Bayesian statistics 1 - Objectives, philosophy and a simple example

Frédéric Barraquand (CNRS, IMB)

October 18, 2022

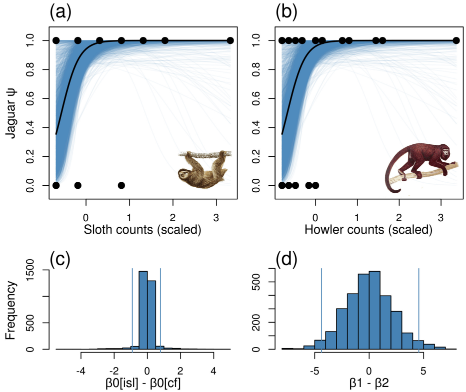
Bayesian statistics, what are they?

Thomas Bayes

Pierre-Simon de Laplace

- An application of Bayes' theorem (read to the Royal Society in 1763 after Bayes's death, rediscovered by Laplace among others)
- A way of doing statistics that's complementary to *frequentist* statistics (aka classical statistics) // sometimes I just run two sets of analyses...
- A good way to represent uncertainty in statistical results

Motivating examples



Predictors of jaguar occupancy across sample sites in Mamirauá's floodplain forest in Central Amazonia, taken from the paper below by Rabelo et al.

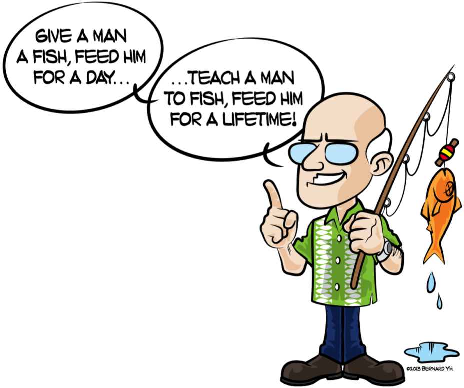
- Prey abundance drives habitat occupancy by jaguars in Amazonian floodplain river islands
- Imperial College's modelling of COVID-19
- Local Biomass Baselines and the Recovery Potential for Hawaiian Coral Reef Fish Communities

Course objectives

At the end of the course I hope you'll have

- Acquired familiarity with the concepts
- Be able to formulate (simple) models in BUGS (JAGS)
- Diagnose whether your model converges and fits
- Be able to read the literature on more complex models
- Have a more comprehensive view of uncertainty in statistics
  // no more “but wait – is this significant?"

Do’s and don’ts of this course



We will

- Do a mixture of theory/maths and practice examples/code
- Use case studies that illustrate general issues

We won’t

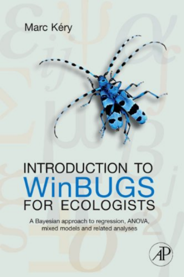
- Analyse one’s own data related to one’s own specific problems → better to learn
- Do cookbook stats (in situation S, apply method M)

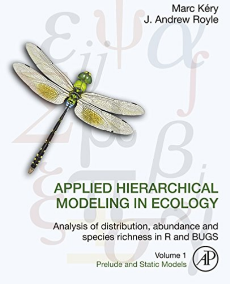
Plan for the whole course

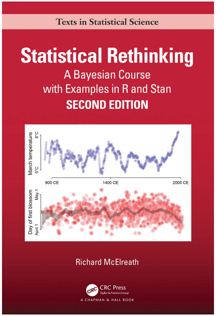
1. Objectives and philosophy of Bayesian statistics. TD1 Bayesian estimation of a proportion
2. Revisiting the ANOVA in a Bayesian framework. TD2 Getting acquainted with software (JAGS), coding the first models
3. Markov Chain Monte Carlo (i.e., algorithms for Bayesian statistics). Practicals within the course: Monte Carlo integration, rejection sampling, Metropolis algorithm.
4. From fixed to random effects, introduction to mixed models. TD4 variance partitioning (with thorough convergence diagnostics).
5. Mixed models. A hint of Poisson GLMs. TD5 mixed models following up on TD4 (done first)
6. Generalized linear models for counts. TD6 GLM(Ms) Poisson LN (fitting diagnostics, posterior predictive checks).
7. Binomial/Bernoulli GLM(Ms) (importance of priors in original and transformed scale). TD7 Binomial ANOVA.
8. Nonlinear models (organism growth, population growth). TD8 Gompertz organism growth.
9. Latent variable models. TD9 occupancy model (0/1 data with added observation process).
10. Model selection in a Bayesian setting. TD10 Linear and nonlinear model comparison.

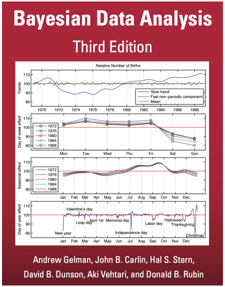
Credits: colleagues, books,...

Inspired from classes by Boris Hejblum (first three), Juliette Archambeau (this lecture), Olivier Gimenez. As well as









and perhaps many others...

Bayes theorem applied – mind gymnastics

Conditional probabilities

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\text{Card}(A \cap B)}{\text{Card}(B)}$$

A = it rains, B = Roger has an umbrella. Number of events leading to both Roger with an umbrella and rain, divided by the number of events where Roger has an umbrella.

Bayes theorem applied – mind gymnastics

Conditional probabilities

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathrm{Card}(A \cap B)}{\mathrm{Card}(B)}$$

A = it rains, B = Roger has an umbrella. Number of events leading to both Roger with an umbrella and rain, divided by the number of events where Roger has an umbrella. So you can write

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A)$$

and

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$$

Bayes theorem applied – mind gymnastics

Conditional probabilities

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\text{Card}(A \cap B)}{\text{Card}(B)}$$

A = it rains, B = Roger has an umbrella. Number of events leading to both Roger with an umbrella and rain, divided by the number of events where Roger has an umbrella. So you can write

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A)$$

and

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$$

leading to

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Bayes theorem for Bayesian statistics

$$\mathbb{P}(\textit{hypothesis}|\textit{data}) = \frac{\mathbb{P}(\textit{data}|\textit{hypothesis})\mathbb{P}(\textit{hypothesis})}{\mathbb{P}(\textit{data})}$$

aka $$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

Bayes theorem for Bayesian statistics

$$\mathbb{P}(\textit{hypothesis}|\textit{data}) = \frac{\mathbb{P}(\textit{data}|\textit{hypothesis})\mathbb{P}(\textit{hypothesis})}{\mathbb{P}(\textit{data})}$$

aka Posterior $\propto$ Likelihood $\times$ Prior

If your hypothesis materializes in a parameter set θ, then

$$\mathbb{P}(\theta|\textit{data}) = \frac{\mathbb{P}(\textit{data}|\theta)\mathbb{P}(\theta)}{\mathbb{P}(\textit{data})}$$

Posterior probability of θ $\propto$ Likelihood $\times$ probability of θ

Bayes theorem with random variables

Let's Z and Y be random variables with observed values z and y. First case, z and y take discrete values e.g. in $\{0, 1\}$ // heads or tails, $X = 1$ if heads $X = 0$ if tails. Examples: Bernoulli, Binomial, Geometric,... distributions. Then replacing A by $\{Z = z\}$ and B by $\{Y = y\}$ we get

$$\mathbb{P}(Z = z|Y = y) = \frac{\mathbb{P}(Y = y|Z = z)\mathbb{P}(Z = z)}{\mathbb{P}(Y = y)} = \frac{\mathbb{P}(Y = y, Z = z)}{\mathbb{P}(Y = y)}$$

Bayes theorem with random variables

Let's Z and Y be random variables with observed values z and y. First case, z and y take discrete values e.g. in $\{0,1\}$ // heads or tails, $X=1$ if heads $X=0$ if tails. Examples: Bernoulli, Binomial, Geometric,... distributions. Then replacing A by $\{Z=z\}$ and B by $\{Y=y\}$ we get

$$\mathbb{P}(Z=z|Y=y) = \frac{\mathbb{P}(Y=y|Z=z)\mathbb{P}(Z=z)}{\mathbb{P}(Y=y)} = \frac{\mathbb{P}(Y=y,Z=z)}{\mathbb{P}(Y=y)}$$

Second case, Z and Y are absolutely continuous random variables. Examples: Gaussian, Exponential, Gamma, Log-Normal, ... Then then admit *probability density functions* $f_Z(z)$ and $f_Y(y)$ so that

$$\mathbb{P}(a<Y<b) = \int_a^b f_Y(y)dy$$

We define similarly *conditional densities* $f_{Y|Z}(y|z) = \frac{f_{Y,Z}(y,z)}{f_Z(z)}$.

Bayes theorem with random variables

Let's Z and Y be random variables with observed values z and y. First case, z and y take discrete values e.g. in $\{0,1\}$ // heads or tails, $X=1$ if heads $X=0$ if tails. Examples: Bernoulli, Binomial, Geometric,... distributions. Then replacing A by $\{Z=z\}$ and B by $\{Y=y\}$ we get

$$\mathbb{P}(Z=z|Y=y)=\frac{\mathbb{P}(Y=y|Z=z)\mathbb{P}(Z=z)}{\mathbb{P}(Y=y)}=\frac{\mathbb{P}(Y=y,Z=z)}{\mathbb{P}(Y=y)}$$

Second case, Z and Y are absolutely continuous random variables. Examples: Gaussian, Exponential, Gamma, Log-Normal, ... Then then admit *probability density functions* $f_Z(z)$ and $f_Y(y)$ so that

$$\mathbb{P}(a<Y<b)=\int_a^b f_Y(y)dy$$

We define similarly *conditional densities* $f_{Y|Z}(y|z)=\frac{f_{Y,Z}(y,z)}{f_Z(z)}$. So that we get

$$f_{Z|Y}(z|y)=\frac{f_{Y|Z}(y|z)\times f_Z(z)}{f_Y(y)}$$

Why all the maths?

Now let's assume that Y represents your data measurements from an experiments, and we replace Z by Θ your parameter values. Apply the formula

$$\mathbb{P}(\Theta = \theta | Y = y) = \mathbb{P}(Y = y | \Theta = \theta) \times \mathbb{P}(\Theta = \theta) \times \frac{1}{\mathbb{P}(Y = y)}$$

which is equivalent to

$$\underbrace{\mathbb{P}(\text{parameters have value } \theta | \text{data})}_{\text{Posterior}} = \underbrace{\mathbb{P}(\text{data} | \theta)}_{\text{Likelihood}} \times \underbrace{\mathbb{P}(\text{parameters have value } \theta)}_{\text{Prior}} \times \frac{1}{\mathbb{P}(\text{data})}$$

or for continuously distributed parameters and data (dropping cumbersome indices)

$$\underbrace{f(\theta|y)}_{\text{Posterior}} = \underbrace{f(y|\theta)}_{\text{Likelihood}} \times \underbrace{f(\theta)}_{\text{Prior}} \times \frac{1}{f(y)}$$

We don't know the last bit, but it is a constant independent from the model, which will be useful in the computation.

Probability: several *interpretations* but only one *definition*

Often one hears:

- Frequentists view probability has a proportion of an event occurring over many experiments
- Bayesian view probability has degree of beliefs

Partially true (works when thinking about uncertainty intervals around parameters) – these views can be exchanged, combined,...

Only one *definition* of probability, introduced by Andreï Kolmogorov in 1933!
Third axiom (most important):

$$\mathbb{P}(E_1 \cup E_2 \cup E_3 \cup ...) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) + ...$$

if $E_1, E_2, ...$ are *mutually exclusive*. Probabilities of disjoint sets of events add up.
So what's the big difference between frequentist and Bayesian analysis?

Back to frequentist analyses

Linear regression example, parameter set $\theta = (a, b)$.

$$Y_i = a + bx_i + E_i,\; E_i \sim \mathcal{N}(0, \sigma^2)$$

In frequentist analyses, *a* and *b* are fixed–they have a single value. You will then test $H_0 : \{b = 0\}$ against $H_1 : \{b \neq 0\}$.

Pause: what's a p-value? I color-code where the data is in red.

Back to frequentist analyses

Linear regression example, parameter set $\theta = (a, b)$.

$$Y_i = a + bx_i + E_i, \; E_i \sim \mathcal{N}(0, \sigma^2)$$

In frequentist analyses, *a* and *b* are fixed–they have a single value. You will then test $H_0 : \{b = 0\}$ against $H_1 : \{b \neq 0\}$.

Pause: what's a p-value? I color-code where the data is in red.

$\mathbb{P}(T > t_{\text{obs}} | H_0)$ where T is a test statistic, or again
$\mathbb{P}(\text{data more extreme than observed}|\text{null hypothesis})$.

Here $p = \mathbb{P}(T > t_{\text{obs}} | b = 0)$. Completely different from $\mathbb{P}(b \neq 0|\text{data}) \rightarrow$ what most people are interested in.

A parenthesis on p-values and confidence intervals

So many critiques I can't list them – just one recent example (Amrhein et al., 2019):

> *Let's be clear about what must stop: we should never conclude there is 'no difference' or 'no association' just because a P value is larger than a threshold such as 0.05 or, equivalently, because a confidence interval includes zero. Neither should we conclude that two studies conflict because one had a statistically significant result and the other did not. These errors waste research efforts and misinform policy decisions.*

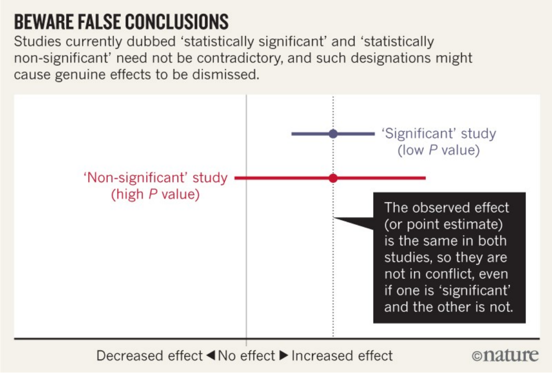
No problem with p-values per se, just that interpretation is so different from what people believe...

See also this Perspective by Regina Nuzzo.

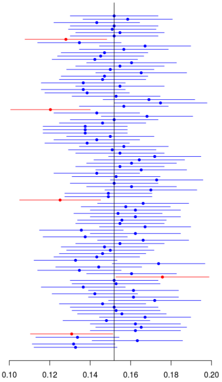
A parenthesis on p-values and confidence intervals

We know theoretically stuff about p-values *under the null hypothesis*: they are uniformly distributed between 0 and 1 (so $p < \alpha = 5\%$ gives you 1 false positives in 20).

However, H_1 is often true (true $b \neq 0$). Then p-values are affected by sample size, power of the test, ... Few data points $\rightarrow$ large p-values even with reasonably sized effects.



A parenthesis on p-values and confidence intervals



Black line = true value. https://freakonometrics.hypotheses.org/18117

A confidence interval at a 95% significance level will contain the true value in 95% of the experiments. By contrast a *Bayesian credible interval* $[\theta_{\text{low}}, \theta_{\text{high}}]$ means that $\mathbb{P}(\theta \in [\theta_{\text{low}}, \theta_{\text{high}}] | \text{data}) = 95\%$. For large datasets this often matches.

In Bayesian analysis, everything is is random variable

Linear regression example

$$Y_i = a + bx_i + E_i,\ E_i \sim \mathcal{N}(0, \sigma^2)$$

We have *priors*: $a \sim \mathcal{N}(0, \sigma_a^2)$, $b \sim \mathcal{N}(0, \sigma_b^2)$ (for instance).

Then you can obtain $\mathbb{P}(b > 0)$ or $\mathbb{P}(b > 0.1)$, from the density for $\theta = (a, b)$,
$\mathbb{P}(b > 0.1) = \int_{\infty}^{\infty} \int_{0.01}^{\infty} f(a, b) dadb$.

Also *a posteriori* probabilities $\mathbb{P}(b > 0|\text{data})$ or $\mathbb{P}(b > 0.1|\text{data})$, using the Bayes formula

$$f(\theta|\text{data}) = \underbrace{f(\text{data}|\theta)}_{\text{likelihood}} \underbrace{f(\theta)}_{\text{prior}} \underbrace{\frac{1}{f(\text{data})}}_{\text{constant}}$$

We will see how to obtain those in practice in lecture 2, using the JAGS software. How exactly we deal with the annoying constant using MCMC (Markov Chain Monte Carlo) will be tackled in lecture 3.

In Bayesian analysis, everything is is random variable II

Wait, does that mean I should write my regression

$$Y_i = A + Bx_i + E_i, \; E_i \sim \mathcal{N}(0, \sigma^2)$$

because A and B are random variables? Is this a random effect or mixed model??

In Bayesian analysis, everything is is random variable II

Wait, does that mean I should write my regression

$$Y_i = A + Bx_i + E_i,\; E_i \sim \mathcal{N}(0, \sigma^2)$$

because *A* and *B* are random variables? Is this a random effect or mixed model??

Tentative answer: there are no more mixed models in Bayesian analysis – or rather, it depends on your data and your priors.

We will see more in lecture 4 and 5.

A simple example: Bayesian estimation of a proportion

Gyrfalcon provenance

Groenland Falcon & Iceland Falcon 1915 - Artiste Archibald Thorburn (1860 - 1935)

Gyrfalcons are the largest of all falcons and were fancied by European courts for hunting.

Nielsen & Pétursson (1995) reported that 4,848 gyrfalcons were exported from Iceland to Denmark during 1731–1793. Of these, 4,318 were grey birds, 374 white and 156 half-white (we sum these to 530).

We know that $\mathbb{P}(\text{white}) \approx \mathbb{P}(\text{from Greenland})$. What is the proportion of the birds caught in Iceland that were actually migrants from Greenland?

A simple example: Bayesian estimation of a proportion

We need three things

1. A question. What is $\mathbb{P}$(white)?
2. A model for the data $\rightarrow$ Likelihood part
3. Prior distribution for the parameter(s) of interest.

A simple example: Bayesian estimation of a proportion

We need three things

1. A question. What is $\mathbb{P}(\text{white})$?
2. A model for the data $\rightarrow$ Likelihood part
   $X_i \sim \text{Bernoulli}(\theta)$, with $X_i = 1$ if white, $X_i = 0$ if grey.
3. Prior distribution for the parameter(s) of interest. For now, $\theta \sim \text{Uniform}([0, 1])$

A simple example: Bayesian estimation of a proportion

We need three things

1. A question. What is $\mathbb{P}(\text{white})$?
2. A model for the data $\rightarrow$ Likelihood part
   $X_i \sim \text{Bernoulli}(\theta)$, with $X_i = 1$ if white, $X_i = 0$ if grey.
   $l(x|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{y}(1-\theta)^{n-y}$ with $y = x_1 + ... + x_n$.
3. Prior distribution for the parameter(s) of interest. For now, $\theta \sim \text{Uniform}([0,1])$
   $\pi(\theta) = \frac{1}{b-a} = 1$ with $a = 0$ and $b = 1$.

Getting the posterior

$$\mathbb{P}(\theta|data) = \frac{\mathbb{P}(data|\theta)\mathbb{P}(\theta)}{\mathbb{P}(data)}$$

Here, what do we write?

Getting the posterior

$$\mathbb{P}(\theta|data) = \frac{\mathbb{P}(data|\theta)\mathbb{P}(\theta)}{\mathbb{P}(data)}$$

$$p(\theta|x) = \frac{l(x|\theta)\pi(\theta)}{m(x)}$$

Marginal or integrated likelihood $m(x) = \int l(x|\theta)\pi(\theta)d\theta$. In general very hard to compute especially if θ is high-dimensional.

Getting the posterior

$$\mathbb{P}(\theta|data) = \frac{\mathbb{P}(data|\theta)\mathbb{P}(\theta)}{\mathbb{P}(data)}$$

$$p(\theta|x) = \frac{l(x|\theta)\pi(\theta)}{m(x)}$$

Marginal or integrated likelihood $m(x) = \int l(x|\theta)\pi(\theta)d\theta$. In general very hard to compute especially if θ is high-dimensional.

Here $m(x) = m(y) = \int \theta^y(1-\theta)^{n-y}d\theta$ we are lucky – this is known as the Beta function $m(y) = B(y+1, n-y+1) = \frac{(y)!(n-y)!}{(n+1)!} = \frac{1}{n+1}\frac{1}{\binom{n}{y}}$.

Getting the posterior

$$\mathbb{P}(\theta|data) = \frac{\mathbb{P}(data|\theta)\mathbb{P}(\theta)}{\mathbb{P}(data)}$$

$$p(\theta|x) = \frac{l(x|\theta)\pi(\theta)}{m(x)}$$

Marginal or integrated likelihood $m(x) = \int l(x|\theta)\pi(\theta)d\theta$. In general very hard to compute especially if θ is high-dimensional.

Here $m(x) = m(y) = \int \theta^y (1-\theta)^{n-y} d\theta$ we are lucky – this is known as the Beta function $m(y) = B(y+1, n-y+1) = \frac{(y)!(n-y)!}{(n+1)!} = \frac{1}{n+1}\frac{1}{\binom{n}{y}}$.

Eventually we obtain

$$p(\theta|x) = (n+1)\binom{n}{y}\theta^y(1-\theta)^{n-y}$$

which can also be written

$$p(\theta|x) = \frac{1}{B(y+1, n-y+1)}\theta^y(1-\theta)^{n-y}$$

so that the posterior distribution follows a Beta$(y+1, n-y+1)$ distribution.

Finding estimates

- MAP: Maximum A Posteriori. $\hat{\theta}$ so that $p(\theta|y)$ is maximal. $\hat{\theta} = \frac{y}{n}$. (Proof at the end of the presentation). Same as max likelihood.

Finding estimates

- MAP: Maximum A Posteriori. $\hat{\theta}$ so that $p(\theta|y)$ is maximal. $\hat{\theta} = \frac{y}{n}$. (Proof at the end of the presentation). Same as max likelihood.
- Mean $\mathbb{E}(\theta) = \int \theta p(\theta|y) d\theta$, in this case $\theta = \frac{y+1}{n+2}$.

Finding estimates

- MAP: Maximum A Posteriori. $\hat{\theta}$ so that $p(\theta|y)$ is maximal. $\hat{\theta} = \frac{y}{n}$. (Proof at the end of the presentation). Same as max likelihood.
- Mean $\mathbb{E}(\theta) = \int \theta p(\theta|y) d\theta$, in this case $\theta = \frac{y+1}{n+2}$.
- Median ...

New priors and new data

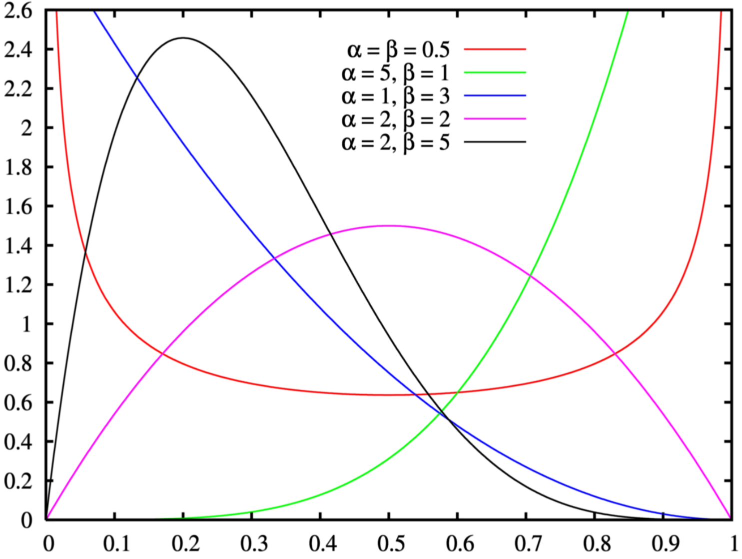
My colleague Olafur studies falcons nowadays. Every now and then he sees some whites, but perhaps less. Let's assume this year he has seen 74 grey and 5 whites. What's the percentage nowadays ?

We consider three scenarios for priors:

1. Uniform
2. Prior = posterior for 1731–1793
3. Prior loosely concentrated around previous mean (10.9%) // since we're not in 1794.

This time we'll plot the whole prior and posterior distributions to compare.

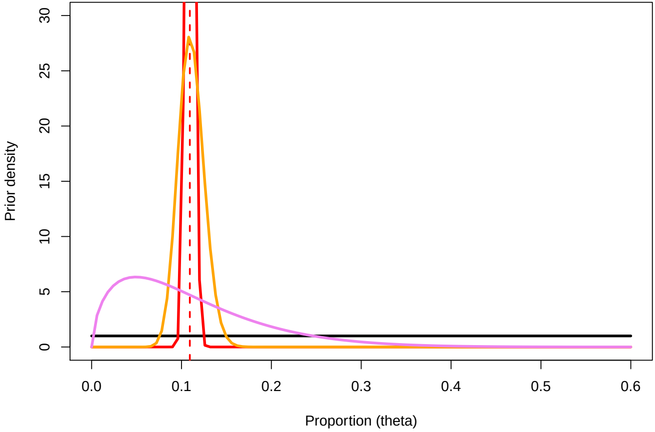
The Beta distribution $B(\alpha, \beta)$



Beta distribution probability density

Priors

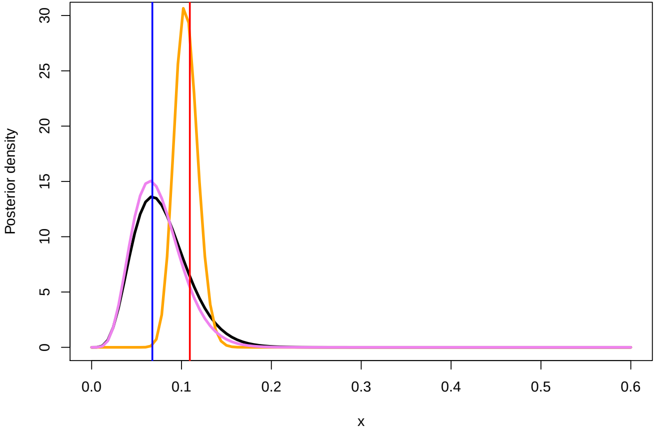
Beta$(1, 1)$ = uniform, Beta$(53 + 1, 485 + 1)$, Beta$(1.63, 13.36)$



Priors 1 (black), 2 (red = previous posterior / orange 10 times less data), 3 (violet).

Posteriors

If the prior is Beta$(1, 1)$ the posterior is Beta$(y + 1, n - y + 1)$. More generally, if the prior is Beta(a, b), the posterior is Beta$(y + a, n - y + b)$.



Posteriors 1 (black), 2 (orange), 3 (violet).
Red = estimate 18th century. Blue = observed now.

Exercise

Let's code these priors and posteriors in R (i.e., reproduce the two previous figures).

Hint: use `dbeta(x,a,b)` for B(a,b) as well as the `curve()` function.

If you have time, compute the 95% credibility intervals for the three posteriors. Hint: use `qbeta()`.

Nota bene Using new prior = previous posterior is a way to add up the data. This can be shown mathematically:

Consider a first experiment with prior Beta(a, b). Then the posterior has a Beta$(y + a, n - y + b)$ distribution. You do a second experiment using new prior = previous posterior. Consequently, the new prior is Beta(a', b') with $a' = y + a$ and $b' = n - y + b$. The new posterior has then distribution Beta$(y' + a', n' - y' + b') =$ Beta$(y' + y + a, n' - y' + n - y + b) =$ Beta$((y' + y) + a, (n' + n) - (y' + y) + b)$. Observations add up.

To wrap up

- Bayes theorem applied to random variables representing your data and parameters leads to

$$\mathbb{P}(\theta|data) = \frac{\mathbb{P}(data|\theta)\mathbb{P}(\theta)}{\mathbb{P}(data)}$$

- Bayesian analysis differs essentially from frequentist analysis in that both observations and parameters are assumed to arise from random variables
- Credible intervals are more complicated to compute (for simple models) than confidence intervals but logically more straightforward
- Bayesian and frequentist statistics give similar estimates when n is very large, but prior influence grows as sample size decreases.

Formulas for θ estimates

- MAP: Maximum A Posteriori. $\hat{\theta}$ so that $p(\theta|y)$ is maximal. $\hat{\theta} = \frac{y}{n}$.

$$\frac{\partial p}{\partial \theta} = \theta^{y-1}(1-\theta)^{n-y-1}(y-n\theta)$$

Formulas for θ estimates

- MAP: Maximum A Posteriori. $\hat{\theta}$ so that $p(\theta|y)$ is maximal. $\hat{\theta} = \frac{y}{n}$.

$$\frac{\partial p}{\partial \theta} = \theta^{y-1}(1-\theta)^{n-y-1}(y - n\theta)$$

- Mean $\mathbb{E}(\theta) = \int \theta p(\theta|y) d\theta$, in this case $\theta = \frac{y+1}{n+2}$.

$$\int \theta \times (n+1)\binom{n}{y}\theta^y(1-\theta)^{n-y}d\theta = (n+1)\binom{n}{y} \times \int \theta^{y+1}(1-\theta)^{n-y}d\theta$$

$$= (n+1)\binom{n}{y}B(y+2, n-y+1)$$

We have $B(a+1, b) = B(a, b)\frac{a}{a+b}$ (property of the Beta function) so that $(n+1)\binom{n}{y}B(y+2, n-y+1) = (n+1)\binom{n}{y}B(y+1, n-y+1)\frac{y+1}{n+2}$ $= (n+1)\binom{n}{y}\frac{1}{n+1}\frac{1}{\binom{n}{y}}\frac{y+1}{n+2}$

Some refs.

Amrhein, V., Greenland, S. & McShane, B. (2019) Scientists rise up against statistical significance. Nature **567**, 305–308.

Nielsen, Ó.K. & Pétursson, G. (1995) Population fluctuations of gyrfalcon and rock ptarmigan: analysis of export figures from iceland. Wildlife Biology **1**, 65–71.