# Introduction to Bayesian statistics with R
# 4. Priors

Olivier Gimenez

last updated: 2025-03-10

# To-do list

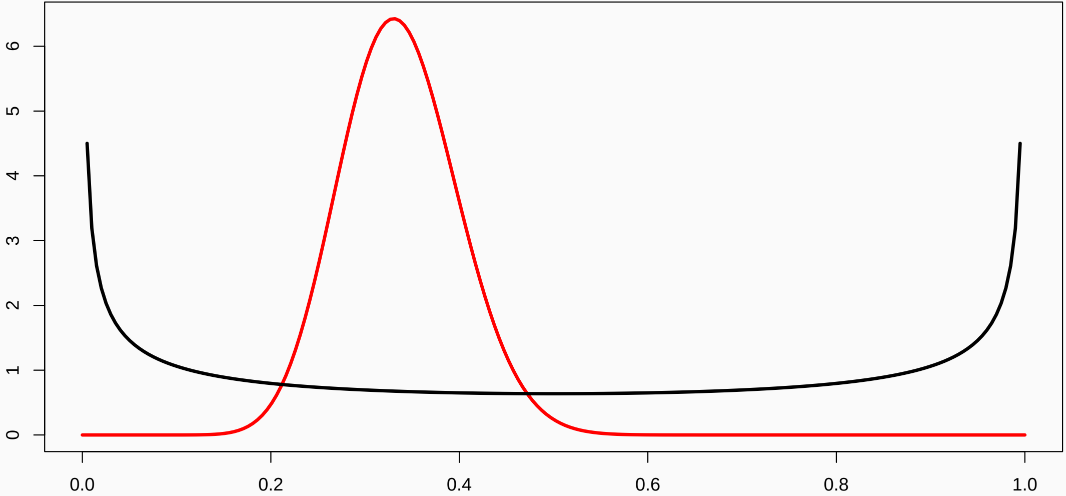
# To-do list

- add graph with gamma distribution
- check McCarthy paper w/ example ANOVA
- check paper big bad ugly prior
- check tips paper
- add further reading (McCarthy JAE, paper bad and ugly and paper on problem with priors in occupancy)
- check to-do list on github page of stats course https://github.com/oliviergimenez/bayesian-stats-with-R?tab=readme-ov-file#to-do-list
- prior elicitation (voir papiers JOSS suggérés par Osvaldo pour papier tips ; moment matching, voir aussi https://umr-astre.pages.mia.inra.fr/training/crashbayes/articles/crashbayes.html#prior-elicitation).

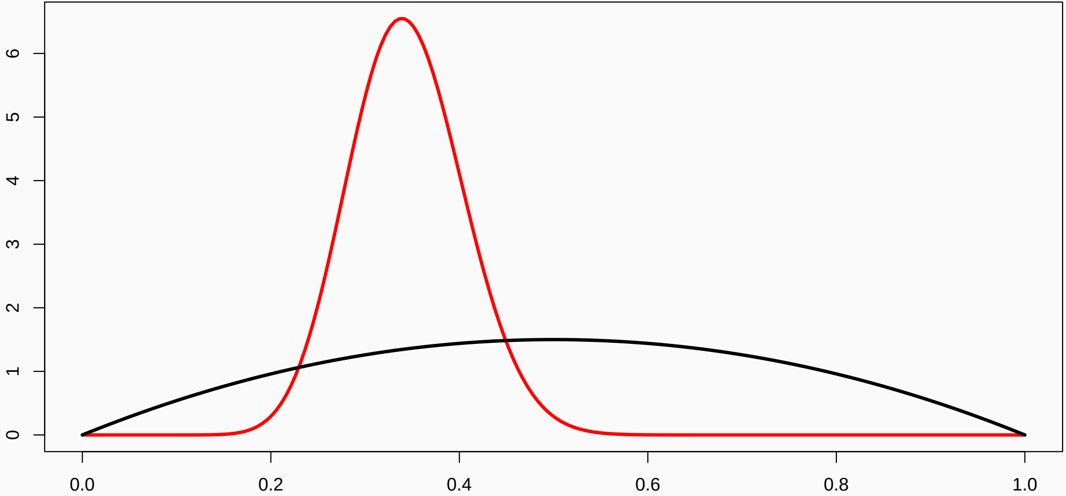
# A detour to explore priors

# Influence of the prior

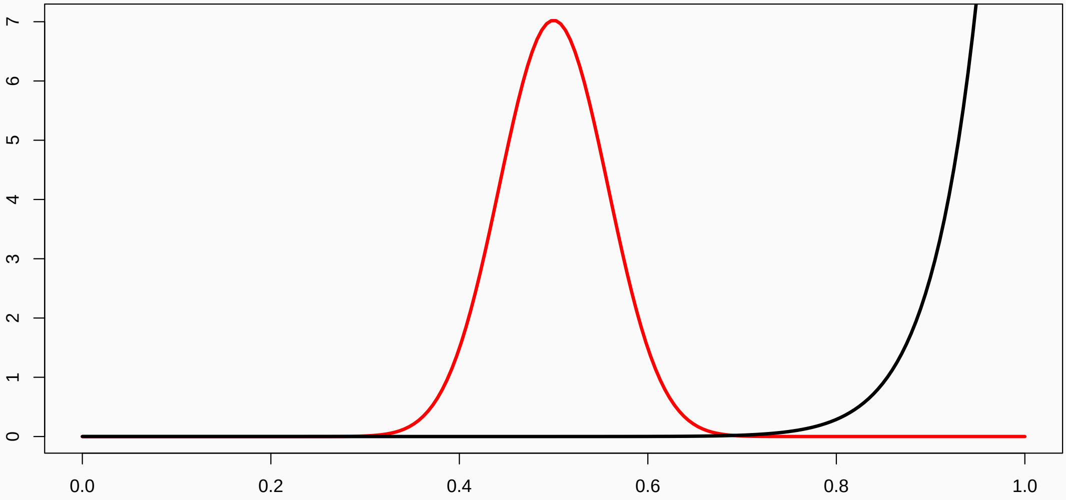
# Prior $Beta(0.5, 0.5)$ and posterior survival $Beta(19.5, 38.5)$



# 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.

# 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 $\phi$ 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 $\phi$ 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 $\phi$ 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?

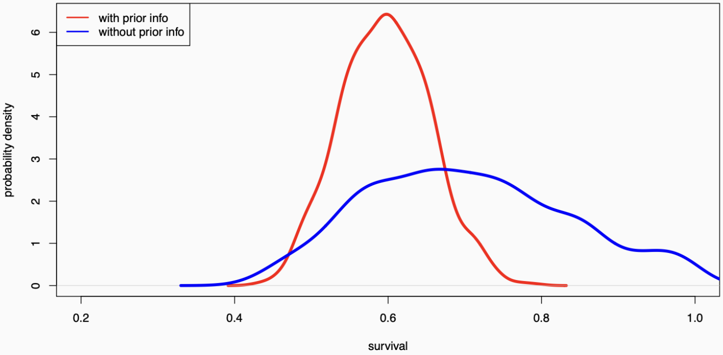
# How to incorporate prior information?

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

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

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

# 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 $\mu$ and $\sigma^2$ are seen as the moments of a *Beta*$(\alpha, \beta)$ 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 $\mu$ and $\sigma^2$ are seen as the moments of a *Beta*$(\alpha, \beta)$ distribution.
- Now we look for values of $\alpha$ and $\beta$ that match the observed moments of the Beta distribution ($\mu$ and $\sigma^2$).

# 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 $\mu$ and $\sigma^2$ are seen as the moments of a *Beta*$(\alpha, \beta)$ distribution.
- Now we look for values of $\alpha$ and $\beta$ that match the observed moments of the Beta distribution ($\mu$ and $\sigma^2$).
- 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:

```
(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:

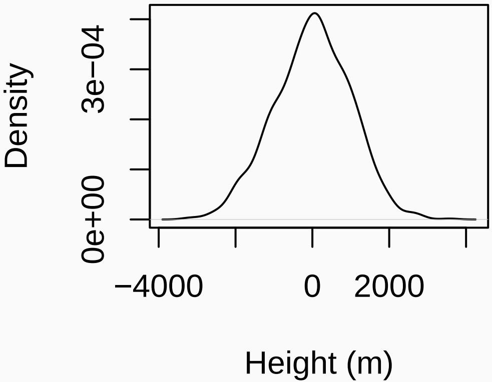
```
(alpha <- ( (1 - 0.57)/(0.073*0.073) - (1/0.57) )*0.57^2)
#> [1] 25.64636
(beta <- alpha * ( (1/0.57) - 1))
#> [1] 19.34726
```

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

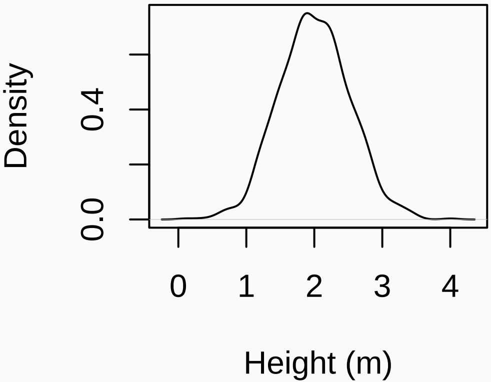
# Prior predictive checks

## Linear regression

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

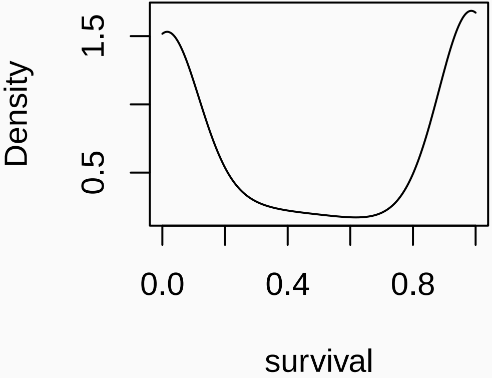



Reasonable prior $\beta \sim N(2, 0.5^2)$



# Logistic regression

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




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