Bayesian statistics with R
4. A detour with priors

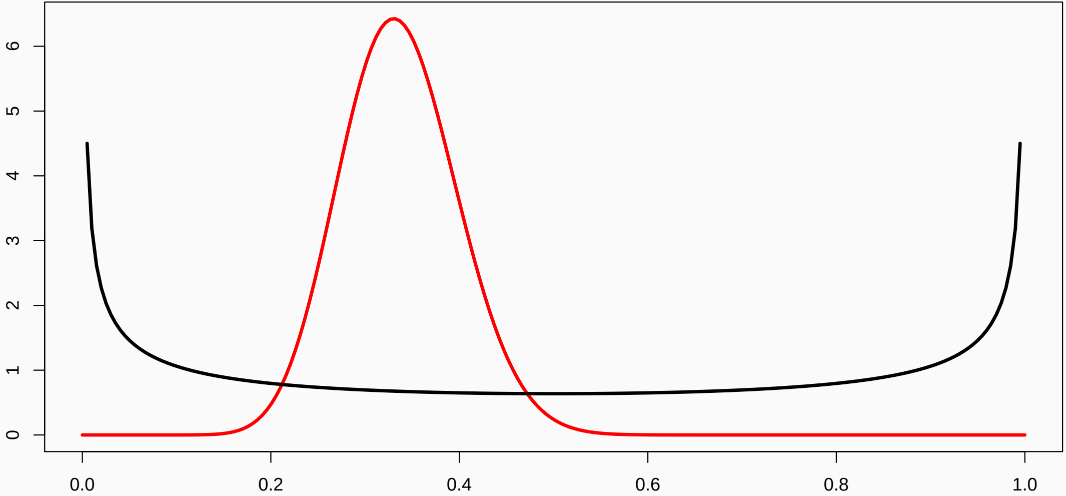
Olivier Gimenez

March 2021

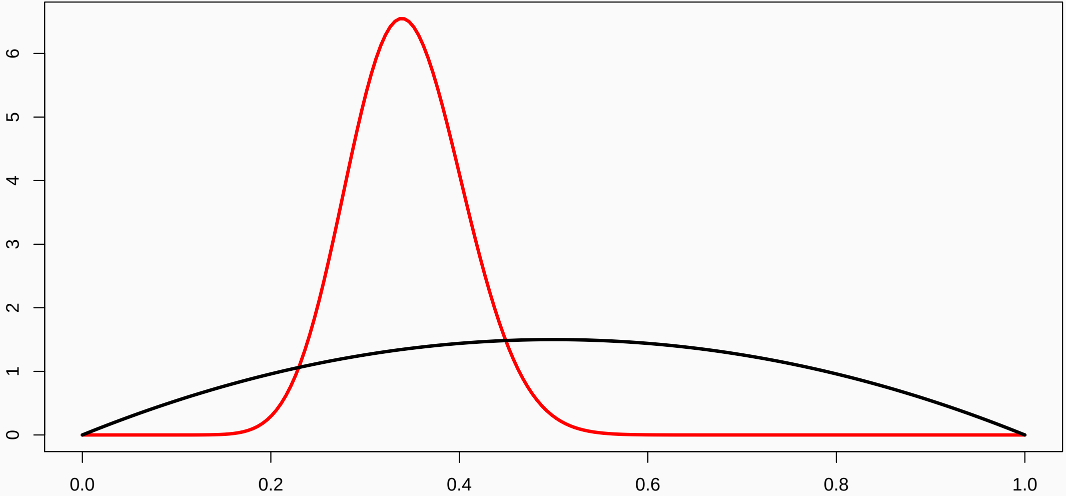
A detour to explore priors

Influence of the prior

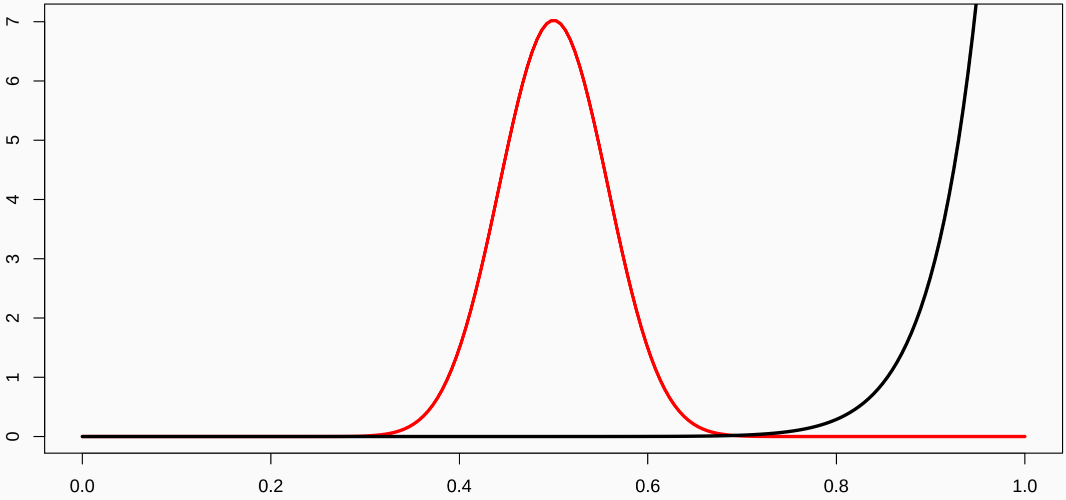
Prior $Beta(0.5, 0.5)$ and posterior survival $Beta(19.5, 38.5)$



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.
- Always perform 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.

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.
- Always perform a sensitivity analysis.


WE ARE ALL BAYESIANS,...
Based on my priors
I will be just fine

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:

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

Notation

- $y_{i,t} = 1$ if individual i detected at occasion t and 0 otherwise
- $z_{i,t} = 1$ if individual i alive between occasions t and $t+1$ and 0 otherwise

$$
\begin{aligned}
y_{i,t} \mid z_{i,t} &\sim \text{Bernoulli}(p \, z_{i,t}) && \text{[likelihood (observation eq.)]} \\
z_{i,t+1} \mid z_{i,t} &\sim \text{Bernoulli}(\phi \, z_{i,t}) && \text{[likelihood (state eq.)]} \\
\phi &\sim \text{Beta}(1,1) && \text{[prior for } \phi \text{]} \\
p &\sim \text{Beta}(1,1) && \text{[prior for p]}
\end{aligned}
$$

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?

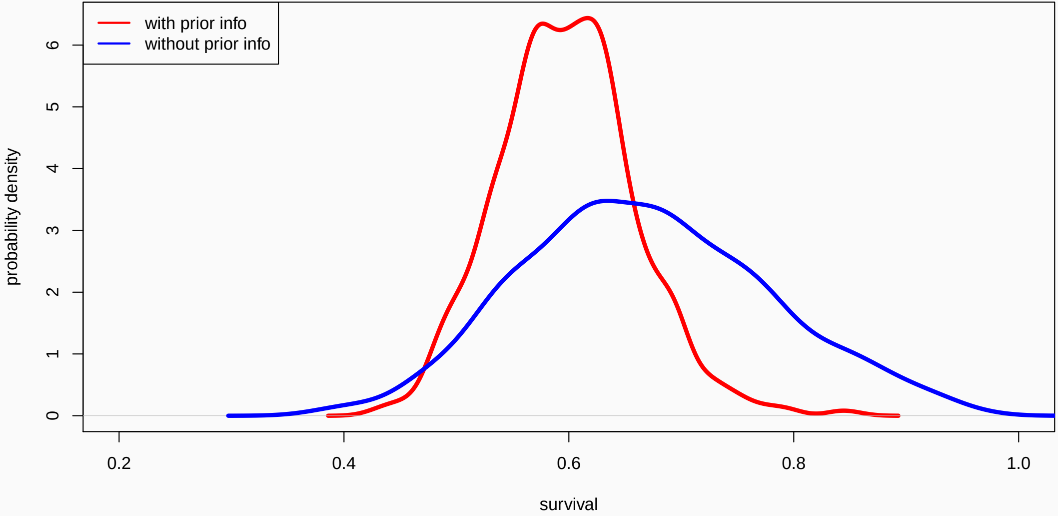
How to incorporate prior information?

- 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 $X \sim \mathit{Beta}(\alpha, \beta)$, then the first and second moments of X are:

$$\mu = \mathsf{E}(X) = \frac{\alpha}{\alpha + \beta}$$

$$\sigma^2 = \mathsf{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

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*(α, β) 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*(α, β) distribution.
- Now we look for values of α and β 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*(α, β) distribution.
- Now we look for values of α and β that match the observed moments of the Beta distribution (μ and σ^2).
- We need another set of equations:

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

$$\beta = \alpha\left(\frac{1}{\mu} - 1\right)$$

- For our model, that means:

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

- For our model, that means:

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

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

Your turn

Question

Use simulations to check that our estimates are correct.

Solution

R code

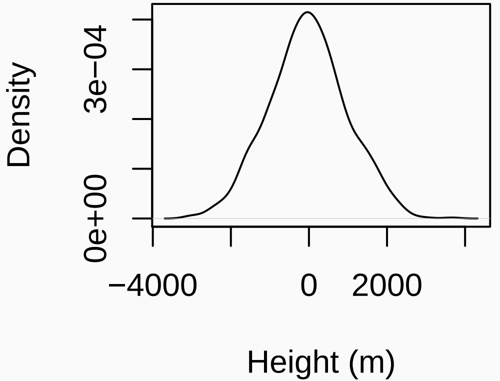
```r
alpha <- ( (1 - 0.57)/(0.073*0.073) - (1/0.57) )*0.57^2
beta <- alpha * ( (1/0.57) - 1)
n <- 10000
samp <- rbeta(n, alpha, beta)
(mu <- mean(samp))
(sigma <- sqrt(var(samp)))
```

Prior predictive checks

Linear regression

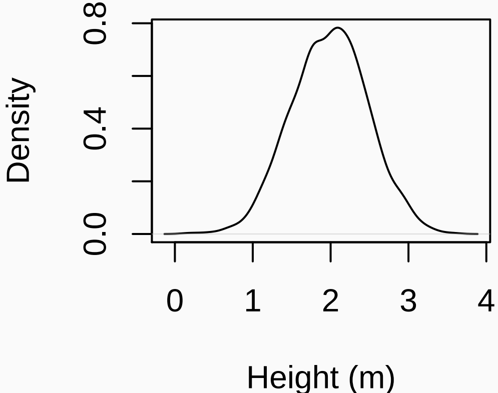
Unreasonable prior $\beta \sim N(0, 1000^2)$

```r
plot(density(rnorm(1000, 0, 1000)),
     main="", xlab="Height (m)")
```



Reasonable prior $\beta \sim N(2, 0.5^2)$

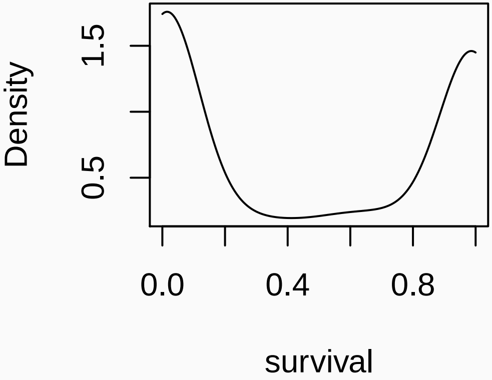
```r
plot(density(rnorm(1000, 2, 0.5)),
     main="", xlab="Height (m)")
```



Logistic regression

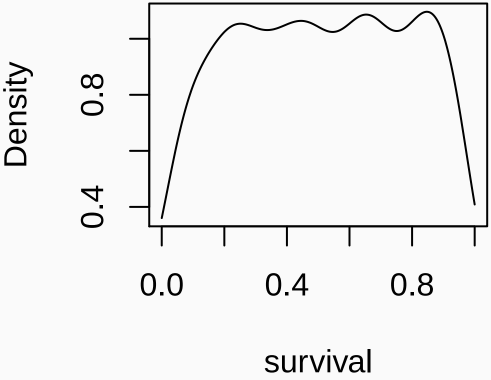
Unreasonable $\text{logit}(\phi) = \beta \sim N(0, 10^2)$

```
plot(density(plogis(rnorm(1000,0,10)),
from = 0, to = 1), main='', xlab='surv
```



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

```
plot(density(plogis(rnorm(1000,0,1.5))
from = 0, to = 1), main='', xlab='surv
```



Dynamic updating

Today's posterior is tomorrow's prior

If you obtain more data, no need to redo all of the analysis. Your posterior from the first analysis simply becomes your prior for the next analysis (and so on).

Today's posterior is tomorrow's prior

If you obtain more data, no need to redo all of the analysis. Your posterior from the first analysis simply becomes your prior for the next analysis (and so on).

- Stage 0. Prior $p(\theta) \sim \text{Beta}(1, 1)$.

Today's posterior is tomorrow's prior

If you obtain more data, no need to redo all of the analysis. Your posterior from the first analysis simply becomes your prior for the next analysis (and so on).

- Stage 0. Prior $p(\theta) \sim \text{Beta}(1, 1)$.
- Stage 1. Observe $y_1 = 22$ successes from $n_1 = 29$ trials.
    - Likelihood is $p(y_1|\theta) \sim \text{Binomial}(n_1 = 29, \theta)$.
    - Posterior is $p(\theta|y_1) \sim \text{Beta}(23, 8)$ with mean $23/31 = 0.74$.

Today's posterior is tomorrow's prior

If you obtain more data, no need to redo all of the analysis. Your posterior from the first analysis simply becomes your prior for the next analysis (and so on).

- Stage 0. Prior $p(\theta) \sim \text{Beta}(1, 1)$.
- Stage 1. Observe $y_1 = 22$ successes from $n_1 = 29$ trials.
  - Likelihood is $p(y_1|\theta) \sim \text{Binomial}(n_1 = 29, \theta)$.
  - Posterior is $p(\theta|y_1) \sim \text{Beta}(23, 8)$ with mean $23/31 = 0.74$.
- Stage 2. Observe $y_2 = 5$ successed from $n_2 = 10$ new trials.
  - Likelihood is $p(y_2|\theta) \sim \text{Binomial}(n_2 = 10, \theta)$.
  - Prior is $p(\theta) \sim \text{Beta}(23, 8)$ from stage 1.
  - Posterior is $p(\theta|y_1 \text{ and } y_2) \propto p(\theta|y_1)p(y_2|\theta) = \text{Beta}(28, 13)$ with mean $28/41 = 0.68$.