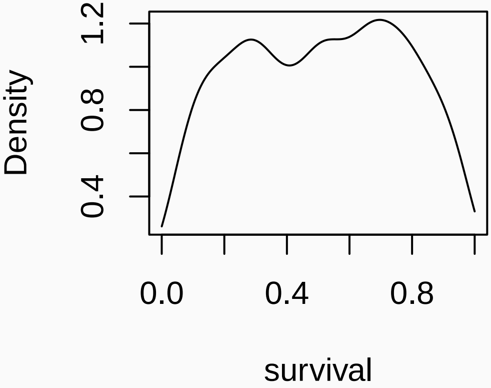
Beware of so-called non-informative priors

Logistic regression

Unreasonable $\text{logit}(\phi) = \beta \sim N(0, 10^2)$



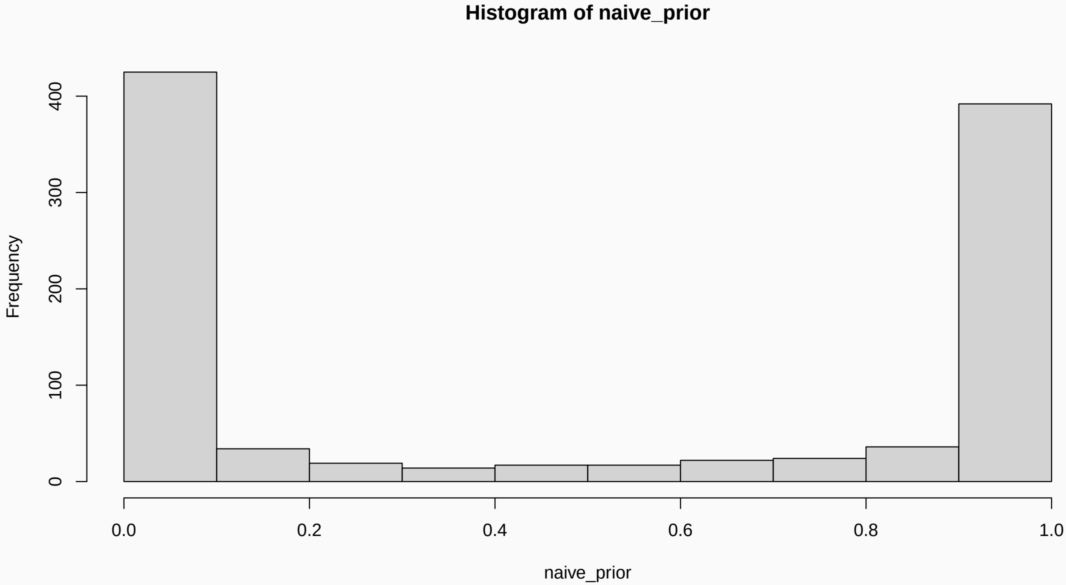
Reasonable $\text{logit}(\phi) = \beta \sim N(0, 1.5^2)$



Prior predictive checks: Use simulations to check your priors.

```
naive_prior_logit <- rnorm(1000,0,10) # on logit scale
naive_prior <- plogis(naive_prior_logit) # on natural scale bw 0 and 1
hist(naive_prior)
```

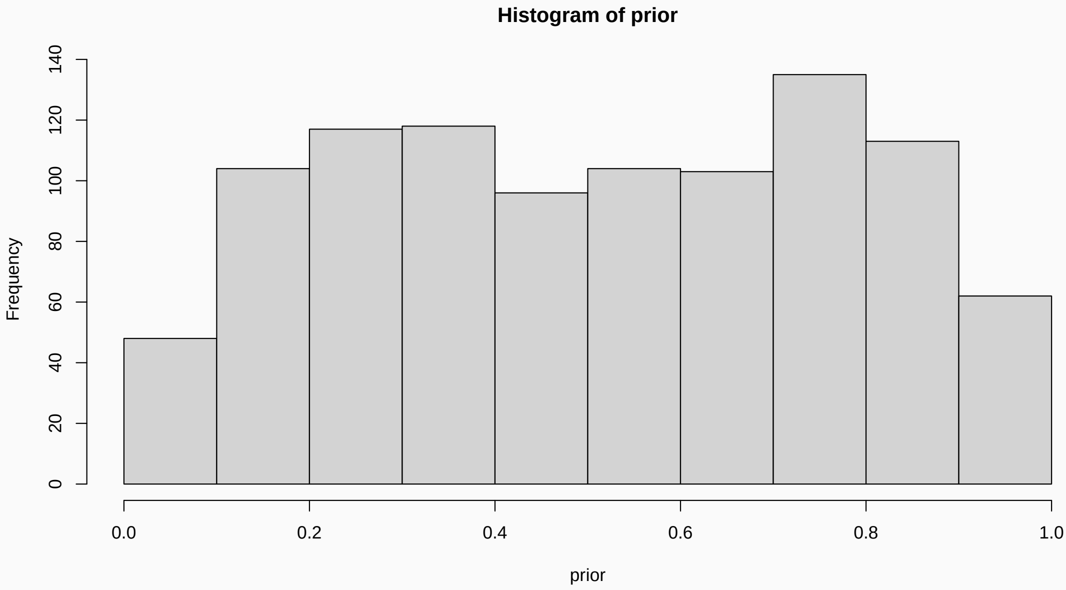
Prior predictive checks: Use simulations to check your priors.



Prior predictive checks: Use simulations to check your priors.

```
prior_logit <- rnorm(1000,0,1.5) # on logit scale
prior <- plogis(prior_logit) # on natural scale bw 0 and 1
hist(prior)
```

Prior predictive checks: Use simulations to check your priors.



Further reading

Further reading

- Banner KM, Irvine KM, Rodhouse TJ (2020). The use of Bayesian priors in Ecology: The good, the bad and the not great. *Methods Ecol Evol* 11: 882–889.
- Lemoine NP (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos* 128: 912–928.
- McCarthy MA, Masters P (2005). Profiting from prior information in Bayesian analyses of ecological data. *Journal of Applied Ecology* 42: 1012–1019.
- Mikkola P et al. (2024). Prior Knowledge Elicitation: The Past, Present, and Future. *Bayesian Analysis* 19: 1129-1161.
- Northrup JM, Gerber BD (2018). A comment on priors for Bayesian occupancy models. *PLoS ONE* 13(2): e0192819