MCMC 傻瓜入门

丑高武

前言

在使用 r 语言的时候，你需要与各种各样的分布函数和随机数打交道。

最简单的当然是均匀分布，uniform distribution。

```
x <- runif(n = 1000, min = -1, max = 4)
plot(x)
hist(x)
```

除开均匀分布之外，还有正态分布，学生 t 分布，F 分布，gamma 分布等等

以正态分布为例，R 语言针对每个函数都有四个对应的 “子” 函数, 分别是 d/p/q/n

d 返回密度函数的值，例如,

```
dnorm(x = 0, mean = 0, sd = 1)
1/sqrt(2*pi)
```

p 返回累积分布函数的值，例如,

```
pnorm(q = 0, mean = 0, sd = 1)
pnorm(q = 3, mean = 0, sd = 1)
```

q 是 p 的反函数，例如,

```
qnorm(p = 0.5, mean = 0, sd = 1)
qnorm(p = 0.9986501, mean = 0, sd = 1)
```

r 生成随机数, 例如,

```r
x <- rnorm(1000)
hist(x,nclass = 30)
```

贝叶斯

数据 Y 来自于某分布其中 θ 是未知向量参数。

有很多方法可以估计 θ，例如最大似然估计，找到似然函数 $L(y|\theta)$, 并估计一个 θ 值但是这并不一定是利用数据的最 “好” 方式。

MLE 的思想概括:

- 1. 先估计参数，参数的估计是最大化样本产生概率的参数，$\pi_{MLE} = argmax_{\pi} P(data|\pi)$
- 2. 再产生新数据，新数据的产生使用估计的参数，$P(newdata|data) = P(newdata|\pi_{MLE})$

考虑下面这个例子，

投掷一枚骰子 10 次，其中正面朝上的次数为 4 次，求正面朝上的概率 θ 是多少。

```r
# MLE method
f1 <- function(x) return(dbinom(4, 10, x))
plot(seq(0,1,by = 0.01),sapply(seq(0,1,by = 0.01), f1),type = "l")
```

假设 θ 来自于先验分布 $\pi_0(\theta)$ ，利用贝叶斯定理我们可以获得后验分布 $\pi_1(\theta|y) \propto L(y|\theta) * \pi_0(\theta)$

假设 $Y_i \sim N(\theta, 1)$

假设 θ 的先验分布是 $\pi_0(\theta) = \frac{1}{\pi(1+\theta^2)}$

我们可以求出

$\pi_1(\theta|y) \propto exp - \frac{\Sigma_{i=1}^{n}(y_i-\theta)^2}{2} * \frac{1}{1+\theta^2}$

蒙特卡洛模拟

一个基本的应用是估计积分。

$E_\pi[h(X)] = \int h(x)\pi(x)dx,$

例如，很多例子。

但是如果我们能够从分布 $\pi(x)$ 中抽取一堆样本 X_i，算一个平均值就可以获得这个积分。

但是抽样也是很麻烦的一件事情，一些简单的分布可以抽样，但是难一点的分布就不行了。

抽样

从 0-1 分布中抽样很简单。用函数 runif()。

```
hist(runif(100000),nclass = 20)
```

从指数分布中抽样叶很简单。用函数 rexp()。

```
hist(rexp(100000),nclass = 20)
```

但是这个指数函数随机数的函数中发生了什么？我们考虑指数函数的 pdf $p(x) = \lambda e^{-\lambda x}$

如何利用 0-1 分布来生成指数分布呢，计算指数函数的 cdf $h(x) = 1 - e^{-\lambda x}$ 然后计算反函数 $x = -\lambda^{-1} ln(1-z)$ 这样就可以生成指数函数了。

```
z <- runif(100000)


lambda <- 1

```

```
x <- -lambda^{-1}*log(1-z)


par(mfrow = c(1,2))
hist(x,nclass = 20)
hist(rexp(1000,lambda),nclass = 20)
```

但是有的时候这个反函数是找不到的，那么这个时候我们就需要寻求其他的方法。

我们?rgamma 一下

```
?rgamma
```

里面有句话说:

> rgamma for shape >= 1 uses Ahrens, J. H. and Dieter, U. (1982). Generating gamma variates by a modified *rejection technique*. Communications of the ACM, 25, 47–54,

拒绝抽样 (Reject Sampling)

考虑一个奇怪的分布 $p(x) = \lambda e^{-\lambda x}$

其实就是刚刚的指数分布

```
par(mfrow = c(1,1))
plot(seq(0,10,by=0.001),dexp(seq(0,10,by=0.001)),type = "l")
lines(seq(0,10,by=0.001),rep(1,length(seq(0,10,by=0.001))),col = "red")


iteration <- 100000
lambda <- 1
RejSamp <- function() {
```

```r
  # roll a number from the uniform distribution
  x <- runif(1,0.000000000001,10)

  # calculate the corresponding density
  px <- lambda*exp(-lambda*x)

  # goddice is the judgement variable to determine
  # whether or not keep the number
  goddice <- runif(1)

  if (goddice > px/1) {
    return(NA)
  } else {
    return(x)
  }
}

x <- sapply(1:iteration,function(x) RejSamp())

par(mfrow = c(1,2))
hist(x[!is.na(x)],nclass = 20)
hist(rexp(length(x[!is.na(x)])),nclass = 20)
```

当然指数分布不应该用这个方法。

马尔科夫链

有的时候那根横线很难找到，即那个分布很难找到。所以需要用到 MCMC 的方法。

马尔科夫链有很多性质。

- Stationarity : converges to its stationary distribution
- Irreducibility : Assuming a stationary distribution exists, it is unique if the chain is irreducible.
- Aperiodicity : A Markov chain taking only finite number of values is aperiodic
- Ergodicity

不一一叙述。

如何找到一个平稳分布是目标分布 $\pi(x)$ 的马尔可夫链呢?

Metropolis-Hastings algorithm

在每一次循环,

- 步骤 1 : 从 $y \sim q(y|x^t)$ 中抽取 y，其中，$q()$ 是 proposal distribution, x^t 是现在的状态。
- 步骤 2 : 以概率 $\alpha(x^t, y) = min\{1, \frac{\pi(y)q(x^t|y)}{\pi(x^t)q(y|x^t)}\}$ 接受 $x^{t+1} = y$, 否则 $x^{t+1} = x^t$

一个例子

proposal 有一个新的性质。

symmetric : $q(x|y) = q(y|x)$

在这个情况下

$\alpha(x^t, y) = min\{1, \frac{\pi(y)q(x^t|y)}{\pi(x^t)q(y|x^t)}\} = min\{1, \frac{\pi(y)}{\pi(x^t)}\}$

假设 $\pi(x) \propto exp - \frac{x^2}{2}$

让 proposal $q(y|x) \propto exp - \frac{(y-x)^2}{2*0.5^2}$

```
par(mfrow=c(1,1))
plot(seq(-10,10,by=0.01),dnorm(seq(-10,10,by=0.01)),type = "l")
lines(seq(-10,10,by=0.01),dnorm(seq(-10,10,by=0.01),mean = -2),col = "red")
lines(seq(-10,10,by=0.01),dnorm(seq(-10,10,by=0.01),mean = 2),col = "blue")
```

Gibbs Sampler

Gibbs Sampler 的原理和之前的抽样很类似，不过它的大致思路是当你需要抽取多个值的时候, 例如 $z_1, z_2, z_3......$，不是一次性抽取它们，而是一个个抽取它们。

以 Bivariate Normal Distribution 二元正态分布为例子。

$P(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp[-\frac{z}{2(1-\rho^2)}]$

假设 x 和 y 的均值和方差均为 0 和 1，但是它们之间的相关系数不为 0，如何抽样。

一种简单的方法是，先抽 x，再根据 ρ 抽 y

```
BiNromSimple <- function (n, rho) {
  x <- rnorm(n, 0, 1)
  y <- rnorm(n, rho*x, sqrt(1-rho^2))
  return(data.frame(x,y,stringsAsFactors = F))
}


par(mfrow=c(1,1))
```

```r
df1 <- BiNromSimple(1000,0.9)
plot(df1$x,df1$y)
hist(df1$x)
hist(df1$y)

BiNromGibbs <- function (n, rho) {
  x <- 0
  y <- 0
  df <- data.frame(x = rep(0,n), y = rep(0,n))
  df[1,] <- c(x,y)
  for (i in 2:n) {
    x <- rnorm(1, rho*y, sqrt(1 - rho^2))
    y <- rnorm(1, rho*x, sqrt(1 - rho^2))

    df[i,] <- c(x,y)
  }
  return(df)
}

df2 <- BiNromGibbs(1000,0.9)
plot(df2$x,df2$y)
hist(df2$x)
hist(df2$y)
```

参考资料

1. A simple Gibbs sampler, Darren Wilkinson, http://www.mas.ncl.ac.uk/~ndjw1/teaching/sim/gibbs/gibbs.html.

2. GIBBS SAMPLING FOR THE UNINITIATED, Philip Resnik, Eric Hardisty

3. Introduction to Markov chain Monte Carlo with examples from Bayesian statistics, Hakon Tjelmeland

4. Tutorial Lectures on MCMC I, Sujit Sahu

5. Pattern Recognition and Machine Learning, Chapter 11