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):

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

Empirical Bayes

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*: $p(\theta|\boldsymbol{y}, \hat{\eta})$

Empirical Bayes

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

- Combines Bayesian and frequentist frameworks
- Concentrated *posterior* ($\searrow$ variance) but $\nearrow$ bias (data used twice !)
- 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:

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

$\Rightarrow$ *posterior* distribution updates as new observations are aquired/available (*online updates*)