Markov Chain Monte Carlo

22S:138, Bayesian Statistics

Lecture 10
Sept. 25, 2006

Kate Cowles
374 SH, 335-0727

The Poisson distribution (one more one-parameter distribution)

- The Poisson distribution may be appropriate when the data are counts of rare events.
- events occurring at random at a constant rate per unit time, distance, volume, or whatever
- assumption that the number of events that occur in any interval is independent of the number of events occurring in a disjoint interval

- examples:
  - the number of cases of a rare form of cancer occurring in Johnson County in each calendar year
  - the number of flaws occurring in each 100-foot length of yarn produced by a spinning machine
  - the number of particles of pollen per cubic foot of air in this room
- Since the values of a random variable following a Poisson distribution are *counts*, what are the possible values?
- probability mass function for a Poisson random variable

$$p(y|\lambda) = \frac{e^{-\lambda}\,\lambda^y}{y!}, \quad y = 0, 1, \ldots$$

- the count of the number of events occurring in m time units also follows a Poisson distribution, but with parameter $m\lambda$

- The conjugate prior distribution for the Poisson rate parameter is the gamma family.

Markov Chain Monte Carlo Methods

- Goals
  - to make inference about model parameters
  - to make predictions
- Requires
  - integration over possibly high-dimensional integrand
  - and we may not know the integrating constant

Monte Carlo integration and MCMC

- Monte Carlo integration
  - draw independent samples from required distribution
  - use sample averages to approximate expectations
- Markov chain Monte Carlo (MCMC)
  - draws samples by running a Markov chain that is constructed so that its limiting distribution is the joint distribution of interest

Markov chains

- A Markov chain is a sequence of random variables $X_0, X_1, X_2, \ldots$
- At each time $t \geq 0$ the next state X_{t+1} is sampled from a distribution

  $$P(X_{t+1}|X_t)$$

  that depends only on the state at time t
  - called "transition kernel"
- Under certain regularity conditions, the iterates from a Markov chain will gradually converge to draws from a unique *stationary* or *invariant* distribution
  - i.e. chain will "forget' its initial state
  - as t increases, sampled points X_t will look increasingly like (correlated) samples from the stationary distribution

- Suppose:
  - MC is run for N (large number) iterations
  - we throw away output from first m iterations
  - regularity conditions are met
- then by *ergodic theorem*
  - we can use averages of remaining samples to estimate means

$$E[f(X)] \simeq \frac{1}{N-m} \sum_{t=m+1}^{N} f(X_t)$$

Gibbs sampling: one way to construct the transition kernel

- seminal references
  - Geman and Geman (*IEEE Trans. Pattn. Anal. Mach. Intel.*, 1984)
  - Gelfand and Smith (*JASA*, 1990)
  - Hastings (*Biometrika*, 1970)
  - Metropolis, Rosenbluth, et al. (*J. Chem. Phys*, 1953)
- subject to regularity conditions, joint distribution is uniquely determined by "full conditional distributions"
  - full conditional distribution for a model quantity is distribution of that quantity conditional on assumed known values of all the other quantities in the model
- break complicated, high-dimensional problem into a large number of simpler, low-dimensional problems

Example: Inference about normal mean and variance, both unknown

- model

$$y_i|\mu, \sigma^2 \sim N(\mu, \sigma^2)$$
$$i = 1, \ldots, N$$

- priors

$$\mu \sim N(\mu_0, \sigma_0^2)$$
$$\sigma^2 \sim IG(a_1, b_1)$$

- We want posterior means, posterior medians, posterior credible sets for μ, σ^2

Full Conditional Distributions for Normal Model

- to extract mathematical form of full conditional for a parameter:
  - write out expression to which joint posterior is proportional
  - pull out all terms containing the parameter of interest