March 9
Friday March 9, 2018 09:00

MCMC (Markov Chain Monte Carlo)

Inspiration:

Tiger's tale: Bayes modeling (partial membership model).

Tiger's teacher: choice of prior distribution

① Bayes modeling

θ is unknown parameter — random variable.

$\theta \sim \pi(\theta)$ prior distribution.

X — random sample

Bayes' Theorem: $\pi(\theta|x) = \frac{p(x|\theta)\cdot\pi(\theta)}{p(x)} = \frac{p(x|\theta)\cdot\pi(\theta)}{\int p(x|\theta)\pi(\theta)\,d\theta}$

posterior distribution!

$\pi(\theta|x) \propto p(x|\theta)\cdot\pi(\theta)$ proportional relationship

$\int p(x|\theta)\pi(\theta)d\theta$ ← normalization constant.

Ideally: obtain samples from $\pi(\theta|x)$, without knowing constant, just by $p(x|\theta)\cdot\pi(\theta)$ ← what exactly MCMC does.

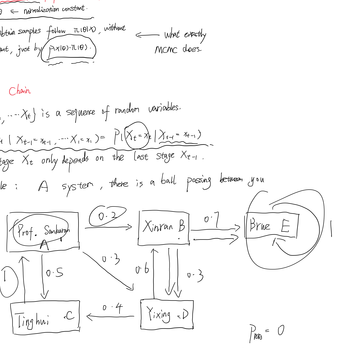
② Markov Chain

Def: $\{X_1, X_2, \cdots X_n\}$ is a sequence of random variables.

$P(X_t = x_t \mid X_{t-1} = x_{t-1}, \cdots X_1 = x_1) = P(X_t = x_t \mid X_{t-1} = x_{t-1})$

current stage X_t only depends on the last stage X_{t-1}.

Example: A system, there is a ball passing between y ∞



$P_{AA} = 0$

5 stages A, B, C, D, E

$P(X_t = B \mid X_{t-1} = A) = 0.2$

$P(X_t = C \mid X_{t-1} = A) = 0.3$

$P(X_t = A \mid X_{t-1} = C, X_{t-2} = D), \cdots)$

$= P(X_t = A \mid X_{t-1} = C) \cdots)$

$=$

generalize transition matrix

$$Q = \begin{pmatrix} P_{AA} & P_{AB} & P_{AC} & P_{AD} & P_{AE} \\ P_{BA} & P_{BB} & \cdots & & \\ \vdots & & & & \\ & & & & P_{EE} \end{pmatrix}_{5\times5} = \begin{pmatrix} 0 & 0.2 & 0.3 & 0.3 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & 1 \end{pmatrix}$$

P_0 initial probability $= (0.1, 0.2, 0.3, 0.2, 0.2)$ at time $= 0$

$P_1 = P_0 \cdot Q = (0.2, 0.3, 0.1, 0.3)$, at time $= 1$

$P_2 = P_1 \cdot Q = P_0 \cdot Q^2$

$\cdots\cdots$

$P_n = P_0 \cdot Q^n$ at time n

when $n \uparrow$, $P_{n+1} = P_n \cdot Q = P_n$ stable

$\pi = \pi \cdot Q$. $(0.2, 0.1, \cdots)$ time $\to \infty$.

stationary distribution → posterior $\pi(\theta|X)$ intuition.

no matter what is your initial stage P_0. (when $n \to \infty$)

Conditions? Doesn't matter right now

Notation: $p(x) = \frac{p'(x)}{Z} = \frac{p(x|\theta)\cdot\pi(\theta)}{\int p(x|\theta)\cdot\pi(\theta)d\theta}$ ← numerator, denominator

Algorithm

Let $X_0 = x$ be current state, repeat following steps.

1. Generate $y \sim Q(\cdot|x)$. Q is a Markov transition matrix. Let y be the generated value.
2. Set $X_{t+1} = y$ with probability $\alpha(x,y) = \min\left\{\frac{p(y)}{p(x)}\cdot\frac{Q(x|y)}{Q(y|x)}, 1\right\}$ (cancel the constant.) $= \min\left\{\frac{p'(y)}{p'(x)}\cdot\frac{Q(x|y)}{Q(y|x)}, 1\right\}$ otherwise $X_{t+1} = x$.
3. generate $u \sim |U(0,1)|$ if $u < \alpha$ accept else reject. $X_{t+1} = y$, $X_{t+1} = x$

$P(u<\alpha) = \alpha$

Q

application

① simulate known distribution directly.

② simulate distribution with $p(x)$ known.

only using the numerator in Bayes theorem.

$\theta \sim U(0,1)$

$B(n, \theta)$ $\theta \in (0,1)$.

$\theta' \sim U(0,1)$.

Question:

① How to choose transition matrix Q

② When will the markov chain converge.

Case 1: Normal $(3, 9)$ is our target distribution.

Q: $(x_t, 1) \to y$.

$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\cdot|x-\mu|^2\right)$

here $\sigma^2 = 9$, $\mu = 3$

Case 2:

prior ~ Beta(1, 1)

observed data ~ Binomial (10, θ).

say we observe two success.

theoretically, posterior will be Beta$(1+2, 1+10-2)$ = Beta$(3, 9)$.

$X \sim$ Binomial $(10, \theta)$. $\pi(\theta) \sim$ Beta$(1,1)$.

$\pi(\theta|x) = \frac{p(x|\theta)\cdot\pi(\theta)}{p(x)} = \frac{p(x|\theta)\cdot\pi(\theta)}{\int_0^1 p(x|\theta)\cdot\pi(\theta)\,d\theta}$ ← $p'(\theta_x)$

$= \frac{C_n^x\cdot\theta^x\cdot(1-\theta)^{n-x}\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\cdot\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int\cdots d\theta}$ actually $\to \theta^{x+\alpha-1}\cdot(1-\theta)^{\beta+n-x-1}$ ← what if unknown

← Beta$(x+\alpha, \beta+n-x)$

$\frac{p(y)\cdot Q(x|y)}{p'(x)\cdot Q(y|x)}$, here $y = \theta_{t+1}$, $x = \theta_t$

$Q(\cdot|\cdot) = 1$

So $= \frac{C_n^x\cdot(\theta_{t+1})^x\cdot(1-\theta_{t+1})^{n-x}}{C_n^x\cdot(\theta_t)^x\cdot(1-\theta_t)^{n-x}}\cdot\frac{Q(x|y)}{Q(y|x)}$

Then the resulting Markov chain is reversible and stationary distribution $p(x) \propto p'(x)/Z_p$

We can therefore sample from $p(x)$ by running the algorithm until stationary is achieved.

We don't need to know Z_p here to sample from $p(x)$.

Questions:

① When stationary is determined. ✓ related to convergence

② How to choose transition matrix Q

Conjugate prior for $p(x|\theta)$

if $p(\theta|x) \propto \pi(\theta)\cdot p(x|\theta)$

the transition follows the posterior, with a parameter α

the type of distribution is not changed

E.g.: beta distribution is conjugate prior for binomial distribution

Target: simulate the posterior without knowing the concept of proportionality.

$\pi(\theta|X) \propto \pi(\theta)\cdot p(x|\theta)$

$\alpha(x,y) = \min\left\{\frac{p(y)}{p(x)}\cdot\frac{Q(x|y)}{Q(y|x)}, 1\right\} = \frac{C_n^x\cdot\theta^x\cdot(1-\theta)^{n-x}}{C_n^x\cdot\theta^x\cdot(1-\theta)^{n-x}}\cdot\frac{C_n^y\cdot\theta^y\cdot(1-\theta)^{n-y}}{C_n^y\cdot\theta^y\cdot(1-\theta)^{n-y}}$

$= \frac{C_n^x}{C_n^x}\cdot\theta^{x-x}\cdot(1-\theta)^{n-x-n+x}$