STA 360/602L: Module 3.7

MCMC and Gibbs sampling I

Dr. Olanrewaju Michael Akande

Bayesian inference (conjugacy recap)

- As we've seen so far, Bayesian inference is based on posterior distributions, that is,

$$\pi(\theta|y) = \frac{\pi(\theta) \cdot p(y|\theta)}{\int_{\Theta} \pi(\tilde{\theta}) \cdot p(y|\tilde{\theta}) \mathrm{d}\tilde{\theta}} = \frac{\pi(\theta) \cdot L(\theta|y)}{L(y)},$$

  where $y = (y_1, \ldots, y_n)$.

- Good news: we have the numerator in this expression.

- Bad news: the denominator is typically not available (may involve high dimensional integral)!

- How have we been getting by? Conjugacy! For conjugate priors, the posterior distribution of θ is available analytically.

- What if a conjugate prior does not represent our prior information well, or we have a more complex model, and our posterior is no longer in a convenient distributional form?

Some conjugate models

- For example, the most common conjugate models are

Prior	Likelihood	Posterior
beta	binomial	beta
gamma	Poisson	gamma
gamma	exponential	gamma
normal-gamma	normal	normal-gamma
beta	negative-binomial	beta
beta	geometric	beta

- There are a few more we have not covered yet, for example, the Dirichlet-multinomial model.

- However, clearly, we cannot restrict ourselves to conjugate models only.

Back to the normal model

- For example, for conjugacy in the normal model, we had

$$\pi(\mu|\tau) = \mathcal{N}\left(\mu_0, \frac{1}{\kappa_0 \tau}\right).$$
$$\pi(\tau) = \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2\tau_0}\right)$$

- Suppose we instead wish to specify our uncertainty about μ as independent of τ, that is, we want $\pi(\mu, \tau) = \pi(\mu)\pi(\tau)$. For example,

$$\pi(\mu) = \mathcal{N}\left(\mu_0, \sigma_0^2\right).$$
$$\pi(\tau) = \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2\tau_0}\right).$$

- When σ_0^2 is not proportional to $\frac{1}{\tau}$, the marginal density of τ is not a gamma density (or a density we can easily sample from).

- Side note: for conjugacy, the joint posterior should also be a product of two independent Normal and Gamma densities in μ and τ respectively.

Non-conjugate priors

- In general, conjugate priors are not available for generalized linear models (GLMs) other than the normal linear model.
- One can potentially rely on an asymptotic normal approximation.
- As $n \to \infty$, the posterior distribution is normal centered on MLE.
- However, even for moderate sample sizes, asymptotic approximations may be inaccurate.
- In logistic regression for example, for rare outcomes or rare binary exposures, posterior can be highly skewed.
- It is appealing to avoid any reliance on large sample assumptions and base inferences on **exact posterior**.

Non-conjugate priors

- Even though we may not be able to sample from the marginal posterior of a particular parameter when using a non-conjugate prior, sometimes, we may still be able to sample from conditional distributions of those parameters given all other parameters and the data.

- These conditional distributions, known as full conditionals, will be very important for us.

- In our normal example with

$$\begin{aligned} \mu &\sim \mathcal{N}\left(\mu_0, \sigma_0^2\right). \\ \tau &\sim \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2\tau_0}\right), \end{aligned}$$

  turns out we will not be able sample easily from $\tau|Y$,

- However, as you will see, we will be able to sample from $\tau|\mu, Y$. That is the full conditional for τ.

- By the way, note that we already know the full conditional for μ, i.e., $\mu|\tau, Y$ from previous modules.

Full conditional distributions

- Goal: try to take advantage of those full conditional distributions (without sampling directly from the marginal posteriors) to obtain samples from the said marginal posteriors.
- In our example, with $\pi(\mu) = \mathcal{N}\left(\mu_0, \sigma_0^2\right)$, we have

$$\mu | Y, \tau \sim \mathcal{N}(\mu_n, \tau_n^{-1}),$$

  where

  - $\mu_n = \dfrac{\frac{\mu_0}{\sigma_0^2} + n\tau\bar{y}}{\frac{1}{\sigma_0^2} + n\tau}$; and
  - $\tau_n = \frac{1}{\sigma_0^2} + n\tau.$
- Review results from previous modules on the normal model if you are not sure why this holds.
- Let's see if we can figure out the other full conditional $\tau | \mu, Y$.

Full conditional distributions

$$
\begin{aligned}
p(\tau|\mu, Y) &= \frac{\Pr[\tau, \mu, Y]}{\Pr[\mu, Y]} = \frac{p(y|\mu,\tau)\pi(\mu,\tau)}{p[\mu, y]} \\
&= \frac{p(y|\mu,\tau)\pi(\mu)\pi(\tau)}{p[\mu, y]} \\
&\propto p(y|\mu,\tau)\pi(\tau) \\
&\propto \underbrace{\tau^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\tau\sum_{i=1}^{n}(y_i-\mu)^2\right\}}_{\propto\, p(y|\mu,\tau)} \times \underbrace{\tau^{\frac{\nu_0}{2}-1} \exp\left\{-\frac{\tau\nu_0}{2\tau_0}\right\}}_{\propto\, \pi(\tau)} \\
&= \underbrace{\tau^{\frac{\nu_0+n}{2}-1} \exp\left\{-\frac{1}{2}\tau\left[\frac{\nu_0}{\tau_0} + \sum_{i=1}^{n}(y_i-\mu)^2\right]\right\}}_{\text{Gamma Kernel}}.
\end{aligned}
$$

Full conditional distributions

$$p(\tau|\mu, Y) \propto \underbrace{\tau^{\frac{\nu_0+n}{2}-1} \exp\left\{-\frac{1}{2}\tau\left[\frac{\nu_0}{\tau_0} + \sum_{i=1}^{n}(y_i-\mu)^2\right]\right\}}_{\text{Gamma Kernel}}$$
$$= \text{Gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n}{2\tau_n(\mu)}\right) \quad \text{OR} \quad \text{Gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n\sigma_n^2(\mu)}{2}\right),$$

where

$$\nu_n = \nu_0 + n$$
$$\sigma_n^2(\mu) = \frac{1}{\nu_n}\left[\frac{\nu_0}{\tau_0} + \sum_{i=1}^{n}(y_i-\mu)^2\right] = \frac{1}{\nu_n}\left[\frac{\nu_0}{\tau_0} + ns_n^2(\mu)\right]$$
$$\text{OR} \quad \tau_n(\mu) = \frac{\nu_n}{\left[\frac{\nu_0}{\tau_0} + \sum_{i=1}^{n}(y_i-\mu)^2\right]} = \frac{\nu_n}{\left[\frac{\nu_0}{\tau_0} + ns_n^2(\mu)\right]};$$
$$\text{with} \quad s_n^2(\mu) = \frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2.$$

Iterative scheme

- Now we have two full conditional distributions but what we really need is to sample from $\pi(\tau|Y)$.
- Actually, if we could sample from $\pi(\mu, \tau|Y)$, we already know that the draws for μ and τ will be from the two marginal posterior distributions. So, we just need a scheme to sample from $\pi(\mu, \tau|Y)$.
- Suppose we had a single sample, say $\tau^{(1)}$ from the marginal posterior distribution $\pi(\tau|Y)$. Then we could sample

$$\mu^{(1)} \sim p(\mu|\tau^{(1)}, Y).$$

- This is what we did in the last class, so that the pair $\{\mu^{(1)}, \tau^{(1)}\}$ is a sample from the joint posterior $\pi(\mu, \tau|Y)$.
- $\Rightarrow$ $\mu^{(1)}$ can be considered a sample from the marginal distribution of μ, which again means we can use it to sample

$$\tau^{(2)} \sim p(\tau|\mu^{(1)}, Y),$$

and so forth.

Gibbs sampling

- So, we can use two **full conditional distributions** to generate samples from the **joint distribution**, once we have a starting value $\tau^{(1)}$.
- Formally, this sampling scheme is known as Gibbs sampling.
  - Purpose: Draw from a joint distribution, say $p(\mu, \tau | Y)$.
  - Method: Iterative conditional sampling
    - Draw $\tau^{(1)} \sim p(\tau | \mu^{(0)}, Y)$
    - Draw $\mu^{(1)} \sim p(\mu | \tau^{(1)}, Y)$
  - Purpose: Full conditional distributions have known forms, with sampling from the full conditional distributions fairly easy.
- More generally, we can use this method to generate samples of $\theta = (\theta_1, \ldots, \theta_p)$, the vector of p parameters of interest, from the joint density.

Gibbs sampling

- Procedure:
  - Start with initial value $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_p^{(0)})$.
  - For iterations $s = 1, \ldots, S$,
    1. Sample $\theta_1^{(s)}$ from the conditional posterior distribution

       $$\pi(\theta_1 | \theta_2 = \theta_2^{(s-1)}, \ldots, \theta_p = \theta_p^{(s-1)}, Y)$$

    2. Sample $\theta_2^{(s)}$ from the conditional posterior distribution

       $$\pi(\theta_2 | \theta_1 = \theta_1^{(s)}, \theta_3 = \theta_3^{(s-1)}, \ldots, \theta_p = \theta_p^{(s-1)}, Y)$$

    3. Similarly, sample $\theta_3^{(s)}, \ldots, \theta_p^{(s)}$ from the conditional posterior distributions given current values of other parameters.
  - This generates a **dependent** sequence of parameter values.
- In the next module, we will look into a simple example of how this works, before going back to the normal model.

What's next?

Move on to the readings for the next module!