Explaining approximate deterministic inference methods

HIV Inference Lab Group Meeting

Adam Howes

Imperial College London

December 2022

Bayes

- As a statistical modeller, our whole job (approximately) is to construct a generative model for data y using parameters ϑ
- This is the joint distribution $p(y, \vartheta) = p(y \mid \vartheta)p(\vartheta)$
- What we actually want is the posterior $p(\vartheta \mid y)$ which is **just**[1]

$$p(\vartheta \mid y) = \frac{p(y, \vartheta)}{p(y)} = \frac{p(y \mid \vartheta)p(\vartheta)}{p(y)}$$

- The **core problem** of Bayesian inference is that we can't compute $p(y)$

$$p(y) = \int p(y, \vartheta)\mathrm{d}\vartheta$$

[1] I've bolded this for sarcasm.

How might you do it?

- Usually if you wanted to integrate something, you could use numerical integration methods
- Pick nodes ϑ_i and weights ω_i then compute the sum

$$\tilde{p}(y) = \sum_i p(y, \vartheta_i)\omega_i$$

Monte Carlo as an example of numerical integration

- Suppose we can sample $\vartheta_i \sim p(y, \vartheta)$ for $i = 1, \ldots, N$
- If we set $\omega_i = 1/N$ for all i then

$$\tilde{p}(y) = \frac{1}{N} \sum_i p(y, \vartheta_i)$$

Monte Carlo is fundamentally unsound

- "Monte Carlo ignores information" according to O'Hagan (1987)
- Suppose $N = 3$ and we sample $\vartheta_1, \vartheta_2, \vartheta_3$ with $\vartheta_2 = \vartheta_3$ then our MC estimate is
$$\tilde{p}(y) = \frac{1}{3}\left(p(y, \vartheta_1) + p(y, \vartheta_2) + p(y, \vartheta_3)\right)$$
- This is despite the fact that nothing new about the function has been learned by adding $\{\vartheta_3, p(y, \vartheta_3)\}$

Application to HIV survey sampling

- Suppose we're running a household survey, and sample the same individual twice
- We didn't learn anything new about HIV by surveying them again!
- This doesn't just bite for nodes or individuals which are exactly the same: an analogous argument can be made if they are close together and we expect their function evaluations to be similar

$\implies$ Bayesian quadrature, Bayesian survey design (end of digression)

Latent variables and hyperparameters

- Quadrature doesn't work very well when $\dim(\vartheta)$ gets even moderately sized
- Previously I had all of the parameters under the symbol ϑ
- What if we split them up as being $\vartheta = (x, \theta)$
- The key part about this is that $\dim(x)$ is big and $\dim(\theta)$ is small

Names for x	Names for θ
Latent variables, random effects, latent field	Hyperparameters, fixed effects

Spatio-temporal statistics

- There is nothing inherently special about spatio-temporal statistics
- We have observations indexed by space $s \in \mathcal{S}$ and time $t \in \mathcal{T}$
- Usually we associate parameters to spatio-temporal locations as well as observations
- This ends up with us having something like $\{x_{s,t}\}$

What's important about this?

1. There might be **a lot** of spatio-temporal locations, so $\dim(x)$ might be pretty big!

- If you have 100 districts and 10 years, that's already $100 \times 10 = 1000$ parameters

2. Perhaps we're willing to make assumptions about how things vary over space-time

Latent Gaussian models

$$
\begin{aligned}
&\text{(Observations)} && y \sim p(y \mid x, \theta), \\
&\text{(Latent field)} && x \sim \mathcal{N}(x \mid \mu(\theta), Q(\theta)^{-1}), \\
&\text{(Hyperparameters)} && \theta \sim p(\theta),
\end{aligned}
$$

where $y = (y_1, \ldots, y_n)$, $x = (x_1, \ldots, x_n)$, $\theta = (\theta_1, \ldots, \theta_m)$

References I

O'Hagan, Anthony. 1987. "Monte Carlo Is Fundamentally Unsound." *The Statistician*, 247–49.