

统计计算 张楠 2019秋: The Gibbs Sampler

(返回 [统计计算 张楠 2019秋](#))

Gibbs 抽样是MH算法的一个特例, 其经常用于目标分布是多元分布的场合. 假设所有的一元条件分布 (每个分量对其他分量的条件分布)都是可以确定的, Gibbs抽样使用这些一元条件分布进行抽样.

令 $X = (X_1, \dots, X_d)$ 为 R^d 中的随机变量, 定义 $d-1$ 维的随机变量 $X_{-j} = (X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_d)$, 并记 $X_j|X_{-j}$ 的条件密度为 $f(X_j|X_{-j})$. 则Gibbs抽样是从这 d 个条件分布中产生候选点. 算法如下:

1. 在 $t = 0$ 时, 初始化 $X(0)$;

2. 对 $t = 1, 2, \dots, T$,

(a) 令 $x_1 = X_1(t-1)$.

(b) 对每个分量 $j = 1, \dots, d$,

(i) 从 $f(X_j|x_{-j})$ 中产生候选点 $X_j^*(t)$.

(ii) 更新 $x_j = X_j^*(t)$.

(c) 令 $X(t) = (X_1^*(t), \dots, X_d^*(t))$ (每个候选点都被接受)

(d) 增加 t 注意在上述算法(b)步抽样中, 各个分量依次被更新:

$$\begin{aligned} x_1(t) &\sim f(x_1|x_2(t-1), \dots, x_d(t-1)); \\ x_2(t) &\sim f(x_2|x_1(t), x_3(t-1), \dots, x_d(t-1)) \\ &\vdots \\ x_d(t) &\sim f(x_d|x_1(t), \dots, x_{d-1}(t)). \end{aligned}$$

从一

元分 布 $f(x_j|x_1(t), x_2(t), \dots, x_{j-1}(t), x_{j+1}(t-1), \dots, x_d(t-1))$ 中抽样是比较容易的, 因为 $f(x_j|x_{-j}) \propto f(x)$, 其中除了变量 x_j 外, 其他变量都是常数.

例 1 (Gibbs 抽样: 二元分布) 使用Gibbs抽样产生二元正态分布 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 的随机数

在二元正态场合, $X_1|X_2$ 以及 $X_2|X_1$ 仍然服从正态分布, 且易知

$$E[X_1|X_2 = x_2] = \mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2),$$

类似可得 $X_2|X_1$ 的分布. 因此

$$Var[X_1|X_2 = x_2] = (1 - \rho^2)\sigma_1^2$$

$$f(x_1|x_2) \sim N(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), (1 - \rho^2)\sigma_1^2)$$

$$f(x_2|x_1) \sim N(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), (1 - \rho^2)\sigma_2^2)$$

因此, 使用Gibbs算法如下

1. 令 $(x_1, x_2) = X(t-1)$;

2. 从 $f(x_1|x_2)$ 中产生候选点 $X_1^*(t)$.

3. 更新 $x_1 = X_1^*(t)$.

4. 从 $f(x_2|x_1)$ 中产生 $X_2^*(t)$.

5. 令 $X(t) = (X_1^*(t), X_2^*(t))$.

R 代码如下

```
#initialize constants and parameters
N <- 5000           #length of chain
burn<- 1000        #burn-in length
X <- matrix(0, N, 2) #the chain, a bivariate sample

rho <- -.75         #correlation
mu1 <- 0
mu2 <- 2
sigma1 <- 1
sigma2 <- .5
s1 <- sqrt(1-rho^2)*sigma1
s2 <- sqrt(1-rho^2)*sigma2

##### generate the chain #####

X[1, ] <- c(mu1, mu2) #initialize

for (i in 2:N) {
  x2 <- X[i-1, 2]
  m1 <- mu1 + rho * (x2 - mu2) * sigma1/sigma2
  X[i, 1] <- rnorm(1, m1, s1)
  x1 <- X[i, 1]
  m2 <- mu2 + rho * (x1 - mu1) * sigma2/sigma1
  X[i, 2] <- rnorm(1, m2, s2)
}

b <- burn + 1
x <- X[b:N, ]
```

产生的链开始的1000个观测被丢弃掉, 剩下的观测存在x中, 对此样本 计算均值和协方差矩阵如下. 各参数的样本估计离真值很近, 散点图也显示出 二元正态所具有的球面对称性和负相关性特征.

```
# compare sample statistics to parameters
colMeans(x)
cov(x)
cor(x)

plot(x, main="", cex=.5, xlab=bquote(X[1]),
      ylab=bquote(X[2]), ylim=range(x[, 2]))
```

取自 ["http://shjcx.wang/index.php?title=统计计算_张楠_2019秋:_The_Gibbs_Sampler&oldid=166874"](http://shjcx.wang/index.php?title=统计计算_张楠_2019秋:_The_Gibbs_Sampler&oldid=166874)

本页面最后编辑于2019年12月4日 (星期三) 10:14。