STAT 431 — Applied Bayesian Analysis — Course Notes

More About Conjugacy

Fall 2022

Perspective

Let the data model for $\boldsymbol{Y}$ with parameter $\boldsymbol{\theta}$ be a family $\mathcal{F}$ of distributions with densities

$$f(\boldsymbol{y} \mid \boldsymbol{\theta})$$

Consider a family $\mathcal{P}$ of distributions for $\boldsymbol{\theta}$ with densities

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{\phi})$$

where the index $\boldsymbol{\phi}$ is a **hyperparameter**. $\sim (\alpha, \beta)$

Then $\mathcal{P}$ is **conjugate** for $\mathcal{F}$ if:

Using the prior from $\mathcal{P}$ with density

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{\phi}_0)$$

yields a posterior from $\mathcal{P}$ with density

$$\pi\big(\boldsymbol{\theta} \mid \boldsymbol{\phi}_1(\boldsymbol{\phi}_0, \boldsymbol{y})\big)$$

regardless of the observed data values $\boldsymbol{y}$.

no matter what y happens to be

Note that the posterior's hyperparameter

$$\boldsymbol{\phi}_1 \;=\; \boldsymbol{\phi}_1(\boldsymbol{\phi}_0, \boldsymbol{y})$$

depends on $\boldsymbol{\phi}_0$ and $\boldsymbol{y}$.

Example: Normal Mean

Family $\mathcal{F}$ is defined by

$$Y_1, \ldots, Y_n \mid \mu \quad \sim \quad iid \ \mathrm{Normal}(\mu, 1/\tau^2)$$

for convenience

with τ^2 known, so that

$$\theta \quad \equiv \quad \mu \quad \text{only.}$$

Take family $\mathcal{P}$ to be

$$\mathrm{Normal}(\mu_*, 1/\tau_*^2) \qquad -\infty < \mu_* < \infty \qquad \tau_*^2 > 0$$

so that

$$\boldsymbol{\phi} \quad \equiv \quad (\mu_*, \tau_*^2)$$

Recall: Using the prior for μ from $\mathcal{P}$ that has ϕ equal to

$$\boldsymbol{\phi}_0 \quad = \quad (\mu_0, \tau_0^2)$$

led to a posterior from $\mathcal{P}$ that has ϕ equal to

$$\begin{aligned}\boldsymbol{\phi}_1 &= \boldsymbol{\phi}_1(\boldsymbol{\phi}_0, \boldsymbol{y}) \\ &= (\mu_1, \tau_1^2) \quad = \quad \left(\frac{n\tau^2\bar{y} + \tau_0^2\mu_0}{n\tau^2 + \tau_0^2}, n\tau^2 + \tau_0^2\right)\end{aligned}$$

· depends on y through $\bar{y}$

weighted avg

if $n = 0$

(μ_0, τ_0^2)

Posterior = prior

Regard $\boldsymbol{\phi}_1$ as having been "updated" from $\boldsymbol{\phi}_0$ by the data $\boldsymbol{y}$.

Natural Conjugacy

Let $Y_1, \ldots, Y_n$ be iid with common density

$$f(y \mid \boldsymbol{\theta})$$

Consider specifying values

$$\boldsymbol{\phi}_0 \;=\; \left(y_1^0, \ldots, y_m^0, m\right)$$

where m is a positive integer and $y_1^0, \ldots, y_m^0$ are values possible under f, such that

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{\phi}_0) \;\propto\; \prod_{j=1}^{m} f(y_j^0 \mid \boldsymbol{\theta})$$

integrable over θ

is a legitimate density in $\boldsymbol{\theta}$.

Then the class $\mathcal{P}$ of distributions defined by densities of that form is conjugate.

We call it the **natural conjugate** prior family.

See BSM, Section 2.1.5, for proof of conjugacy and application to

- binomial data (Bernoulli trials):

  $$\text{Beta is natural conjugate}$$

- Poisson counts:

  $$\text{Gamma is natural conjugate}$$

- mean-only normal samples:

  $$\text{Normal is natural conjugate}$$

Mixture Densities

II)

Suppose we have two different PDFs for $\boldsymbol{\theta}$:

$$\pi_1(\boldsymbol{\theta}) \qquad\qquad \pi_2(\boldsymbol{\theta})$$

Then, if $0 \leq q \leq 1$,

$$q\,\pi_1(\boldsymbol{\theta}) \;+\; (1-q)\,\pi_2(\boldsymbol{\theta})$$

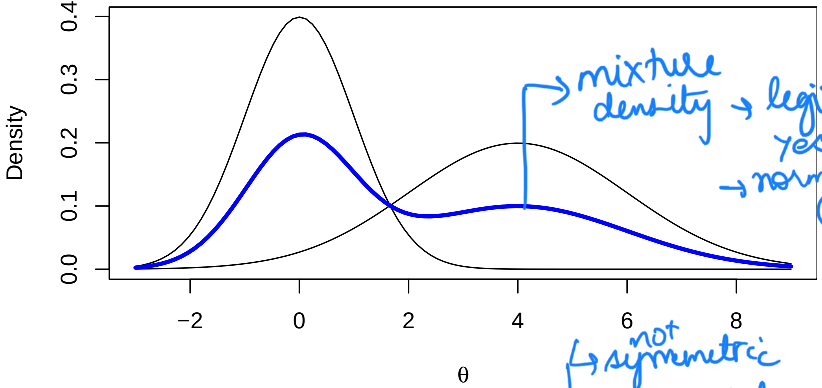
Convex combination or weighted density

is also a PDF for $\boldsymbol{\theta}$. (Similarly for two PMFs.)

We call it a **mixture** density, and may use it as a prior for $\boldsymbol{\theta}$.

For example, a half-and-half mixture of $\mathrm{Normal}(0, 1)$ and $\mathrm{Normal}(4, 4)$ has density

$$\pi(\theta) \;=\; \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\theta^2} \;+\; \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi \cdot 4}}\, e^{-\frac{1}{2\cdot 4}(\theta-4)^2}$$



→ mixture density → legitimate yes
→ normal PDF NO

↳ not symmetric
↳ bi-modal

More generally, a mixture density can have J components:

$$\pi(\boldsymbol{\theta}) \;=\; \sum_{j=1}^{J} q_j \, \pi_j(\boldsymbol{\theta})$$

where all $q_j \geq 0$ and $q_1 + \cdots + q_J = 1$.

$q_1, \ldots, q_J$ are the mixture **weights**.

Conjugate Mixtures

Suppose family $\mathcal{P}$ of distributions for $\boldsymbol{\theta}$ (all having densities) is conjugate for family $\mathcal{F}$ of data distributions for $\boldsymbol{Y}$ (all having densities).

expansion of P

Then the family $\widetilde{\mathcal{P}}$ of distributions with mixture densities

$$\pi(\boldsymbol{\theta}) \;=\; \sum_{j=1}^{J} q_j \, \pi_j(\boldsymbol{\theta})$$

$$q_j \geq 0, \qquad q_1 + \cdots + q_J \;=\; 1, \qquad \pi_j \text{ from } \mathcal{P}$$

is also conjugate for $\mathcal{F}$.

Reason:

mixture prior

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \;\propto\; f(\boldsymbol{y} \mid \boldsymbol{\theta})\,\pi(\boldsymbol{\theta}) \;=\; f(\boldsymbol{y} \mid \boldsymbol{\theta}) \sum_{j=1}^{J} q_j\,\pi_j(\boldsymbol{\theta})$$

Bayes' rule gives $\quad f(\boldsymbol{y} \mid \boldsymbol{\theta})\,\pi_j(\boldsymbol{\theta}) \;=\; m_j(\boldsymbol{y})\,p_j(\boldsymbol{\theta} \mid \boldsymbol{y}),\quad$ so

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \;\propto\; \sum_{j=1}^{J} q_j\,f(\boldsymbol{y} \mid \boldsymbol{\theta})\,\pi_j(\boldsymbol{\theta}) \;=\; \sum_{j=1}^{J} \underbrace{q_j\,m_j(\boldsymbol{y})}_{\text{sum=1 for prob density}}\,p_j(\boldsymbol{\theta} \mid \boldsymbol{y})$$

If $\quad Q_1, \ldots, Q_J \quad$ are $\quad q_1\,m_1(\boldsymbol{y}), \ldots, q_J\,m_J(\boldsymbol{y}) \quad$ after normalizing them to sum to 1, we get

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \;=\; \sum_{j=1}^{J} Q_j\,p_j(\boldsymbol{\theta} \mid \boldsymbol{y})$$

Since each p_j is from $\mathcal{P}$ (by conjugacy), the posterior is in $\widetilde{\mathcal{P}}$.

See BSM, Section 2.1.8 for an example.

Note: Computing the updated weights $Q_1, \ldots, Q_J$ requires evaluating the marginal density values $m_1(\boldsymbol{y}), \ldots, m_J(\boldsymbol{y})$.

Although evaluating marginal densities is usually hard, it is often possible analytically for "natural" conjugate families.