La loi *a priori* de Jeffreys : application à l'exemple historique

$$f(y|\theta) = \theta^y (1-\theta)^{(1-y)}$$

$$\begin{aligned}
\pi(\theta) &\propto \sqrt{I(\theta)} \\
&\propto \sqrt{-\mathbb{E}_{Y|\theta}\left[\frac{\mathrm{d}^2 \log(f(y|\theta))}{\mathrm{d}\theta^2}\right]} \\
&\propto \dots
\end{aligned}$$

[Rappel : $I_{(Y_1,\dots,Y_n)}(\theta) = n \times I_Y(\theta)$ si les Y_i sont *iid*]

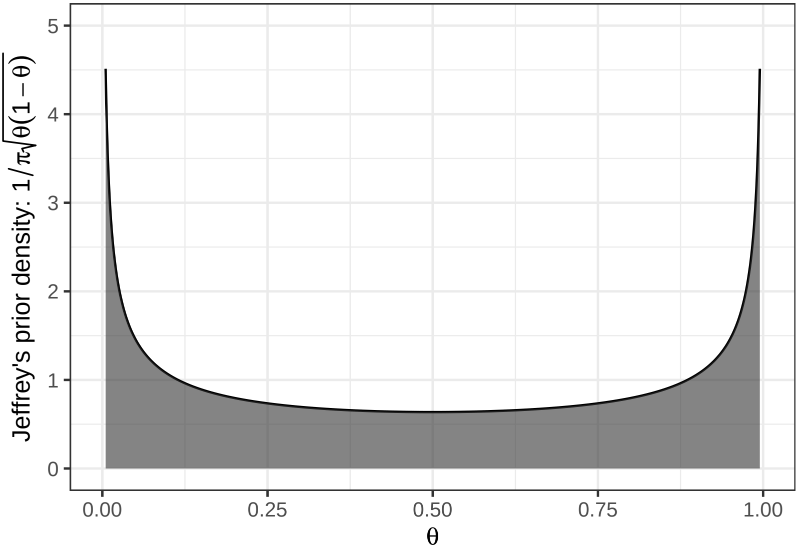
La loi *a priori* de Jeffreys : application à l'exemple historique

$$f(y|\theta) = \theta^y(1-\theta)^{(1-y)}$$

$$\begin{aligned}
\pi(\theta) &\propto \sqrt{I(\theta)} \\
&\propto \sqrt{-\mathbb{E}_{Y|\theta}\left[\frac{\mathrm{d}^2\log(f(y|\theta))}{\mathrm{d}\theta^2}\right]} \\
&\propto \sqrt{\mathbb{E}_{Y|\theta}\left[\frac{y}{\theta^2} + \frac{1-y}{(1-\theta)^2}\right]} \\
&\propto \sqrt{\theta\left(\frac{1}{\theta^2}\right) + (1-\theta)\left(\frac{1}{(1-\theta)^2}\right)} \\
&\propto \sqrt{\frac{1}{\theta} + \frac{1}{1-\theta}} \\
&\propto \frac{1}{\sqrt{\theta(1-\theta)}}
\end{aligned}$$

[Rappel : $I_{(Y_1,\dots,Y_n)}(\theta) = n \times I_Y(\theta)$ si les Y_i sont *iid*]

Loi *a priori* de Jeffreys : illustration dans l'exemple historique



Hyper-priors & modèles hiérarchiques

Niveaux hiérarchiques :

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

Hyper-priors & modèles hiérarchiques

Niveaux hiérarchiques :

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

Hyper-priors & modèles hiérarchiques

Niveaux hiérarchiques :

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 & modèles hiérarchiques

Niveaux hiérarchiques :

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 niveaux hiérarchiques ⇔ 2 niveaux avec $\pi(\theta) = \int \pi(\theta|\eta)h(\eta)d\eta$

Hyper-priors & modèles hiérarchiques

Niveaux hiérarchiques :

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 niveaux hiérarchiques ⇔ 2 niveaux avec $\pi(\theta) = \int \pi(\theta|\eta)h(\eta)d\eta$

⇒ peut **faciliter la modélisation** & l'**élicitation** des lois *a priori*

Utilisation d'hyper-priors dans l'exemple historique

Exemple historique du sexe à la naissance avec un *a priori* Beta
⇒ 2 hyper-priors Gamma pour α et β :

$$\alpha \sim \text{Gamma}(4;0,5)$$
$$\beta \sim \text{Gamma}(4;0,5)$$
$$\theta|\alpha,\beta \sim \text{Beta}(\alpha;\beta)$$
$$Y_i|\theta \overset{iid}{\sim} \text{Bernoulli}(\theta)$$

Bayésien empirique

Élicitation de la loi *a priori* d'un paramètre d'après sa loi marginale empirique

Bayésien empirique

Élicitation de la loi *a priori* d'un paramètre d'après sa loi marginale empirique

$\Rightarrow$ nécessite d'estimer cet *a priori* à partir des données

Bayésien empirique

Élicitation de la loi *a priori* d'un paramètre d'après sa loi marginale empirique

$\Rightarrow$ nécessite d'estimer cet *a priori* à partir des données

1. hyper-paramètres η
2. estimés par $\widehat{\eta}$ grâce à des méthodes fréquentistes (e.g. Max de Vraisemblance ou Méthode des Moments)
3. injectés dans la loi *a priori* : $\pi(\theta|\widehat{\eta})$
4. $\Rightarrow$ loi *a posteriori* : $p(\theta|\boldsymbol{y},\hat{\eta})$

Bayésien empirique

Élicitation de la loi *a priori* d'un paramètre d'après sa loi marginale empirique

$\Rightarrow$ nécessite d'estimer cet *a priori* à partir des données

1. hyper-paramètres η
2. estimés par $\widehat{\eta}$ grâce à des méthodes fréquentistes (e.g. Max de Vraisemblance ou Méthode des Moments)
3. injectés dans la loi *a priori* : $\pi(\theta|\widehat{\eta})$
4. $\Rightarrow$ loi *a posteriori* : $p(\theta|\boldsymbol{y},\hat{\eta})$

- Combine les approches bayésienne et fréquentiste
- Distribution *a posteriori* concentrée (variance $\searrow$), mais biais $\nearrow$ (données utilisées 2x !)
- Approximation d'une approche totalement bayésienne

Bayésien empirique : exemple

Pour une loi Beta(α, β) :

- $\frac{\alpha}{\alpha+\beta}$ est la moyenne
- $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ est la variance

Par **la méthode des moments** on trouve : $\widehat{\alpha}_M = 0,020$ et $\widehat{\beta}_M = 0,021$
puisque $\widehat{\theta} = \overline{y} = \frac{1}{n}\sum_{i=1}^n y_i = 0,49$ et $\widehat{Var}(\theta) = \frac{1}{n-1}\sum_{i=1}^n (y_i - \overline{y})^2 = 0,24$

$$\theta|\alpha,\beta \sim \text{Beta}(\widehat{\alpha};\widehat{\beta})$$

$$Y_i|\theta \overset{iid}{\sim} \text{Bernoulli}(\theta)$$

Bayes séquentiel

Le théorème de Bayes peut être utilisé séquentiellement :

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

Si $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$, alors :

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

Bayes séquentiel

Le théorème de Bayes peut être utilisé séquentiellement :

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

Si $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$, alors :

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

Bayes séquentiel

Le théorème de Bayes peut être utilisé séquentiellement :

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

Si $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$, alors :

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

⇒ mise à jour de la distribution *a posteriori* au fur et à mesure qu'arrive les observations (*online*)

Bayes séquentiel : application à l'exemple historique

Imaginons que l'on commence par observer 20 naissances $\boldsymbol{y}_{1:20}$ début 1745, dont 9 filles, et que l'on ait un *a priori* uniforme sur θ :

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