Bayes AI

Unit 6: Markov Chain Monte Carlo

Vadim Sokolov George Mason University Spring 2025

MCMC Simulation

Suppose that $X \sim F_X(x)$ and let $Y = g(X)$.

How do we find $F_Y(y)$ and $f_Y(y)$?

- von Neumann

Given a uniform U, how do we find $X = g(U)$?

- In the bivariate case $(X, Y) \to (U, V)$.

We need to find $f_{(U,V)}(u, v)$ from $f_{X,Y}(x, y)$

- Applications: Simulation, MCMC and PF.

Transformations

The cdf identity gives

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y)$$

- Hence if the function $g(\cdot)$ is monotone we can invert to get

$$F_Y(y) = \int_{g(x) \le y} f_X(x) dx$$

- If g is increasing $F_Y(y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y))$

If g is decreasing $F_Y(y) = P(X \ge g^{-1}(y)) = 1 - F_X(g^{-1}(y))$

Transformation Identity

1. Theorem 1: Let X have pdf $f_X(x)$ and let $Y = g(X)$. Then if g is a monotone function we have

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

There's also a multivariate version of this that we'll see later.

- Suppose X is a continuous rv, what's the pdf for $Y = X^2$?
- Let $X \sim N(0, 1)$, what's the pdf for $Y = X^2$?

Probability Integral Transform

theorem Suppose that $U \sim U[0,1]$, then for any continuous distribution function F, the random variable $X = F^{-1}(U)$ has distribution function F.

- Remember that for $u \in [0,1]$, $\mathbb{P}\left(U \leq u\right) = u$, so we have

$$\mathbb{P}\left(X \leq x\right) = \mathbb{P}\left(F^{-1}(U) \leq x\right) = \mathbb{P}\left(U \leq F(x)\right) = F(x)$$

Hence, $X = F_X^{-1}(U)$.

Normal

Sometimes thare are short-cut formulas to generate random draws

Normal $N(0, I_2)$: x_1, x_2 uniform on $[0, 1]$ then

$$y_1 = \sqrt{-2 \log x_1} \cos(2\pi x_2)$$
$$y_2 = \sqrt{-2 \log x_1} \sin(2\pi x_2)$$

Simulation and Transformations

An important application is how to transform multiple random variables?

- Suppose that we have random variables:

$$(X, Y) \sim f_{X,Y}(x, y)$$

A transformation of interest given by:

$$U = g(X, Y) \ \text{and} \ V = h(X, Y)$$

- The problem is how to compute $f_{U,V}(u, v)$? Jacobian

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

Bivariate Change of Variable

- *Theorem:* (change of variable)

$$f_{U,V}(u,v) = f_{X,Y}(h_1(u,v), h_2(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right|$$

The last term is the Jacobian.

This can be calculated in two ways.

$$\left| \frac{\partial(x,y)}{\partial(u,v)} \right| = 1 / \left| \frac{\partial(u,v)}{\partial(x,y)} \right|$$

- So we don't always need the inverse transformation $(x,y) = (g^{-1}(u,v), h^{-1}(u,v))$

Inequalities and Identities

1. Markov

$$\mathbb{P}\left(g(X) \geq c\right) \leq \frac{\mathbb{E}(g(X))}{c} \;\; \text{where} \;\; g(X) \geq 0$$

2. Chebyshev

$$\mathbb{P}\left(|X - \mu| \geq c\right) \leq \frac{Var(X)}{c^2}$$

3. Jensen

$$\mathbb{E}\left(\phi(X)\right) \leq \phi\left(\mathbb{E}(X)\right)$$

4. Cauchy-Schwarz

$$corr(X, Y) \leq 1$$

Chebyshev follows from Markov. Mike Steele and Cauchy-Schwarz.

Markov Inequality

Let f be non-decreasing

$$\begin{aligned}
P(Z > t) &= P(f(Z) \geq f(t)) \\
&= E\left(\mathbb{I}(f(Z) \geq f(t))\right) \\
&\leq E\left(\mathbb{I}(f(Z) \geq f(t))\frac{f(Z)}{f(t)}\right) \\
&= E\left(\frac{f(Z)}{f(t)}\right)
\end{aligned}$$

Concentration Inequalities

Law of Large Numbers

$$\lim_{n\to\infty} \mathbb{P}\left(|Z - E(Z)| > n\epsilon\right) = 0 \;\; \forall \epsilon > 0$$

Central Limt Theorem (CLT)

$$\lim_{n\to\infty} \mathbb{P}\left(n^{-1/2}(|Z - E(Z)|) > \epsilon\right) = \Phi(x)$$

Posterior Concentration

Hoeffding and Bernstein

Let $Z = \sum_{i=1}^{n} X_i$.

Hoeffding

$$P(Z > E(Z) + t) \leq \exp\left(-\frac{t^2}{2n}\right)$$

Bernstein

$$P(Z > E(Z) + t) \leq \exp\left(-\frac{t^2}{2(Var(Z) + t/3)}\right)$$

Large Deviations (Varadhan)

Special Distributions

See *Common Distributions*

1. Bernoulli and Binomial
2. Hypergeometric
3. Poisson
4. Negative Binomial
5. Normal Distribution
6. Gamma Distribution
7. Beta Distribution
8. Multinomial Distribution
9. Bivariate Normal Distribution
10. Wishart Distribution

...

Example: Markov Dependence

- We can always factor a joint distribution as

$$p(X_n, X_{n-1}, \ldots, X_1) = p(X_n | X_{n-1}, \ldots, X_1) \ldots p(X_2 | X_1) p(X_1)$$

example - A process has the *Markov Property* if

$$p(X_n | X_{n-1}, \ldots, X_1) = p(X_n | X_{n-1})$$

- Only the current history matter when determining the probabilities.

:

A real world probability model: Hidden Markov Models

Are stock returns a random walk?

Hidden Markov Models (Baum-Welch, Viterbi)

- Daily returns on the SP500 stock market index.

Build a hidden Markov model to predict the ups and downs.

- Suppose that stock market returns on the next four days are $X_1, \ldots, X_4$.
- Let's empirical determine conditionals and marginals

SP500 Data

Marginal and Bivariate Distributions

- Empirically, what do we get? Daily returns from 1948 – 2007.

center x Down Up ———————- ———- ———- $P(X_i) = x$ 0.474 0.526

- Finding $p(X_2|X_1)$ is twice as much computational effort: counting UU, UD, DU, DD transitions.

center X_i Down Up ————————— ———- ———- $X_{i-1} = Down$ 0.519 0.481 $X_{i-1} = Up$ 0.433 0.567

Conditioned on two days

- Let's do $p(X_3|X_2, X_1)$

center X_{i-2} X_{i-1} Down Up ——— ——— ——- ——- Down Down 0.501 0.499 Down Up 0.412 0.588 Up Down 0.539 0.461 Up Up 0.449 0.551

- We could do the distribution $p(X_2, X_3|X_1)$. This is a joint, marginal and conditional distribution all at the same time.

Joint because more than one variable (X_2, X_3), *marginal* because it ignores X_4 and *conditional* because its given X_1.

Joint Probabilities

- Under Markov dependence

$$
\begin{aligned}
P(UUD) &= p(X_1 = U)p(X_2 = U|X_1 = U)p(X_3|X_2 = U, X_1 = U) \\
&= (0.526)(0.567)(0.433)
\end{aligned}
$$

- Under independence we would have got

$$
\begin{aligned}
P(UUD) &= P(X_1 = U)p(X_2 = U)p(X_3 = D) \\
&= (.526)(.526)(.474) \\
&= 0.131
\end{aligned}
$$