DATA130004: Homework 9

董晴园 14300680173

2017.12.24

1. Exercise 9.1: Repeat Example 9.1 for the target distribution Rayleigh($\sigma = 2$). Compare the performance of the Metropolis-Hastings sampler for Example 9.1 and this problem. In particular, what differences are obvious from the plot corresponding to Figure 9.1?

```
f <- function(x, sigma) {
  if (any(x < 0)) return (0)
  stopifnot(sigma > 0)
  return((x / sigma^2) * exp(-x^2 / (2*sigma^2)))
}

m <- 10000
sigma <- 2
x <- numeric(m)
x[1] <- rchisq(1, df=1)
k <- 0
u <- runif(m)

for (i in 2:m) {
  xt <- x[i-1]
  y <- rchisq(1, df = xt)
  num <- f(y, sigma) * dchisq(xt, df = y)
  den <- f(xt, sigma) * dchisq(y, df = xt)
  if (u[i] <= num/den) x[i] <- y else {
    x[i] <- xt
    k <- k+1     #y is rejected
  }
}

print(k)

index <- 5000:5500
y1 <- x[index]
plot(index, y1, type="l", main="", ylab="x")
```

Here's the result:

```
> print(k)
[1] 5257
```

The plot is showed in Figure 1. Comparing to Figure 2 (which is the result in Exercise 9.1), the obvious differences are:

(a) More candidate points are rejected when $\sigma = 2$.

(b) The overall variance of the chain is smaller when $\sigma = 2$.

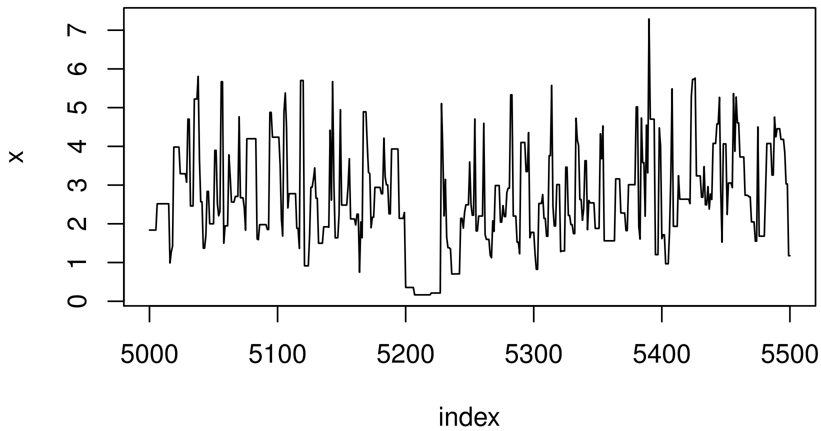

Figure 1: MH sampler for Rayleigh($\sigma = 2$) from $\chi^2(X_t)$

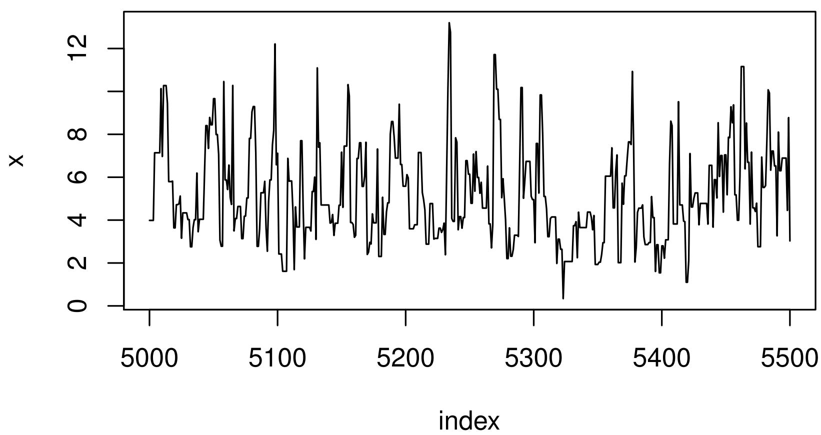

Figure 2: MH sampler for Rayleigh($\sigma = 4$) from $\chi^2(X_t)$

2. Exercise 9.2: Repeat Example 9.1 using the proposal distribution $Y \sim Gamma(X_t, 1)$ (shape parameter X_t and rate parameter 1).

```
f <- function(x, sigma) {
  if (any(x < 0)) return (0)
  stopifnot(sigma > 0)
  return((x / sigma^2) * exp(-x^2 / (2*sigma^2)))
}

m <- 10000
sigma <- 4
x <- numeric(m)
x[1] <- rgamma(1, 1)
k <- 0
u <- runif(m)

for (i in 2:m) {
  xt <- x[i-1]
  y <- rgamma(1, xt)
  num <- f(y, sigma) * dgamma(xt, y)
  den <- f(xt, sigma) * dgamma(y, xt)
  if (u[i] <= num/den) x[i] <- y else {
    x[i] <- xt
    k <- k+1     #y is rejected
  }
}

print(k)

index <- 5000:5500
y1 <- x[index]
plot(index, y1, type="l", main="", ylab="x")
```

Here's the result:

```
> print(k)
[1] 3002
```

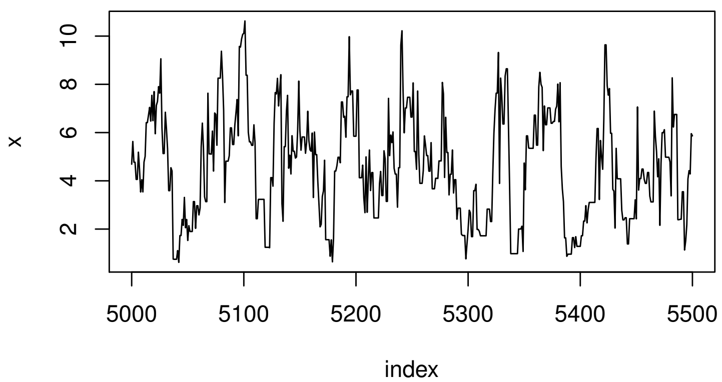

Figure 3: MH sampler for Rayleigh($\sigma = 4$) from Gamma distribution

3. Exercise 9.3: Use the Metropolis-Hastings sampler to generate random variables from a standard Cauchy distribution. Discard the first 1000 of the chain, and compare the deciles of the generated observations with the deciles of the standard Cauchy distribution (see `qcauchy` or `qt` with df=1). Recall that a Cauchy(θ, η) distribution has density function

$$f(x) = \frac{1}{\theta\pi(1 + [(x-\eta)/\theta]^2)}, \qquad -\infty < x < \infty, \theta > 0.$$

The standard Cauchy has the Cauchy($\theta = 1, \eta = 0$) density. (Note that the standard Cauchy density is equal to the Student t density with one degree of freedom.)

For the proposal distribution, I use Normal distribution with $mean = X_t$, $var = 1$. The implementation of the Metropolis-Hastings sampler is as follows:

```
# target density
f <- function(x, theta, eta) {
  stopifnot(theta > 0)
  return(1/(theta * pi * (1 + ((x-eta)/theta)^2)))
}
# MH sampler
m <- 10000
theta <- 1
eta <- 0
x <- numeric(m)
x[1] <- rnorm(1, 0, theta)
k <- 0
u <- runif(m)
for (i in 2:m) {
  xt <- x[i-1]
  y <- rnorm(1, xt, theta)
  num <- f(y, theta, eta) * dnorm(xt, y, theta)
  den <- f(xt, theta, eta) * dnorm(y, xt, theta)
  if (u[i] <= num/den) x[i] <- y else {
    x[i] <- xt
    k <- k+1
  }
}
print(k)
```

Here's the result:

```
> print(k)
[1] 2362
```

To get a plot of the truncated generated chain (see Figure 4):

```
index <- 5000:5500
y1 <- x[index]
plot(index, y1, type="l", main="", ylab="x")
```

To compare the deciles of the sample and theory:

```
b <- 1001      #discard the burnin sample
y <- x[b:m]
a <- seq(0, 1, 0.1)
QC <- qcauchy(a)
Q <- quantile(y, a)
cbind(Q, QC)
```

Here's the result:

```
                Q          QC
0%    -19.10049498        -Inf
10%    -2.47506340 -3.0776835
20%    -1.17297009 -1.3763819
30%    -0.61298864 -0.7265425
40%    -0.25396347 -0.3249197
50%     0.04512704  0.0000000
60%     0.32776135  0.3249197
70%     0.70005034  0.7265425
80%     1.26024047  1.3763819
90%     2.44678662  3.0776835
100%   16.55789425        Inf
```

Figure 4: Part of a chain generated by a Metropolis-Hastings sampler of a Cauchy distribution

To see the QQ plot of quantile distribution (Figure 5) and the histogram of the sample distribution (Figure 6):

```
# qqplot & histogram
a <- ppoints(100)
QC <- qcauchy(a)
Q <- quantile(y, a)
qqplot(QC, Q, main="",
       xlab="Cauchy␣Quantiles", ylab="Sample␣Quantiles")
abline(0,1, col='red')
hist(y, breaks="scott", main="", xlab="", freq=FALSE)
lines(QC, f(QC, theta, eta))
```

4. Exercise 9.7: Implement a Gibbs sampler to generate a bivariate normal chain (X_t, Y_t) with zero means, unit standard deviations, and correlation 0.9. Plot the generated sample after discarding a suitable burn-in sample. Fit a simple linear regression model $Y = \beta_0 + \beta_1 X$ to the sample and check the residuals of the model for normality and constant variance.

```
#initialize constants and parameters
N <- 5000               #length of chain
```

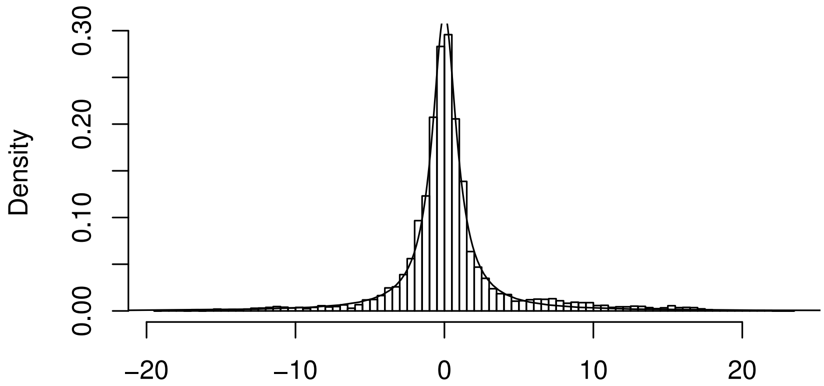

Figure 5: Histogram with target Cauchy density

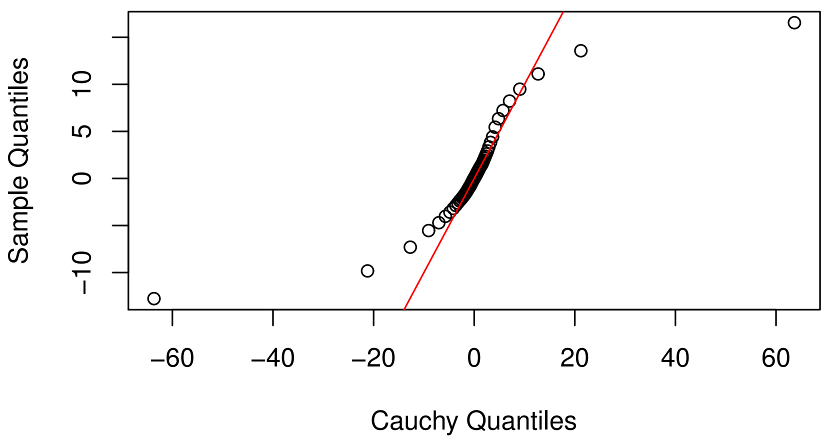

Figure 6: QQ plot for a MH chain in Exercise 9.3

```
burn<- 1000                #burn-in length
X <- matrix(0, N, 2)        #the chain, a bivariate sample

rho <- .9                  #correlation
mu1 <- mu2 <- 0
sigma1 <- sigma2 <- 1
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, ]
```

The corresponding plot of the generated sample after discarding the first 1000 burn-in samples is showed in Figure 7.

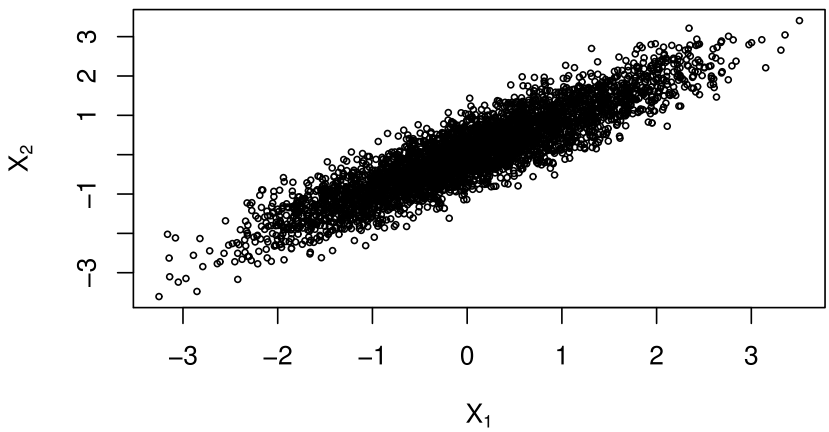

Figure 7: Plot of the generated sample from Exercise 9.7

Using linear regression model to check the residuals:

```
reg.mod <- lm(x[,2] ~ x[,1])
resi <- reg.mod$residuals
c <- seq(min(resi), max(resi), 0.01)
hist(resi, breaks="scott", main="", xlab="", freq=FALSE)
lines(c, dnorm(c, mean = mean(resi), sd = sqrt(var(resi))))
```

As indicated in Figure 8, the distribution of the residuals fit in well with a normal distribution, where mean = `mean(resi) = 1.133425e-18`, variance = `var(resi) = 0.1943319`.

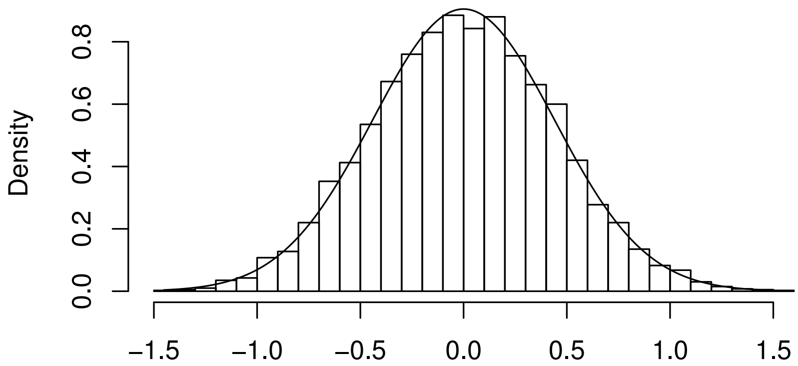

Figure 8: Histogram of the residuals of the simple regression model in Exercise 9.7

5. In Bayesian inference, if the posterior distributions $p(\theta|x)$ are in the same family as the prior probability distribution $\pi(\theta)$, the prior and posterior are then called conjugate distributions, and the prior is called a conjugate prior for the likelihood function. Let $x = (x_1, ..., x_n)^\top$ consist of n i.i.d. random samples. Prove the following well-know conjugate results by specifying the posterior distribution type and parameters.

   (a) When the variance parameter is known, the Gaussian family is conjugate to itself (or self-conjugate) with respect to a Gaussian likelihood function. That is, if the likelihood function is Gaussian, choosing a Gaussian prior over the mean will ensure that the posterior distribution is also Gaussian.

   Assume that the likelihood function follows $\mathrm{N}(\theta, \sigma_\theta^2)$, and the prior distribution follows $\mathrm{N}(\mu_\theta, \sigma_\theta^2)$, donated as

$$
\begin{aligned}
f(x|\theta, \sigma_x) &= \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\theta)^2}{2\sigma_x^2}} \\
\pi(\theta|\mu_\theta, \sigma_\theta) &= \frac{1}{\sqrt{2\pi\sigma_\theta^2}} e^{-\frac{(\theta-\mu_\theta)^2}{2\sigma_\theta^2}}.
\end{aligned}
$$

   In Bayesian inference, the posterior distribution can be computed as

$$
p(\theta|x) \propto \pi(\theta) f(x|\theta).
$$

   Thus, the posterior distribution of θ in this case is

$$
\begin{aligned}
p(\theta|x, \mu_\theta, \sigma_\theta, \sigma_x) &\propto e^{-\frac{(\theta-\mu_\theta)^2}{2\sigma_\theta^2} + \frac{(x-\theta)^2}{2\sigma_x^2}} \\
&\propto e^{-\frac{\left(\sigma_x^2+\sigma_\theta^2\right)\theta^2 - 2\left(\sigma_x^2\mu_\theta+\sigma_\theta^2 x\right)\theta + \sigma_x^2\mu_\theta^2 + \sigma_\theta^2 x^2}{2\sigma_\theta^2\sigma_x^2}} \\
&\propto e^{\frac{\left(\theta - \frac{\sigma_x^2\mu_\theta+\sigma_\theta^2 x}{\sigma_x^2+\sigma_\theta^2}\right)^2}{2\frac{\sigma_x^2\sigma_\theta^2}{\sigma_x^2+\sigma_\theta^2}}}.
\end{aligned}
$$

As proved before, the posterior distribution of θ follows $\mathrm{N}(\mu, \sigma^2)$, where

$$\begin{aligned}
\mu &= \frac{\sigma_x^2 \mu_\theta + \sigma_\theta^2 x}{\sigma_x^2 + \sigma_\theta^2}, \\
\sigma^2 &= \frac{\sigma_x^2 \sigma_\theta^2}{\sigma_x^2 + \sigma_\theta^2}.
\end{aligned}$$

(b) Gamma distribution is the conjugate prior for the rate parameter of Poisson distribution. That is, given data are generated according to Poisson(λ), specify a Gamma(α, β) prior for λ, then the posterior distribution $p(\lambda|x)$ is still Gamma.

Given that

$$\begin{aligned}
f(x|\lambda) &= \frac{\lambda^x}{x!} e^{-\lambda} \\
\pi(\lambda) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda},
\end{aligned}$$

The posterior distribution of λ

$$\begin{aligned}
p(\lambda|x, \alpha, \beta) &\propto \lambda^{x+\alpha-1} e^{-(\alpha+1)\lambda} \\
&\sim Gamma(x+\alpha, \alpha+1).
\end{aligned}$$