Hyper-*priors* & hierarchical models

Hierarchical levels:

1. $\pi(\theta)$
2. $f(\boldsymbol{y}|\theta)$

Hyper-*priors* & hierarchical models

Hierarchical levels:

1. $\eta \sim h(\eta)$
2. $\pi(\theta|\eta)$
3. $f(\boldsymbol{y}|\theta)$

Hyper-*priors* & hierarchical models

Hierarchical levels:

1. $\eta \sim h(\eta)$
2. $\pi(\theta|\eta)$
3. $f(\boldsymbol{y}|\theta)$

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

Hyper-*priors* & hierarchical models

Hierarchical levels:

1. $\eta \sim h(\eta)$
2. $\pi(\theta|\eta)$
3. $f(\boldsymbol{y}|\theta)$

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

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

Hyper-*priors* & hierarchical models

Hierarchical levels:

1. $\eta \sim h(\eta)$
2. $\pi(\theta|\eta)$
3. $f(\boldsymbol{y}|\theta)$

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

NB: 3 hierarchical levels ⇔ 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*

⇒ two Gamma hyper-*priors* for α and β (conjugated):

$$\alpha \sim \text{Gamma}(4, 0.5)$$
$$\beta \sim \text{Gamma}(4, 0.5)$$
$$\theta | \alpha, \beta \sim \text{Beta}(\alpha, \beta)$$
$$Y_i | \theta \overset{iid}{\sim} \text{Bernoulli}(\theta)$$

Empirical Bayes

Eliciting the *prior* according to its empirical marginal distribution

⇒ 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. ⇒ *posterior*: $p(\theta|\mathbf{y},\hat{\eta})$

Empirical Bayes

Eliciting the *prior* according to its empirical marginal distribution

⇒ 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. ⇒ *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

Sequential Bayes

Bayes' theorem can be used sequentially:

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

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

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

⇒ *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 $\boldsymbol{y}_{1:20}$ at the start of 1745, including 9 girls, and that we have a uniform *prior* on θ :

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

Sequential Bayes in the historical example

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

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

Then we observe $\boldsymbol{y}_{21:493\,472}$ the remaining 493 452 births between 1745 and 1770, including 241 936 girls, and we then uses this $\text{Beta}(10, 12)$ *prior* for θ :

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

Sequential Bayes in the historical example

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

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

Then we observe $\boldsymbol{y}_{21:493\,472}$ the remaining 493 452 births between 1745 and 1770, including 241 936 girls, and we then uses this $\text{Beta}(10, 12)$ *prior* for θ :

$$\begin{aligned}\theta|\boldsymbol{y}_{1:20}, \boldsymbol{y}_{21:493\,472} &\sim \text{Beta}(10 + 241\,936, 12 + 251\,516)\\ &\sim \text{Beta}(241\,946, 251\,528)\end{aligned}$$

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