STAT 4010 Bayesian Learning

Tutorial 9

Spring 2022

Cheuk Hin (Andy) CHENG (Email | Homepage)

Di SU (Email | Homepage)

1 Reweighting Importance Sampling

Recall that in importance sampling, we aim at simulating

$$I = \int_\Theta g(\theta)\pi_u(\theta)\mathrm{d}\theta = \int_\Theta g(\theta)\frac{\pi_u(\theta)}{p(\theta)}p(\theta)\mathrm{d}\theta = \int_\Theta g(\theta)w(\theta)p(\theta)\mathrm{d}\theta,$$

where $w(\theta) = \pi_u(\theta)/p(\theta)$. We can use the reweighting importance sampling (RIS) to generate samples from target density.

Algorithm 1: Reweighting Importance Sampling.

Input: (i) simulation size J; (ii) proposed PDF $p(\cdot)$; (iii) (unnormalized) target density $\pi_u(\cdot)$.

begin

(1) Generate $\tilde\theta_1, \ldots, \tilde\theta_J \overset{\text{IID}}{\sim} p(\cdot)$.

(2) Compute $w_j = \pi_u(\tilde\theta_j)/p(\tilde\theta_j)$, for $j = 1, \ldots, J$.

(3) **for** j *in* $\{1, \ldots, J\}$ **do**

Draw one element from $\{\tilde\theta_1, \ldots, \tilde\theta_J\}$ with probabilities $\{w_1, \ldots, w_J\}$.

Saved the sampled element to θ_j

end

end

Output: $\theta_{1:J}$

Theorem 1.1. *(Justification of the RIS) If* $\mathsf{Var}_p(w(\theta)) < \infty$*, then as* $J \to \infty$ *for all* $j = 1, \ldots, J$ *and* $t \in \mathbb{R}$*,*

$$\Pr(\theta_j \le t) \to \int_{-\infty}^t \pi(\theta)\mathrm{d}\theta.$$

Remark 1.1. For estimating I using the RIS, we may consider

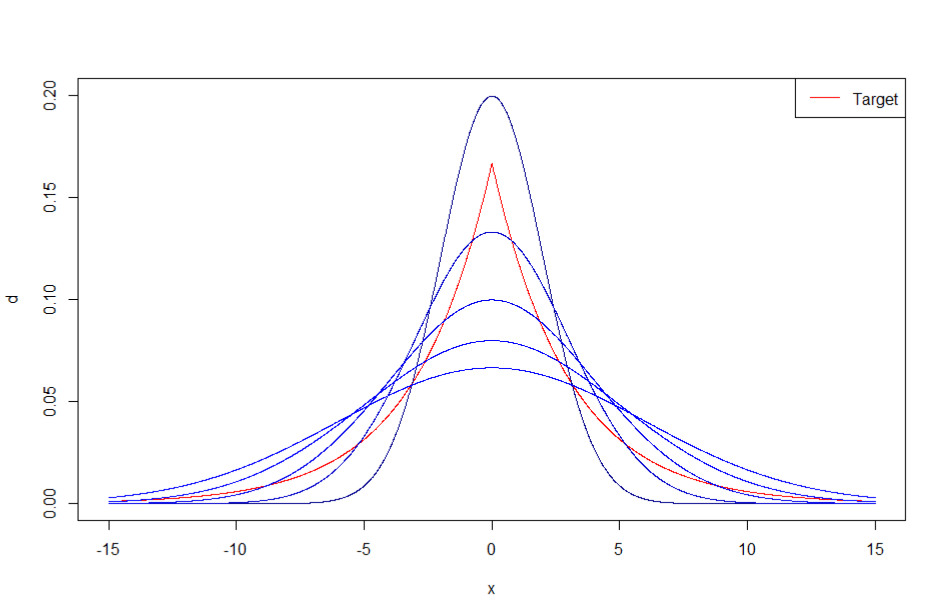
$$\hat I_{RIS} = \frac{1}{J}\sum_{j=1}^J g(\theta_j),$$

where $\{\theta_1, \cdots, \theta_J\}$ are the outputs of RIS.

Example 1.1. Consider $\theta \sim F(\theta)$ and $f(\theta) \propto \exp\{-|\theta|/3\}\mathbb{1}(\theta \in \mathbb{R})$. Simulate $\mathsf{E}\theta^2$ using RIS.

Solution: Note that we use a Normal distribution as the proposal following last tutorial. Also, $\mathsf{E}\theta^2 = 18$.

```r
target_den <- function(theta,b=3){
  log_d = -abs(theta)/b
  exp(log_d - max(log_d))
}

target_g <- function(theta){
  theta^2
}

##RIS
RIS <- function(J,sd = 4, b= 3){ ##sd being the sd of the proposal
  tilde_theta = rnorm(J,0,sd)
  w = target_den(tilde_theta,b)/dnorm(tilde_theta,0,sd)
  sample(tilde_theta,size = J,replace=T,prob = w) ##output the theta_{1;J}
}

sim_I_RIS <- function(J,sd=4,b=3){
  theta_sim = RIS(J,sd,b)
  mean(target_g(theta_sim))
}

J_sim = 2^14
sim_I_RIS(J_sim,sd = 2)
[1] 10.91392
sim_I_RIS(J_sim,sd = 3)
[1] 13.19787
sim_I_RIS(J_sim,sd = 4)
[1] 15.00232
sim_I_RIS(J_sim,sd = 5)
[1] 17.36904
sim_I_RIS(J_sim,sd = 6)
[1] 18.6439
sim_I_RIS(J_sim,sd = 7)
[1] 18.78481
sim_I_RIS(J_sim,sd = 8)
[1] 17.99314

x = seq(-15,15,0.01)
d = exp(-abs(x)/3)/(2*3)
plot(x,d,type='l',ylim=c(0,0.2),col='red',main = 'Target Density vs Proposal')
lines(x,dnorm(x,0,2),col='blue4')
lines(x,dnorm(x,0,3),col='blue3')
lines(x,dnorm(x,0,4),col='blue2')
lines(x,dnorm(x,0,5),col='blue1')
lines(x,dnorm(x,0,6),col='blue')
legend('topright',legend=c('Target'),col='red',lwd=1.5)
```



2 Markov Chain Monte Carlo (MCMC)

2.1 Theoretical Background

Definition 1. *Let θ_J be a Θ-valued RV for each $j = 1, \ldots, J$.*

1. *A collection of* RVs $\{\theta_j : j = 1, 2, \ldots\}$ *is said to be a Markov chain if*

$$\forall j \in \{1, \ldots, J\}, \quad \left[\theta_j \mid \theta_{1:(j-1)}\right] = \left[\theta_j \mid \theta_{j-1}\right].$$

2. *A Markov chain $\{\theta_j\}$ is said to be time-homogeneous (or simply homogenous) if*

$$\forall \delta, \quad \forall k, \quad \forall j_0 \leq j_1 \leq \ldots \leq j_k, \quad [\theta_{j_1}, \ldots, \theta_{j_k} \mid \theta_{j_0}] = [\theta_{j_1+\delta}, \ldots, \theta_{j_k+\delta} \mid \theta_{j_0+\delta}]$$

3. *The transition kernel $K(\cdot \mid \cdot)$ of a homogeneous Markov chain $\{\theta_j\}$ is the conditional density of $[\theta_j \mid \theta_{j-1}]$, i.e.,*

$$K\left(\theta_j \mid \theta_{j-1}\right) = f\left(\theta_j \mid \theta_{j-1}\right).$$

Theorem 2.1. *(Important distribution related to a Markov Chain)*

1. *A Markov chain $\{\theta_j\}$ is said to have a stationary distribution π if*

$$\pi(\theta) = \int_\Theta K\left(\theta \mid \theta'\right) \pi\left(\theta'\right) \mathrm{d}\theta'.$$

2. *A Markov chain $\{\theta_j\}$ is said to satisfy the detailed balance condition with a probability density π if*

$$\forall \theta, \theta' \in \Theta, \quad K\left(\theta \mid \theta'\right) \pi\left(\theta'\right) = K\left(\theta' \mid \theta\right) \pi(\theta).$$

3. *(Detailed balance implies stationarity). If a Markov chain $\{\theta_j\}$ satisfies the detailed balance condition with π, then π is the stationary distribution.*

4. *(Convergence of Markov chain). Let $\{\theta_j\}$ be a Markov chain with $\theta_0 \sim \pi_{initial}(\cdot)$. If (i) π is the stationary distribution of the chain, and (ii) regularity conditions hold, then for large J*

$$[\theta_J] \approx \pi(\cdot).$$

Remark 2.1. The above theorem suggests that, if we can design a suitable K so that the corresponding chain has a limiting distribution π, then when this chain is long enough, the samples from this chain can be regarded as samples from the target distribution.

Theorem 2.2. *(Limiting behaviours of MCMC estimator)*

1. *(Ergodic theorem). Let g be a function such that $\mathrm{E}_\pi|g(\theta)| < \infty$. Let $\{\theta_j\}$ be a Markov chainwith $\theta_0 \sim \pi_{initial}(\cdot)$. Define*

$$\widehat{I}_{\mathrm{MCMC}} := \frac{1}{J}\sum_{j=1}^{J} g(\theta_j) \qquad \textit{and} \qquad I = \mathrm{E}_\pi\{g(\theta)\}.$$

If (i) π is the stationary distribution of the chain, and (ii) regularity conditions hold, then as $J \to \infty$ $\widehat{I}_{\mathrm{MCMC}} \overset{wp1}{\to} I$.

2. *(Central limit theorem). Let g be a function such that $\mathrm{E}_\pi\{g^2(\theta)\} < \infty$. Let $\{\theta_j\}$ be a Markov chain with $\theta_0 \sim \pi_{initial}(\cdot)$. If π is the stationary distribution of $\{\theta_j\}$ and regularity conditions hold, then for any $\pi_{initial}(\cdot)$*

$$\sqrt{J}\left(\widehat{I}_{\mathrm{MCMC}} - I\right) \overset{\mathrm{d}}{\to} \mathrm{N}\left(0, \sigma^2_{\mathrm{MCMC}}\right)$$

where

$$\sigma^2_{\mathrm{MCMC}} = \sum_{k\in\mathbb{Z}} \gamma_k, \qquad \textit{and} \qquad \gamma_k = \mathrm{Cov}_\pi\{g(\theta_0), g(\theta_k)\}$$

2.2 Metropolis–Hastings (MH) Algorithm

Algorithm 2: Metropolis–Hastings Algorithm.

Input: (i) simulation size J; (ii) proposed PDF $p(\cdot \mid \cdot)$; (iii) (unnormalized) target density $\pi_u(\cdot)$; (iv) initialization density $\pi_{\text{initial}}(\cdot)$.

begin

(1) Generate $\theta_0 \overset{\text{IID}}{\sim} \pi_{\text{initial}}(\cdot)$.

(2) **for** j *in* $\{1, \ldots, J\}$ **do**

Generate $\tilde{\theta}_j \overset{\text{IID}}{\sim} p(\cdot \mid \theta_{j-1})$.

Generate $U_j \sim \text{Unif}(0, 1)$.

Compute the acceptance probability:

$$\begin{aligned} a_j &= \min\left\{1, \frac{\pi_u(\tilde{\theta}_j)}{\pi_u(\theta_{j-1})} \Big/ \frac{p(\tilde{\theta}_j \mid \theta_{j-1})}{p(\theta_{j-1} \mid \tilde{\theta}_j)}\right\} \\ &= \min\left\{1, \text{Target odd} \Big/ \text{Proposal odd}\right\}, \\ \text{Target odd} &= \pi_u(\tilde{\theta}_j)/\pi_u(\theta_{j-1}), \\ \text{Proposal odd} &= p(\tilde{\theta}_j \mid \theta_{j-1})/p(\theta_{j-1} \mid \tilde{\theta}_j). \end{aligned}$$

Compute the value in the jth iteration: $\theta_j = \tilde{\theta}_j \mathbb{1}(U_j \le a_j) + \theta_{j-1}\mathbb{1}(U_j > a_j)$

end

end

Output: $\theta_{1:J}$

Remark 2.2. Some remark.

1. If the proposal is symmetric, i.e. $p(\tilde{\theta} \mid \theta) = p(\theta \mid \tilde{\theta})$, then the proposal odd can be dropped.
2. To avoid numerical issue, it is recommended to compute $a_j = \exp(a_j^*)$ where

$$\begin{aligned} a_j^* &= \min\{0, \log \text{Target odd} - \log \text{Proposal odd}\}, \\ \text{Target odd} &= \log \pi_u(\tilde{\theta}_j) - \log \pi_u(\theta_{j-1}), \\ \text{Proposal odd} &= \log p(\tilde{\theta}_j \mid \theta_{j-1}) - \log p(\theta_{j-1} \mid \tilde{\theta}_j). \end{aligned}$$

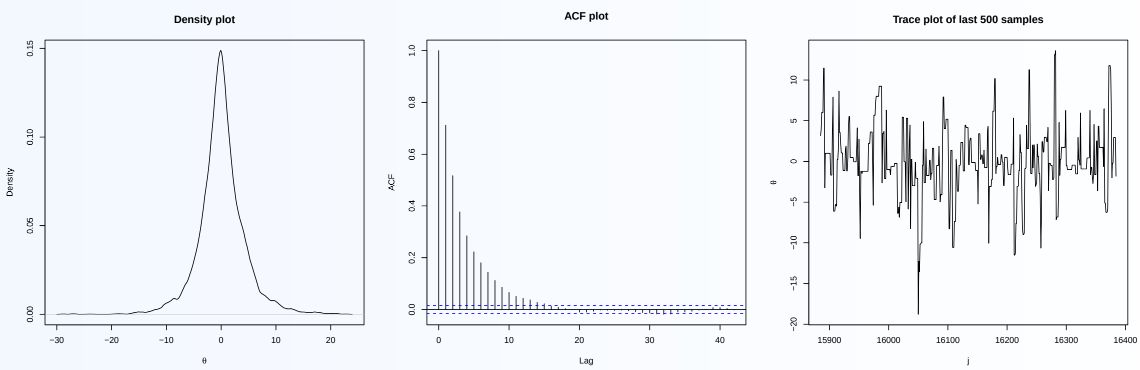
3. Typically, we drop first half (depends on the situation) of the simulated samples and use the remaining for further analysis.

Theorem 2.3 (Justification of MH Algorithm)**.** *The output from MH Algorithm satisfies the detailed balance condition with π.*

Example 2.1. Continue from Example 1.1, estimate $\mathsf{E}\theta^2$ by the MH algorithm.

Solution: Consider the initial proposal being Normal$(0, 8^2)$ and the proposal $\tilde{\theta} \mid \theta \sim$ Normal$(\theta, 8^2)$. Note that the proposal is symmetric in this case.

```
##MH algorithm
##sd0 is the sd of the initial proposal
##sd1 is the sd of the proposal
pi_u <- function(theta,b=3){
  exp(-abs(theta)/b)
}

target_g <- function(theta){
  theta^2
}

MH <- function(J,burn_frac = 0.5, keep = F,sd0 = 8, sd1 = 8,b = 3){
  nsim = ceiling(J/burn_frac)+1
  theta = array(NA,nsim)
  theta[1] = rnorm(1,0,sd0) ##sample from the initial proposal
  for (j in 2:nsim) {
    tilde_theta = rnorm(1,theta[j-1],sd1) ##sample from the proposal
    u = runif(1)
    log_target_odd = log(pi_u(tilde_theta,b)) - log(pi_u(theta[j-1],b))
    a = exp(min(0,log_target_odd))
    theta[j] = tilde_theta*(u <= a) + theta[j-1]*(u> a)
  }
  if (keep) {
    return(theta)
  }else{
    return(tail(theta,J)) ##burn the first fraction
  }
}

J_sim = 2^14
theta_MH = MH(J_sim,sd0 = 8,sd1=8)

##MCMC Diagostic
pi = density(theta_MH, kernel="epanechnikov")
acf = acf(theta_MH,plot=F)
par(mfrow = c(1,3))
plot(pi,main = 'Density plot',ylab = 'Density',xlab = expression(theta))
plot(acf,main = 'ACF plot',ylab = 'ACF',xlab = 'Lag')
plot((J_sim-500+1):J_sim,tail(theta_MH,500),type = 'l',
     main = 'Trace plot of last 500 samples',ylab = expression(theta),xlab = 'j')
##Produce estimate for I
mean(target_g(theta_MH))
[1] 17.82765
```



Example 2.2 (Option Pricing)**.** Black-Scholes Formula is well known for the European option pricing. Specifically, we have the call price c being

$$\begin{aligned} c(\sigma) &= S_0 N(d_1) - Ke^{-rT} N(d_2) \\ d_1 &= \frac{\log(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \\ d_2 &= d_1 - \sigma\sqrt{T}, \end{aligned}$$

where S_0 is current stock price, T is time to maturity, K is the strike price, σ is the volatility of the stock return R and $N(\cdot)$ is the CDF of the standard normal distribution. In practice, the σ is typically unknown. Consider following model,

$$\begin{aligned} R_{1:n} \mid \sigma &\overset{\text{IID}}{\sim} f(R \mid \theta) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}|R|}{\sigma}\right) \mathbb{1}(R \in \mathbb{R}) \\ \sigma &\sim Ga(\alpha)/\beta, \end{aligned}$$

where $\alpha = 0.4$ and $\beta = 2$.

1. Compute the posterior.
2. Using the MH algorithm, give a point estimate for σ.
3. Given the data with $n = 30$ and $A_n = \sum_{i=1}^n |R_i| = 2$. Suggest the call price when $S_0 = 373$, $K = 380$, $T = 0.5$ and $r = 0\%$.

SOLUTION:

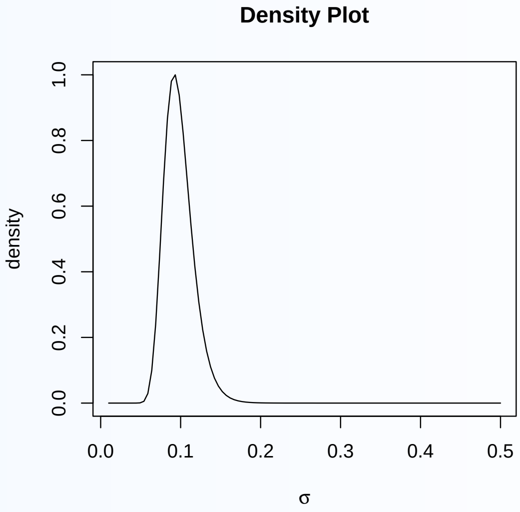
1. By direct computation,

$$f(\sigma \mid R_{1:n}) \propto \sigma^{\alpha-n-1} \exp\left(-\beta\sigma - \frac{\sqrt{2}}{\sigma} A_n\right) \mathbb{1}(\sigma > 0),$$

   where $A_n = \sum_{i=1}^n |R_i|$.

2. We first visualize the target posterior density.

```
An = 2
n = 30

log_pi_u2 = function(sigma,An,n,a= 0.4,b = 2){
  (a-n-1)*log(sigma) - b*sigma - sqrt(2)*An/sigma
}

##visualize the target density
sigma_grid = seq(0.01,0.5,length.out = 101)
log_den = log_pi_u2(sigma_grid,An,n)
plot(sigma_grid,exp(log_den - max(log_den)),type = 'l',
     main = 'Density Plot',ylab ='density',xlab = expression(sigma))
```

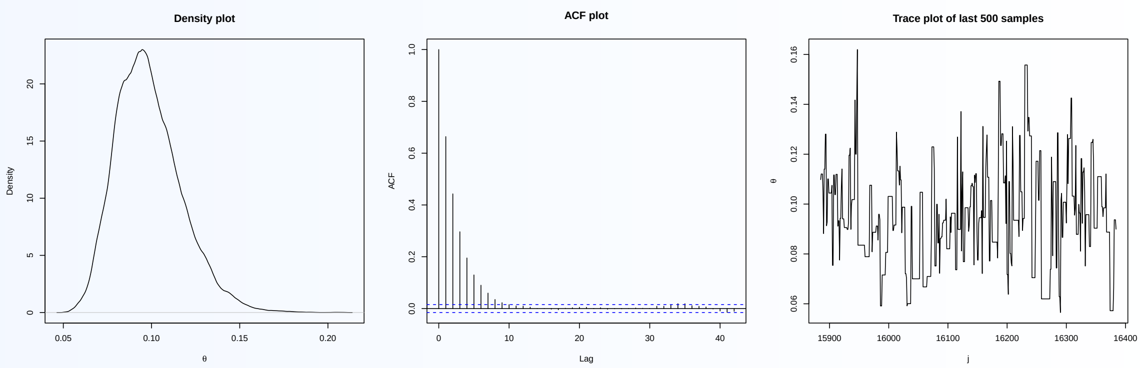


- The plot suggests that we may consider the Gamma distribution as initial point and the proposal.
- From the plot, it suggest that we should consider the initial proposal centers at $a/b = 0.1$ with standard deviation (sd) being $\sqrt{a}/b = 0.06$. Therefore, the initial proposal is $Ga(4)/40$.
- We want the proposal $[\tilde{\theta} \mid \theta]$ to center at θ while keeping the sd same as the initial. Therefore, the proposal is $Ga(400\theta^2)/400\theta$.

The MH algorithm can be implemented by the following code.

```
MH2 <- function(J,An,n,burn_frac = 0.5, keep = F,a0 = 4,b0=40){
  nsim = ceiling(J/burn_frac)+1
  theta = array(NA,nsim)
  theta[1] = rgamma(1,a0,b0) ##sample from the initial proposal
  for (j in 2:nsim) {
    a1 = 400*theta[j-1]^2
    b1 = 400*theta[j-1]
    tilde_theta = rgamma(1,a1,b1) ##sample from the proposal
    u = runif(1)
    log_target_odd = log_pi_u2(tilde_theta,An,n) - log_pi_u2(theta[j-1],
        An,n)
    a2 = 400*tilde_theta^2
    b2 = 400*tilde_theta
    log_proposal_odd = log(dgamma(tilde_theta,a1,b1)) - log(dgamma(theta[
        j-1],a2,b2))
    a = exp(min(0,log_target_odd-log_proposal_odd))
    theta[j] = tilde_theta*(u <= a) + theta[j-1]*(u> a)
  }
  if (keep) {
    return(theta)
  }else{
    return(tail(theta,J)) ##burn the first fraction
  }
}

J_sim = 2^14
```

```r
theta_MH2 = MH2(J_sim,An,n)

##MCMC Diagostic
pi = density(theta_MH2, kernel="epanechnikov")
acf = acf(theta_MH2,plot=F)
windows(height = 5,width = 15)
par(mfrow = c(1,3))
plot(pi,main = 'Density plot',ylab = 'Density',xlab = expression(theta))
plot(acf,main = 'ACF plot',ylab = 'ACF',xlab = 'Lag')
plot((J_sim-500+1):J_sim,tail(theta_MH2,500),type = 'l',
     main = 'Trace plot of last 500 samples',ylab = expression(theta),
        xlab = 'j')
#alternative R functions to produce the same plot
par(mfrow = c(1,3))
plot(pi,main = 'Density plot',ylab = 'Density',xlab = expression(theta))
acf = acf(theta_MH2,plot=T)
ts.plot(tail(theta_MH2,500),type = 'l',
     main = 'Trace plot of last 500 samples',ylab = expression(theta),
        xlab = 'j')
```



3. Under the square loss, the Bayes estimator for the call price $c(\sigma)$ is posterior mean $\mathsf{E}[c(\sigma) \mid R_{1:n}]$ estimated as \$7.2880.

```r
##Option pricing
get_call <- function(sigma_hat,s0 = 373,k = 380,t = 0.5,r = 0){
  d1 = (log(s0/k)+(r+sigma_hat^2/2)*t)/(sigma_hat*sqrt(t))
  d2 = d1-sigma_hat*sqrt(t)
  c = s0*pnorm(d1)-k*exp(-r*t)*pnorm(d2)
  c
}
mean(get_call(theta_MH2))
 [1] 7.287991
```

Remark 2.3. For a comprehensive procedure of using different methods on the same task (together with R codes), you may study Example 7.1 (the model), Example 7.2 (Trapezoidal rule), Example 7.4 (Importance Sampling) 7.6 (Reweighting Importance Sampling) and Example 7.7 (MH Algorithm) in details. It will help assignment 6. However, other examples from the lecture notes are also important.