Thermodynamic integration

Typo in Eq. 21 last time corrected in _V3 on today's Moodle.

Avoiding Calculating Z (Hacks + Swindles)

- Savage - Dickey density ratio (Nested Models)

  calculate Bayes factor $B_{1,2} = \frac{P(\varepsilon = 0 \mid d, M_2)}{P(\varepsilon = 0 \mid M_2)}$

- Defining an Augmented Model

  suppose models M_1 & M_2, this allows us to use MCMC - methods to calculate $B_{1,2}$.

model 1 (M_1) has params x_1

prior $\pi(x_1 \mid M_1)$

likelihood $\mathcal{L}(d \mid x_1, M_1)$

M_2 : x_2

$\pi(x_2 \mid M_2)$

$\mathcal{L}(d \mid x_2, M_2)$

define augmented model M_* with params (ε, x_1, x_2)

<u>Def</u> prior $\pi(\varepsilon, x_1, x_2 \mid M_*)$

$$= \mathbb{1}_{(0,1)}(\varepsilon) \times \pi(x_1 \mid M_1) \times \pi(x_2 \mid M_2)$$

uniform

$0 < \varepsilon < 1$

<u>Def</u> likelihood

not continuous

$$\mathcal{L}(d \mid \varepsilon, x_1, x_2, M_*) = \begin{cases} \mathcal{L}(d \mid x_1, M_1) & \text{if } \varepsilon < \frac{1}{2} \\ \mathcal{L}(d \mid x_2, M_2) & \text{if } \varepsilon > \frac{1}{2} \end{cases}$$

use any MCMC method to sample from augmented Model posterior

$$P(\varepsilon, x_1, x_2 \mid d, M_*) = \frac{\mathcal{L}(d \mid \varepsilon, x_1, x_2, M_*)\, \pi(\varepsilon, x_1, x_2 \mid M_*)}{Z_{M_*}}$$

run a long MCMC chain (and discard burnin)

claim $$B_{1,2} = \frac{\#\text{ samples with } \varepsilon < \frac{1}{2}}{\#\text{ samples with } \varepsilon > \frac{1}{2}}$$

Consider $$\mathrm{Prob}(\varepsilon < \tfrac{1}{2} \mid M_*) = \int_0^{\frac{1}{2}} d\varepsilon \int dx_1 \int dx_2 \; P(\varepsilon, x_1, x_2 \mid d, M_*)$$

$$= \int_0^{\frac{1}{2}} d\varepsilon \int dx_1 \int dx_2 \frac{1}{Z_{M_*}} \mathcal{L}(\cancel{d} \mid x_1, M_1)$$

$$1 \times \pi(x_1 \mid M_1)\, \pi(x_2 \mid M_2)$$

using $\mathbb{I}_{(0,1)}(0 < \varepsilon < \frac{1}{2}) = 1$

$$= \frac{\int_0^{\frac{1}{2}} d\varepsilon \int dx_1\, \mathcal{L}(d \mid x_1, M_1)\, \pi(x_1 \mid M_1)}{Z_{M_*}}$$

$$P\left(\varepsilon < \tfrac{1}{2} \mid M_*\right) = \int_0^{\frac{1}{2}} d\varepsilon \; \frac{Z_{M_1}}{Z_{M_*}}$$

$$= \frac{Z_{M_1}}{2 Z_{M_*}}$$

similarly

$$P\left(\varepsilon > \tfrac{1}{2} \mid M_*\right) = \frac{Z_{M_2}}{2 Z_{M_*}}$$

$$\Rightarrow \underbrace{\frac{P\left(\varepsilon < \frac{1}{2} \mid M_*\right)}{P\left(\varepsilon > \frac{1}{2} \mid M_*\right)}} = \frac{Z_{M_1}}{Z_{M_2}} = B_{1,2}$$

This can be calculated from a single MCMC for M_* model.

drawbacks – scales badly with # models

– difficult to add new models

– if $B_{12} \gg 1$ then a short MCMC chain will only bound the Bayes factor.

— 3rd & final trick

Getting the Bayes from importance sampling

Suppose we have models $\mathcal{M}_1$ and $\mathcal{M}_2$. with identical parameters, and same priors.

$$\pi(x \mid \mathcal{M}_1) = \pi(x \mid \mathcal{M}_2) = \cancel{\pi(x)}$$

The evidence for $\mathcal{M}_2$.

$$\begin{aligned}
Z_{\mathcal{M}_2} &= P(d \mid \mathcal{M}_2) \\
&= \int dx \; \mathcal{L}(d \mid x, \mathcal{M}_2) \, \pi(x \mid \mathcal{M}_2) \\
&= \int dx \; \frac{\mathcal{L}(d \mid x, \mathcal{M}_2)}{\mathcal{L}(d \mid x, \mathcal{M}_1)} \; \mathcal{L}(d \mid x, \mathcal{M}_1) \, \pi(x \mid \mathcal{M}_2) \\
&= \int dx \; \frac{\mathcal{L}(d \mid x, \mathcal{M}_2)}{\mathcal{L}(d \mid x, \mathcal{M}_1)} \; \mathcal{L}(d \mid x, \mathcal{M}_1) \, \pi(x \mid \mathcal{M}_1)
\end{aligned}$$

using Bayes theorem for M_1 $\longrightarrow$

$$Z_{M_2} = Z_{M_1} \int dx \, \frac{\mathcal{L}(d|x, M_2)}{\mathcal{L}(d|x, M_1)} \, P(x|d, M_1)$$

$$\frac{Z_{M_2}}{Z_{M_1}} = \int dx \, P(x|d, M_1) \, \frac{\mathcal{L}(d|x, M_2)}{\mathcal{L}(d|x, M_1)}$$

$$= E_{x \sim P(x|d,M_1)}\left[\frac{\mathcal{L}(d|x, M_2)}{\mathcal{L}(d|x, M_1)} \right]$$

$$B_{2,1} \approx \frac{1}{N} \sum_{i=1}^{N} \frac{\mathcal{L}(d|x_i, M_2)}{\mathcal{L}(d|x_i, M_1)}$$

where $\{x_i\}$ is our MC in model M_1

Nested sampling

- evidence ✓
- even in high dimensions, multimodal/ non-Gaussian posteriors.
- also produces samples from posterior.

differs from stochastic sampling alg. in 2 ways :
- calculates Z.
- uses stochastical sampling internally.