# Hyper-priors & hierarchical models

**Hierarchical levels:** 

1  $\pi(\theta)$ 

**2**  $f(y|\theta)$ 

Course presentation 0000000 Intro to Bayesian stat

Bayesian modeling

Bayesian Infer
0000000000

Conclusion 00

Going further

# Hyper-*priors* & hierarchical models

**Hierarchical levels:** 

1  $\eta \sim h(\eta)$ 

- 2  $\pi(\theta|\eta)$
- $3 f(y|\theta)$

34/50

### Hyper-*priors* & hierarchical models

**Hierarchical levels:** 

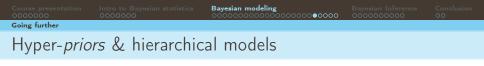
1  $\eta \sim h(\eta)$ 

2  $\pi(\theta|\eta)$ 

 $3 f(y|\theta)$ 

 $p(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{f(\mathbf{y})} = \frac{\int f(\mathbf{y}|\theta,\eta)\pi(\theta|\eta)h(\eta)d\eta}{f(\mathbf{y})}$ 

34/50



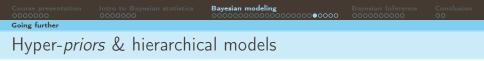
**Hierarchical levels:** 

**1**  $\eta \sim h(\eta)$ 

- 2  $\pi(\theta|\eta)$
- $\Im f(y|\theta)$

 $p(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{f(\mathbf{y})} = \frac{\int f(\mathbf{y}|\theta,\eta)\pi(\theta|\eta)h(\eta)d\eta}{f(\mathbf{y})} = \frac{f(\mathbf{y}|\theta)\int \pi(\theta|\eta)h(\eta)d\eta}{f(\mathbf{y})}$ 

**NB:** 3 hierarchical levels  $\Leftrightarrow$  two levels with prior:  $\pi(\theta) = \int \pi(\theta|\eta) h(\eta) d\eta$ 



**Hierarchical levels:** 

**1**  $\eta \sim h(\eta)$ 

- 2  $\pi(\theta|\eta)$
- $3 f(y|\theta)$

 $p(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{f(\mathbf{y})} = \frac{\int f(\mathbf{y}|\theta,\eta)\pi(\theta|\eta)h(\eta)d\eta}{f(\mathbf{y})} = \frac{f(\mathbf{y}|\theta)\int \pi(\theta|\eta)h(\eta)d\eta}{f(\mathbf{y})}$ 

**NB:** 3 hierarchical levels  $\Leftrightarrow$  two levels with *prior*:  $\pi(\theta) = \int \pi(\theta|\eta) h(\eta) d\eta$ 

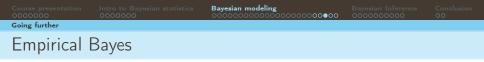
⇒ can ease modeling and elicitation of the prior...

## Hyperprior in the historical example

Historical example of birth sex with a Beta prior

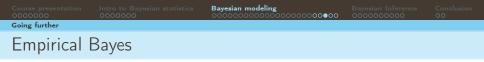
 $\Rightarrow$  two Gamma hyper-*priors* for  $\alpha$  and  $\beta$  (conjugated):

 $\begin{aligned} & \alpha \sim \text{Gamma}(4, 0.5) \\ & \beta \sim \text{Gamma}(4, 0.5) \\ & \theta | \alpha, \beta \sim \text{Beta}(\alpha, \beta) \\ & Y_i | \theta \stackrel{iid}{\sim} \text{Bernoulli}(\theta) \end{aligned}$ 



Eliciting the prior according to its empirical marginal distribution

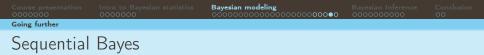
- $\Rightarrow$  estimate the *prior* from the data
  - 1 hyper-parameters
  - 2 estimate them through frequentist methods (e.g. MLE) by  $\hat{\eta}$
  - 3 plug-in estimates into the prior
  - **4**  $\Rightarrow$  posterior:  $p(\theta|\mathbf{y}, \hat{\eta})$



Eliciting the prior according to its empirical marginal distribution

- $\Rightarrow$  estimate the *prior* from the data
  - hyper-parameters
  - 2 estimate them through frequentist methods (e.g. MLE) by  $\hat{\eta}$
  - 3 plug-in estimates into the prior
  - **4**  $\Rightarrow$  posterior:  $p(\theta|\mathbf{y}, \hat{\eta})$
  - Combines Bayesian and frequentist frameworks
  - Concentrated *posterior*: \ variance but / bias (data used twice ⇒ shrinkage around the average!)
  - Approximate a fully Bayesian approach

36/50



Bayes' theorem can be used sequentially:

 $p(\theta|\mathbf{y}) \propto f(\mathbf{y}|\theta) \pi(\theta)$ 

If  $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$ , then:

 $p(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}_2|\boldsymbol{\theta}) f(\boldsymbol{y}_1|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \propto f(\boldsymbol{y}_2|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{y}_1)$ 

⇒ posterior distribution updates as new observations are aquired/available (online updates)

### Sequential Bayes in the historical example

Let's imagine that we start by observing 20 births  $y_{1:20}$  at the start of 1745, including 9 girls, and that we have a uniform *prior* on  $\theta$ :

 $\theta | \boldsymbol{y}_{1:20} \sim \dots$ 



## Sequential Bayes in the historical example

Let's imagine that we start by observing 20 births  $y_{1:20}$  at the start of 1745, including 9 girls, and that we have a uniform *prior* on  $\theta$ :

 $\theta|\mathbf{y}_{1:20} \sim \mathsf{Beta}(10, 12)$ 

Then we observe  $y_{21:493472}$  the remaining 493452 births between 1745 and 1770, including 241 936 girls, and we then uses this Beta(10,12) prior for  $\theta$ :

 $\theta | \mathbf{y}_{1:20}, \mathbf{y}_{21:493\,472} \sim \dots$ 

# Sequential Bayes in the historical example

Let's imagine that we start by observing 20 births  $y_{1:20}$  at the start of 1745, including 9 girls, and that we have a uniform *prior* on  $\theta$ :

 $\theta|\textbf{y}_{1:20} \sim \mathsf{Beta}(10, 12)$ 

Then we observe  $y_{21:493472}$  the remaining 493452 births between 1745 and 1770, including 241 936 girls, and we then uses this Beta(10,12) prior for  $\theta$ :

$$\theta | \mathbf{y}_{1:20}, \mathbf{y}_{21:493472} \sim \text{Beta}(10 + 241936, 12 + 251516)$$
  
~ Beta(241946, 251528)

We get the same *posterior* distribution as with all the observations taken together at once