HW 6

Problem 1

(1)

Find smallest $M > 0$ such that M times the proposal density dominates the target density on their support. For all $\theta > 0$, we have

$$\begin{aligned}
M &\geq \frac{1}{\lambda \exp(-\lambda\theta)} \frac{\phi(\theta + c)}{1 - \Phi(c)} \\
&= \frac{1}{\lambda \exp(-\lambda\theta)} \frac{\exp\left\{-(\theta + c)^2/2\right\}}{\sqrt{2\pi}(1 - \Phi(c))} \\
&= \frac{\exp\left\{-(\theta^2 + 2c\theta + c^2)/2 + \lambda\theta\right\}}{\sqrt{2\pi}\lambda(1 - \Phi(c))} \\
&= \frac{\exp\left\{-(\theta + c - \lambda)^2/2 + (c - \lambda)^2/2 - c^2/2\right\}}{\sqrt{2\pi}\lambda(1 - \Phi(c))}.
\end{aligned}$$

Since $c > 0$ and $\lambda > 0$, we know that $\lambda - c > 0$ so we can choose a $\theta > 0$ such that $\theta = \lambda - c$. This yields a lower bound for M:

$$M \geq \frac{\exp\left\{(c - \lambda)^2/2 - c^2/2\right\}}{\sqrt{2\pi}\lambda(1 - \Phi(c))} = \frac{\exp\left\{(\lambda^2 - 2\lambda c)/2\right\}}{\sqrt{2\pi}\lambda(1 - \Phi(c))}.$$

Since any other $\theta > 0$ will make $-(\theta + c - \lambda)^2$ strictly negative, we conclude that above lower bound is tight.

(2)

Redefine $q(\theta)$ to be the normalized target density.

$$\begin{aligned}
E_g \frac{q(\theta)}{M^* g(\theta)} &= \int_0^\infty \frac{q(\theta)}{M^* g(\theta)} g(\theta)\, d\theta \\
&= \int_0^\infty \frac{q(\theta)}{M^*}\, d\theta \\
&= 1/M^*.
\end{aligned}$$

(3)

To maximize $1/M^*$, we need to minimize $f(\lambda) = \left\{(\lambda^2 - 2\lambda c)/2\right\} - \log \lambda$ on $\lambda > 0$. Taking derivative we have

$$\begin{aligned}
\frac{df}{d\lambda} &= \lambda - c - \frac{1}{\lambda}, \\
\frac{d^2 f}{d\lambda^2} &= 1 + \frac{1}{\lambda^2} > 0.
\end{aligned}$$

Thus it is a strictly convex function on $\lambda > 0$ and by setting the first order derivative to be zero we will find the minimum. Since $\lambda > 0$ we have

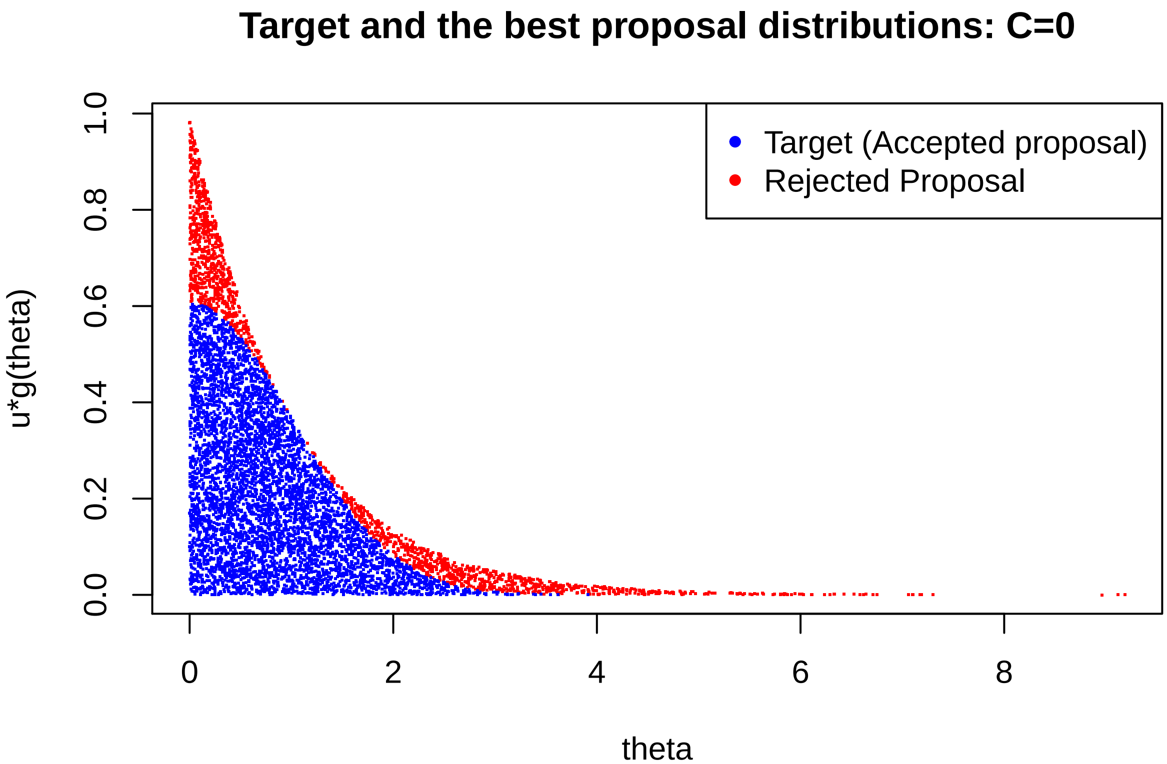
$$\begin{aligned}
\lambda - c - \frac{1}{\lambda} &= 0 \\
\lambda^2 - c\lambda - 1 &= 0 \\
\lambda &= (c + \sqrt{c^2 + 4})/2.
\end{aligned}$$

(4)

```r
library(ggplot2)
set.seed(0)

p <- function(x, C) {
  dnorm(x+C)*(x > 0)/(1-pnorm(C))
}

rej_samp <- function(C, n=7e3) {
  lambda <- (C+sqrt(C^2+4))/2
  M <- exp((lambda^2-2*lambda*C)/2)/(sqrt(2*pi)*lambda*(1-pnorm(C)))

  proposal <- rexp(n, rate=lambda)
  U <- runif(n, 0, 1)
  id <- (log(U) < log(p(proposal, C)) -log(M) - dexp(proposal, rate=lambda, log=T))
  accepted <- proposal[id]

  plot(proposal, U*dexp(proposal, rate=lambda),
    pch=".", col="red", ylab="u*g(theta)", xlab="theta",
    main=paste("Target and the best proposal distributions: C=", C, sep=""))
  points(accepted, (U*dexp(proposal, rate=lambda))[id],
    pch=".", col="blue")
  legend("topright", legend=c("Target (Accepted proposal)", "Rejected Proposal"),
         col=c('blue', 'red'), pch=20)
  cat("Theoretical best acceptance rate is:", 1/M)
  cat("\nEmpirical acceptance rate is:", mean(id))
}
```

$c = 0$

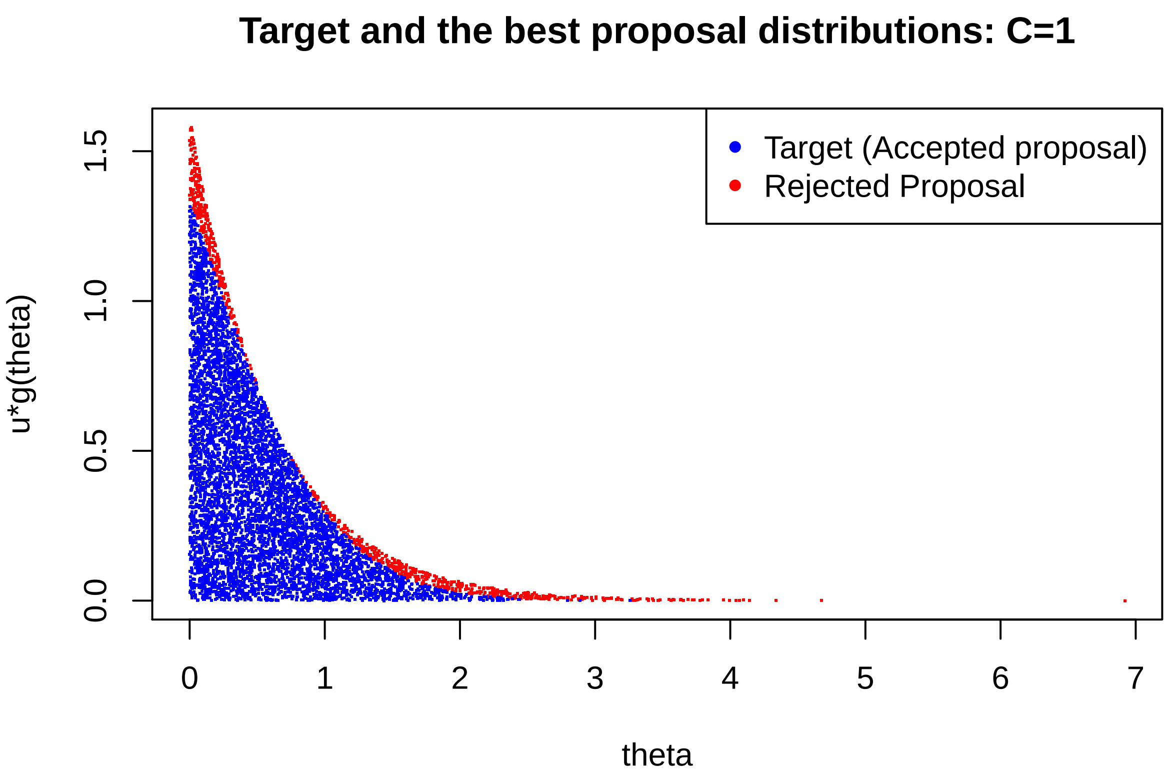
```r
rej_samp(C=0)
```




```
## Theoretical best acceptance rate is: 0.7601735
## Empirical acceptance rate is: 0.7622857
```

$c = 1$

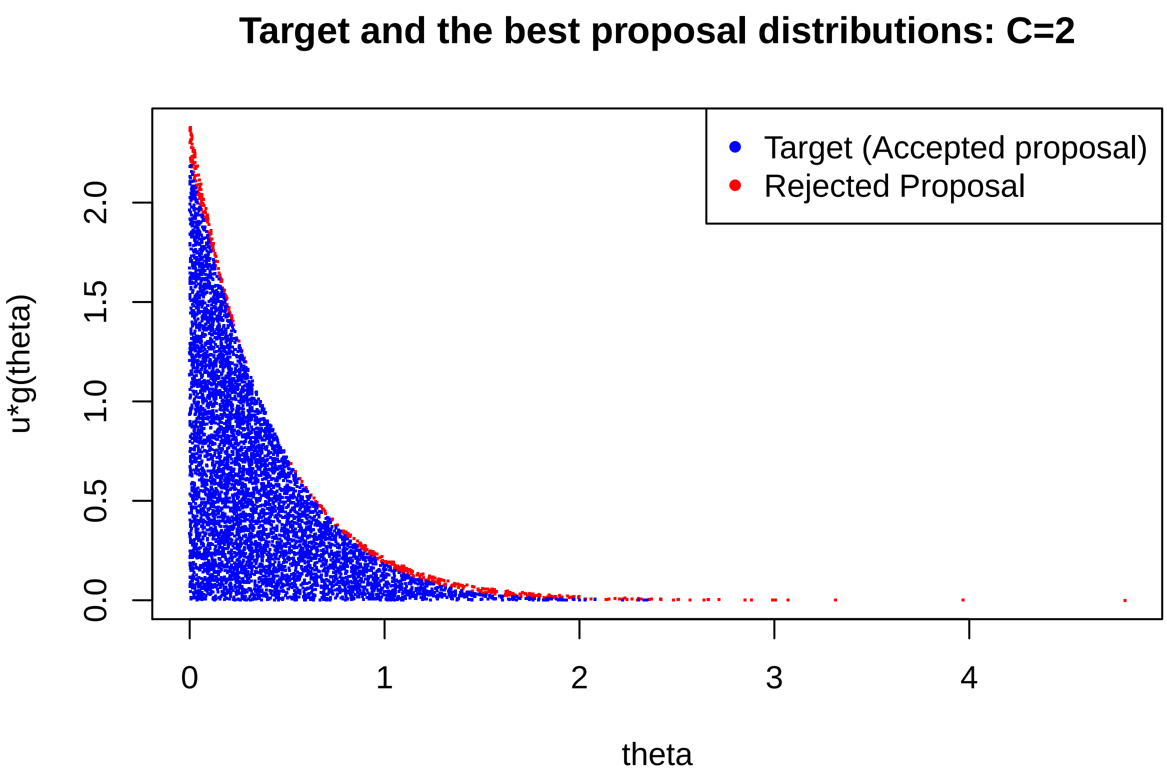
```r
rej_samp(C=1)
```




```
## Theoretical best acceptance rate is: 0.8764687
## Empirical acceptance rate is: 0.8825714
```

$c = 2$

```r
rej_samp(C=2)
```



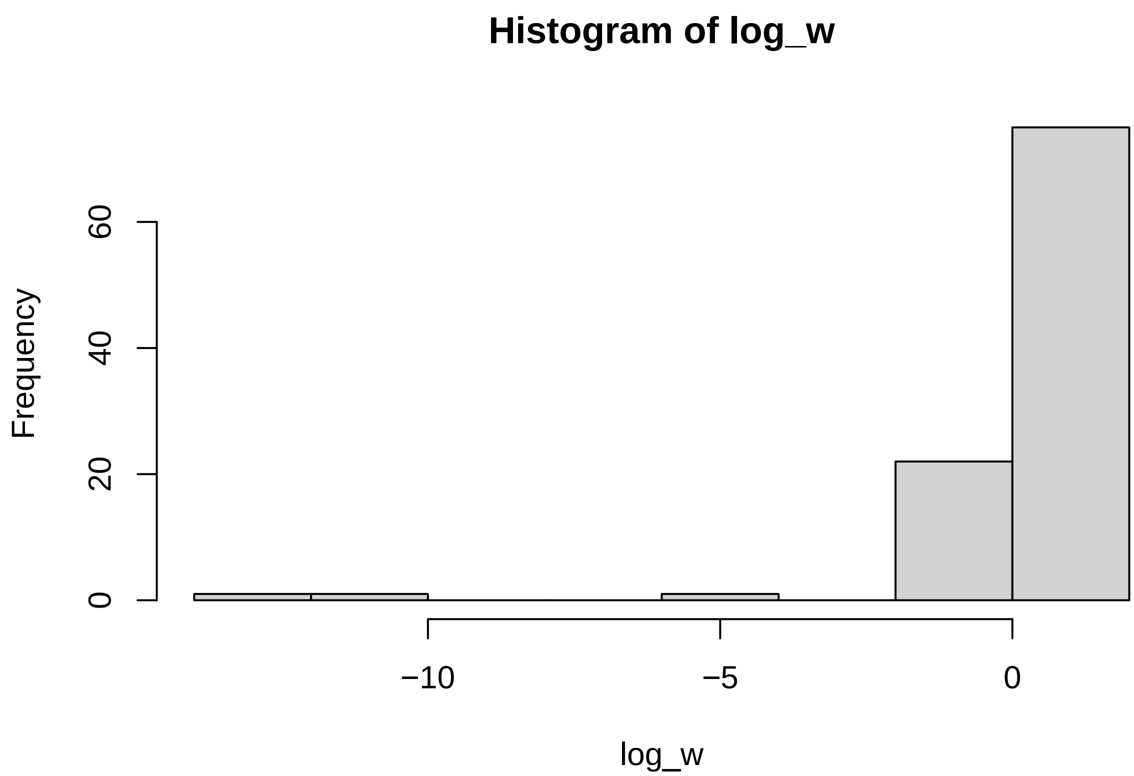

```
## Theoretical best acceptance rate is: 0.9336453
## Empirical acceptance rate is: 0.9317143
```

Problem 2

(1)

Suppose the posterior is standard normal.

```
S <- 1e2
approxi <- rt(S, df=3)
log_w <- dnorm(approxi, log=T) - dt(approxi, df=3, log=T)
hist(log_w)
```



(2)

```
w <- exp(log_w)

mu <- mean(approxi*w)/mean(w)
var <- mean(approxi^2*w)/mean(w) - mu^2

cat("Estimated EX is:" , mu)
```

```
## Estimated EX is: 0.05995679
```

```
cat("\nEstimated VarX is:", var)
```

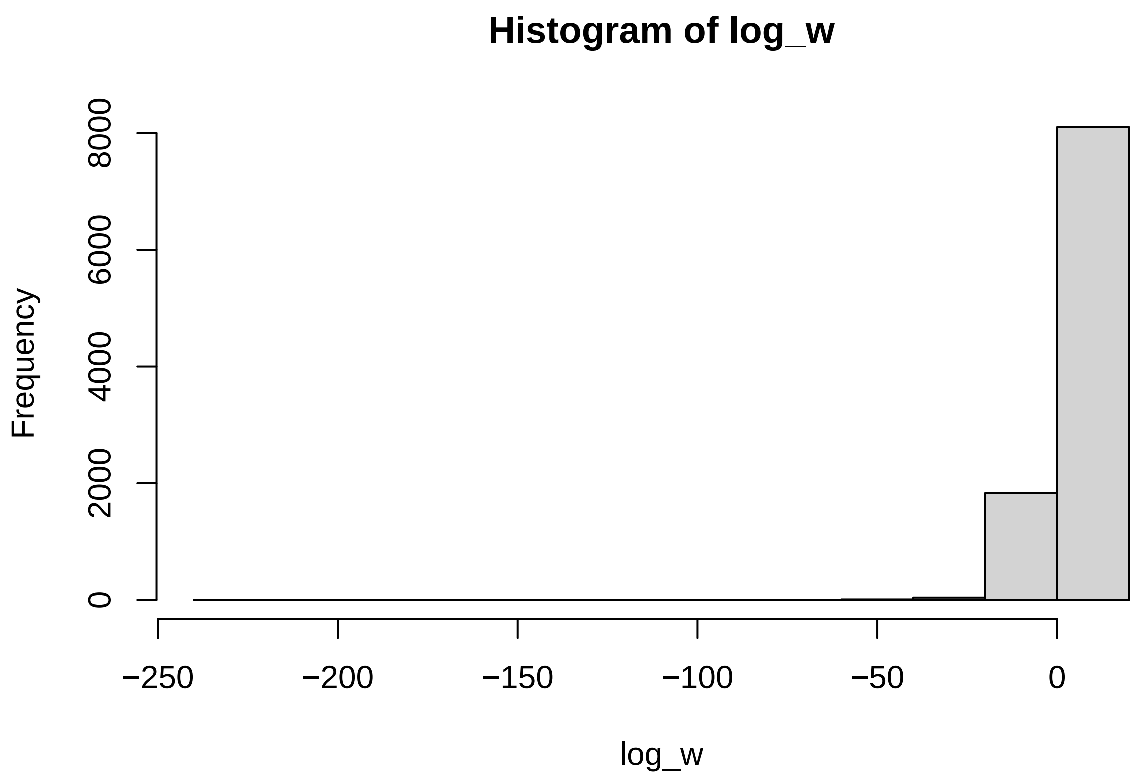
```
##
## Estimated VarX is: 1.147101
```

```
cat("\nTrue values are 0 and 1")
```

```
##
## True values are 0 and 1
```

(3)

```
S <- 1e4
approxi <- rt(S, df=3)
log_w <- dnorm(approxi, log=T) - dt(approxi, df=3, log=T)
hist(log_w)
```




```
w <- exp(log_w)

mu <- mean(approxi*w)/mean(w)
var <- mean(approxi^2*w)/mean(w) - mu^2

cat("Estimated EX is:" , mu)
```

```
## Estimated EX is: 0.01178947
```

```
cat("\nEstimated VarX is:", var)
```

```
##
## Estimated VarX is: 0.9831246
```

```
cat("\nTrue values are 0 and 1")
```

```
##
## True values are 0 and 1
```

(4)

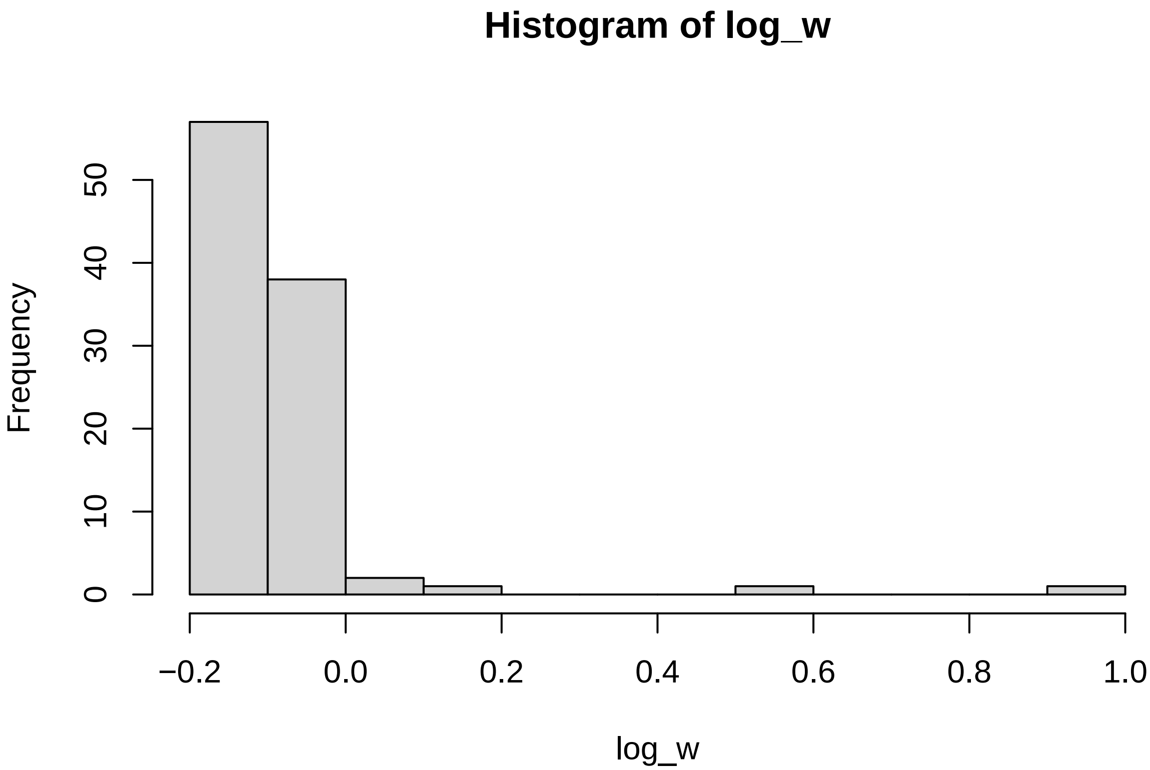
```
w_tilde <- w/sum(w)

S_eff <- 1/sum(w_tilde^2)
S_eff
```

```
## [1] 9170.299
```

Problem 3

Suppose the posterior is t_3. The expectation is still 0, the variance is 3.

```r
S <- 1e2
approxi <- rnorm(S)
log_w <- -dnorm(approxi, log=T) + dt(approxi, df=3, log=T)
hist(log_w)
```




```r
w <- exp(log_w)
mu <- mean(approxi*w)/mean(w)
var <- mean(approxi^2*w)/mean(w) - mu^2

cat("Estimated EX is:" , mu)
```

```
## Estimated EX is: -0.03533384
```

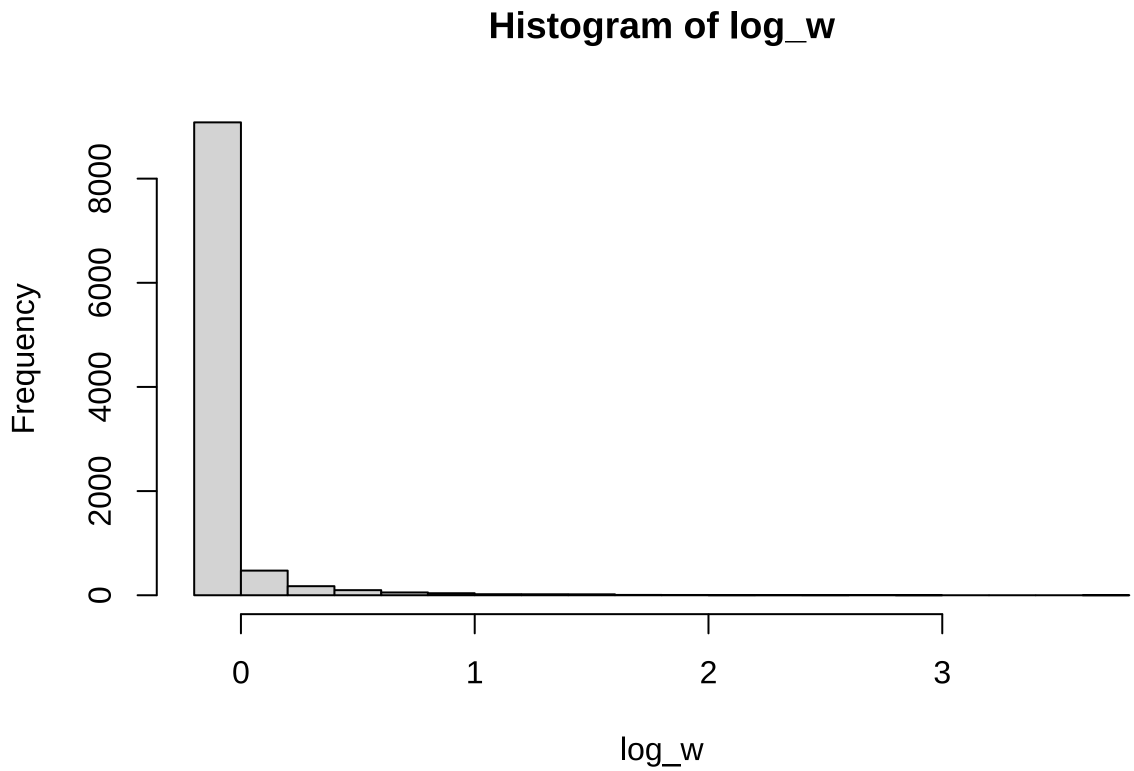
```r
cat("\nEstimated VarX is:", var)
```

```
##
## Estimated VarX is: 0.9002956
```

```r
cat("\nTrue values are 0 and 3")
```

```
##
## True values are 0 and 3
```

```r
S <- 1e4
approxi <- rnorm(S)
log_w <- - dnorm(approxi, log=T) + dt(approxi, df=3, log=T)
hist(log_w)
```




```r
w <- exp(log_w)

mu <- mean(approxi*w)/mean(w)
var <- mean(approxi^2*w)/mean(w) - mu^2

cat("Estimated EX is:" , mu)
```

```
## Estimated EX is: 0.02595918
```

```r
cat("\nEstimated VarX is:", var)
```

```
##
## Estimated VarX is: 1.504327
```

```r
cat("\nTrue values are 0 and 3")
```

```
##
## True values are 0 and 3
```

```r
w_tilde <- w/sum(w)

S_eff <- 1/sum(w_tilde^2)
S_eff
```

```
## [1] 6968.121
```

For simplicity let's say w is normalized. If random variable $w(X)$ concentrates around small numbers then of course your samples of $w(X)$ are mostly small. This happens when the area where g dominates q are assigned high probability under g. As a result the estimator $S^{-1}\sum_{i=1} w(x_i)f(x_i)$ will be smaller than the true value with high probability.

In our example, t_3 is heavy-tailed comparing to standard normal and as a result the area where t_3 dominates normal, that is, the tail are assigned relatively high probability under t_3. So our estimation is systematically low. (We can see that $w(X)$ concentrates around small number from above histograms.)