STAT 578: Advanced Bayesian Modeling

Week 7 – Lesson 1

Sampling with Markov Chains

Fall 2019

Markov Chains

Markov Chains

A sequence of random variables

$$\theta^0, \quad \theta^1, \quad \theta^2, \quad \ldots$$

is a **Markov chain** if, for each $t > 1$,

θ^t is conditionally independent of $\quad \theta^0, \quad \ldots, \quad \theta^{t-2}, \quad$ given θ^{t-1}.

That is,

$$p(\theta^t \mid \theta^{t-1}, \ldots, \theta^0) \quad = \quad p(\theta^t \mid \theta^{t-1}).$$

The value of θ^t is the **state** of the Markov chain at time t.

We will also call θ^t the *draw* at *iteration* t.

Usually, the draws are dependent (despite the conditional independence).

The Markov chains we consider will be *discrete-time*, with only integer values for t, but will usually have a *continuous state space*.

Transition Distributions

The **transition distribution** is the conditional distribution of θ^t given θ^{t-1}.

Its density is the **transition kernel**. For iteration t:

$$T_t(\theta^t \mid \theta^{t-1})$$

Note that it can depend on t.

If the transition distribution does not depend on t, the Markov chain is **time-invariant** (or **time-homogeneous**).

Stationary Distribution

A distribution on states of a time-invariant Markov chain is **stationary** if all subsequent draws have that distribution whenever the first draw does.

That is, if θ^0 has the stationary distribution, then so do θ^1, θ^2, ...

Under certain technical conditions on the transition distribution, there is a unique stationary distribution.

Convergence

A time-invariant Markov chain may converge to a **limiting distribution** $\mathcal{D}$:

$$\text{As } t \to \infty, \qquad \theta^t \longrightarrow_d \mathcal{D}$$

Under certain technical conditions, the limiting distribution is the (unique) stationary distribution.

Key strategy behind Bayesian MCMC:

Simulate sequentially from a Markov chain that has the posterior as its stationary (and hopefully limiting) distribution.

Concerns:

- How do we construct a Markov chain with a given stationary distribution?
- There will be **transient** effects from not starting in the stationary distribution. How do we handle these?
- How do we deal with dependence?