HW 12

Jeff Gould

11/19/2020

1) The file HW12_problem1.txt contains 50 iid samples, $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_{50}$, from a one-dimensional random variable X.. Assume that $X \sim \mathcal{N}(\mu, \sigma^2)$ where σ^2 is known with $\sigma^2 = 1$. Our goal is to estimate μ. Let $\bar{x}$ be the sample mean and $N = 50$.

a) Show that the maximum likelihood estimate (MLE) of μ is given by the sample mean $\bar{x}$.

$\ell(\mu) = P(\hat{X}|\mu) = \sum_{i=1}^{N=50} \log P(\hat{X}_i|\mu) = \sum_{i=1}^{N=50} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\hat{X}_i-\mu)^2/(2\sigma^2)}\right)$

$= \sum_{i=1}^{N=50} \log\left(\frac{1}{\sqrt{2\pi}} e^{-(\hat{X}_i-\mu)^2/2}\right) = N\log\left(\frac{1}{\sqrt{2\pi}}\right) - \sum_{i=1}^{N=50} (\hat{X}_i - \mu)^2/2$

$0 = \frac{d\ell(\mu)}{d\mu} = \sum_{i=1}^{N=50} 2(\hat{X}_i - \mu)/2 = \sum_{i=1}^{N=50} (\hat{X}_i - \mu) = \sum_{i=1}^{N=50} \hat{X}_i - N\mu \Rightarrow$

$0 = \sum_{i=1}^{N=50} \hat{X}_i - N\mu \rightarrow N\mu = \sum_{i=1}^{N=50} \hat{X}_i \rightarrow \mu = \frac{1}{N}\sum_{i=1}^{N=50} \hat{X}_i = \bar{x}$

b) Taking a Bayesian approach, assume a normal prior on μ, $\mu \sim \mathcal{N}(0, \beta^2)$ with $\beta = 10$. Let $f(\mu)$ be the pdf of the prior. Let $p(\mu)$ be the posterior. Show $p(\mu) = \frac{1}{Z} P(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_{50} \mid \mu) f(\mu)$

$\mu \sim \mathcal{N}(0, \beta^2) \rightarrow f(\mu) \sim \frac{1}{\sqrt{2\pi\beta^2}} e^{-(\mu^2)/(2\beta^2)}$

$p(\mu) = P(\mu|\hat{X}_1, \hat{X}_2 \ldots \hat{X}_{50}) = \frac{P(\mu, \hat{X}_1, \hat{X}_2 \ldots \hat{X}_{50})}{P(\hat{X}_1, \hat{X}_2 \ldots \hat{X}_{50})} = \frac{P(\hat{X}_1, \hat{X}_2 \ldots \hat{X}_{50}|\mu) f(\mu)}{P(\hat{X}_1, \hat{X}_2 \ldots \hat{X}_{50})}$

each $X_i \sim \mathcal{N}(\mu, 1^2)$, iid. So $P(\hat{X}_1, \hat{X}_2 \ldots \hat{X}_{50}) = \int_{-\infty}^{\infty} P(\hat{X}_1, \hat{X}_2 \ldots \hat{X}_{50}|\mu) f(\mu) d\mu = Z$

$\frac{P(\hat{X}_1, \hat{X}_2 \ldots \hat{X}_{50}|\mu) P(\mu)}{P(\hat{X}_1, \hat{X}_2 \ldots \hat{X}_{50})} = \frac{1}{Z} P(\hat{X}_1, \hat{X}_2 \ldots \hat{X}_{50}|\mu) f(\mu)$

$P(\hat{X}_1, \hat{X}_2 \ldots \hat{X}_{50}|\mu) P(\mu) = \left[\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-(x_i - \mu)^2/(2\sigma^2)\right] \frac{1}{\sqrt{2\pi\beta^2}} \exp[-\mu^2/(2\beta^2)] = \frac{1}{(2\pi)^{(n+1)/2}\sqrt{\beta^2\sigma^{2n}}} \exp[\frac{-\mu^2}{2\beta^2} - \sum_{i=1}^{N} \frac{x_i^2 - 2\mu x_i + \mu^2}{2\sigma^2}] \propto$

$$\exp\left[\frac{-\mu^2(\sigma^2+N\beta^2) + 2\mu(\sum \beta^2 x_i) - \sum(\beta^2 x_i^2)}{2\sigma^2\beta^2}\right] = \exp\left[\frac{-\mu^2 + 2\mu\frac{\sum \beta^2 x_i}{\sigma^2+N\beta^2} - \left(\frac{\sum \beta^2 x_i}{\sigma^2+N\beta^2}\right)^2}{2\frac{\beta^2\sigma^2}{\sigma^2+N\beta^2}}\right] \cdot \exp\left[\frac{-\sum \beta^2 x_i^2}{2\sigma^2\beta^2}\right] \propto$$

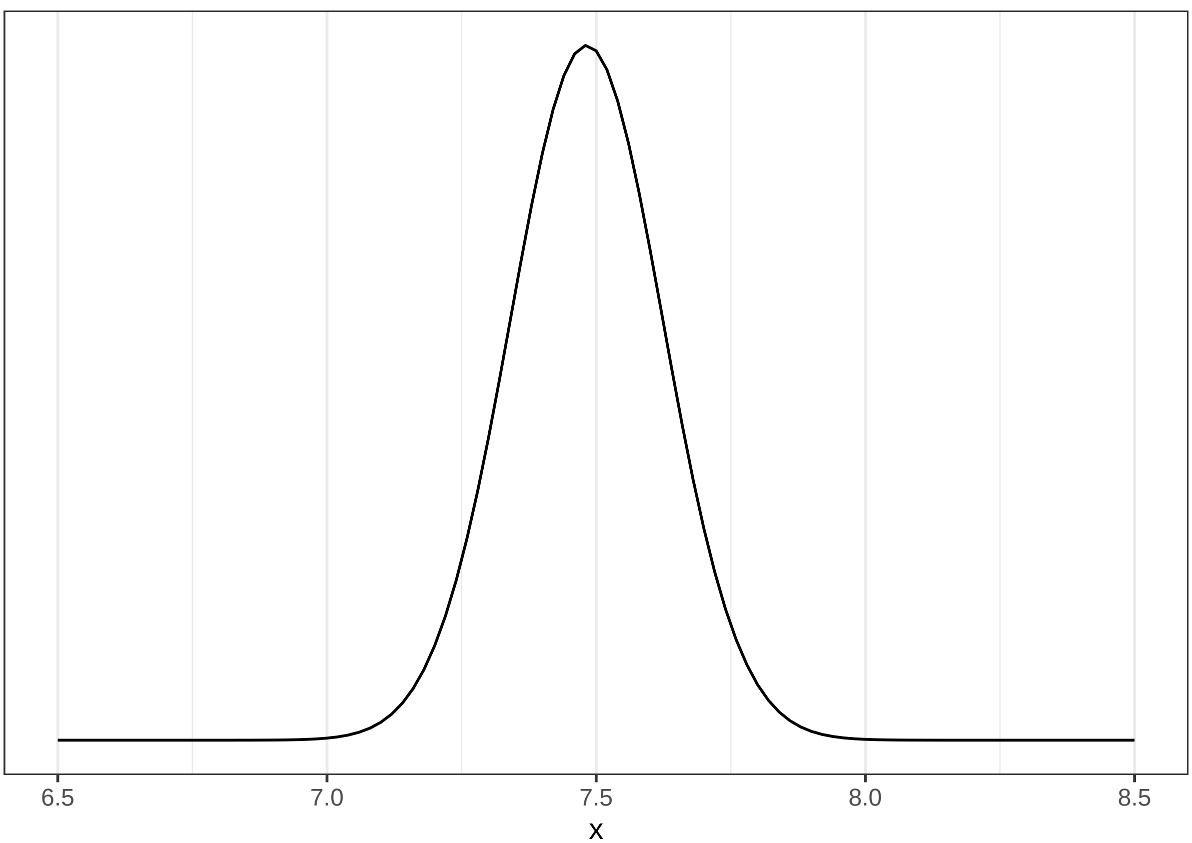
$$\exp\left[-\frac{\left(\mu - \frac{\sum \beta^2 x_i}{\sigma^2+N\beta^2}\right)^2}{2\frac{\beta^2\sigma^2}{\sigma^2+N\beta^2}}\right]$$

$\frac{\beta^2\sum x_i}{\sigma^2+N\beta^2} = \frac{\frac{1}{n}\sum x_i}{\frac{\sigma^2}{N\beta^2}+1} = \frac{\bar{x}}{1+\frac{\sigma^2}{N\beta^2}}$

$$\exp\left[-\frac{\left(\mu - \frac{\sum \beta^2 x_i}{\sigma^2 + N\beta^2}\right)^2}{2\frac{\beta^2\sigma^2}{\sigma^2+N\beta^2}}\right] = \exp\left[\frac{\left(\mu - \frac{\bar{x}}{1+\sigma^2/(N\beta^2)}\right)^2}{2\frac{\beta^2\sigma^2}{\sigma^2+N\beta^2}}\right] \Rightarrow$$

$p(\mu) \sim \mathcal{N}(\frac{\bar{x}}{1+\sigma^2/(N\beta^2)}, \frac{\beta^2\sigma^2}{\sigma^2+N\beta^2})$

```r
X <- read.delim("HW12_problem1.txt")
N <- nrow(X)
sigma <- 1
beta <- 10
x_bar <- mean(X$x)

mu_post <- x_bar / (1 + (sigma / (beta * N)))
var_post <- (sigma * beta) / (N * beta + sigma)

ggplot(data = data.frame(x = c(6.5, 8.5)), aes(x)) +
  stat_function(fun = dnorm,  args = list(mean = mu_post, sd = sqrt(var_post))) +
  scale_y_continuous(breaks = NULL) +
  theme_bw() +
  labs(y = "")
```



c)

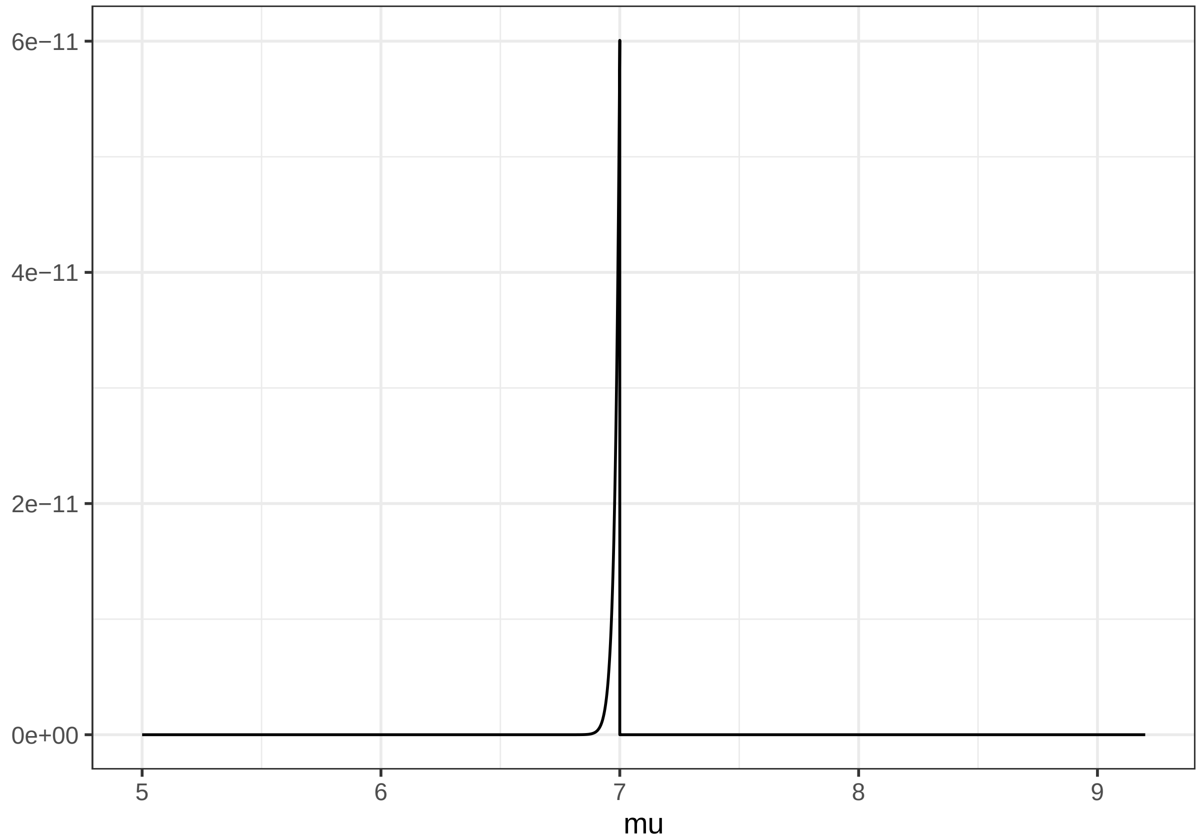
$$f(\mu) = \begin{cases} \frac{1}{10} & \text{if } x \in [5,7] \\ 4 & \text{if } x \in [9, 9.2] \\ 0 & \text{otherwise} \end{cases}$$

$\frac{P(X_1,\ldots,X_{50}|\mu)P(\mu)}{P(X_1,\ldots,X_{50})}$

$Z = P(X_1,\ldots,X_{50}) = \int_5^7 P(X_1,\ldots,X_{50}|\mu) f(\mu) d\mu + \int_9^{9.2} P(X_1,\ldots,X_{50}|\mu) f(\mu) d\mu = \int_5^7 P(X_1,\ldots,X_{50}|\mu) \frac{1}{10} d\mu + \int_9^{9.2} P(X_1,\ldots,X_{50}|\mu)(4) d\mu = \int_5^7 \prod_{i=1}^{50} \frac{1}{\sqrt{2\pi}} \exp(-(x_i-\mu)^2/2) \frac{1}{10} d\mu + \int_9^{9.2} \prod_{i=1}^{50} \frac{1}{\sqrt{2\pi}} \exp(-(x_i-\mu)^2/2)(4) d\mu$

$\frac{P(X_1,\ldots,X_{50}|\mu)P(\mu)}{P(X_1,\ldots,X_{50})} = \frac{\left(\prod_{i=1}^{50} \frac{1}{\sqrt{2\pi}} \exp(-(x_i-\mu)^2/2)\right) f(\mu)}{Z}$

```
x <- X$x
integrand <- function(mu){
  return(prod(1 / sqrt(2 * pi) * exp(-(x - mu)^2 / 2)))
}
#integrand(5)

i1 <- integrate(Vectorize(integrand), lower = 5, upper = 7)
i2 <- integrate(Vectorize(integrand), lower = 9, upper = 9.2)
Z <- (1/(10)) * i1$value + 4 * i2$value

mu_vec <- seq(5, 7, 0.001)
fff <- function(mu){(1 / (2*pi)^25) * exp(-sum((x - mu)^2)) / 2 * 1/10}
mu_dens1 <- sapply(mu_vec, fff) / Z

mu_vec <- seq(9, 9.2, 0.001)
fff <- function(mu){(1 / (2*pi)^25) * exp(-sum((x - mu)^2)) / 2 * 4}
mu_dens2 <- sapply(mu_vec, fff) / Z

pd <- data.frame(mu = c(seq(5, 7, 0.001),7.000001,8.999999, seq(9, 9.2, 0.001)),
                 probs = c(mu_dens1,0,0, mu_dens2))

ggplot(pd, aes(x = mu, y = probs)) + geom_line() + theme_bw() +
  #scale_y_continuous(breaks = NULL) +
  theme_bw() +
  labs(y = "")
```




```
# innn <- function(mu){1/(sqrt(2*pi)) * exp(-(a - mu)^2 / 2) }
#
# P <- 1
# for(i in 1:50){
#   a <- x[i]
#   int1 <- integrate(innn, 5, 7)
#   int2 <- integrate(innn, 9, 9.2)
#   pr <- 0.1 * int1$value + 4 * int2$value
#
#   P <- P * pr
#
# }
# P / Z
# integrate(innn, 5, 7)
#
# density <- function(mu){
#   p <- 1
#   for(q in c(0.1, 4))
#   num <- (1/10 * as.numeric(x >= 5 & x <= 7) + 4 * as.numeric(x >= 9 & x <= 9.2)) * exp(-(x - mu)^2 /
#   return(p / Z)
# }
# density(7)
#
# out <- sapply(seq(6, 10, 0.001), density)
#
# ggplot(data = data.frame(x = c(6.5, 8.5)), aes(x)) +
```

```
#   stat_function(fun = density) +
#   scale_y_continuous(breaks = NULL) +
#   theme_bw() +
#   labs(y = "")
#
#
```

Using a uniform prior, especially one that has bounds so that it takes a 0 value in the range of the data, does not seem useful. Because $posterior \propto likelihood \times prior$, but the $prior$ is just a constant under a uniform prior, or in this case a piecewise uniform prior. So if the $prior$ is a constant, then $posterior \propto likelihood$. We also we have zero mass between $(7, 9)$, even though thats where a lot of the data lives

2

a)

$P(X = x, Y = y, Z = z) = \alpha \exp[\eta_1 x + \eta_2 y + \eta_3 z - w_{12}xy - w_{13}xz]$

The clicks are (X, Y) and (X, Z), since we do not have an edge between Y and Z

From the lectures, in general, $w_{ij} = 0 \iff$ no edge exists between x_i, x_j and then x_i, x_j are conditionally independent given neighbors/all other nodes. So for this example, since $w_{13} = 0$, then given X, Y and Z are conditionall independent.

More rigorously:

For conditional independence, we need:

$(A \perp B)|C \iff P(A \cap B|C) = P(A|C)P(B|C) \iff P(A|B \cap C) = P(A|C)$

here we show $P(A|B \cap C) = P(A|C)$ for Y and Z, given X, thus showing their conditional independence

Let $\psi_1(x, y) = \exp[\frac{\eta_1}{2}x + \eta_2 y - w_{12}xy]$, and let $\psi_2(x, z) = \exp[\frac{\eta_1}{2}x + \eta_3 z - w_{13}xz]$

Then $\alpha\psi_1(x, y)\psi_2(x, z) = \alpha \exp[\frac{\eta_1}{2}x + \eta_2 y - w_{12}xy] \exp[\frac{\eta_1}{2}x + \eta_3 z - w_{13}xz] = \alpha \exp[\eta_1 x + \eta_2 y + \eta_3 z - w_{12}xy - w_{13}xz] = P(X = x, Y = y, Z = z)$

$Y \perp Z|X$:

$P(Y|X = x_j) = \alpha \sum_{z \in \{-1,1\}} \psi_1(Y, x_j)\psi_2(x_j, z) = \alpha\psi_1(Y, x_j) \sum_{z \in \{-1,1\}} \psi_2(x_j, z) \Rightarrow$

$P(Y = y_i|X = x) = \frac{\psi_1(x, y_i) \sum_{z \in \{-1,1\}} \psi_2(x, z)}{(\psi_1(x, y_i) + \psi_1(x, y_j)) \sum_{z \in \{-1,1\}} \psi_2(x, z)} = \frac{\psi_1(x, y_i)}{\psi_1(x, y_i) + \psi_1(x, y_j)}$

This is free of Z, so clearly Y is not dependent on Z when given X

$P(Z = z_k|X = x_j) = \alpha \sum_{y \in \{-1,1\}} \psi_1(y, x_j)\psi_2(x_j, z_k) = \alpha\psi_2(x_j, z_k) \sum_{y \in \{-1,1\}} \psi_1(x_j, y)$

So likewise,

$P(Z = y_i|X = x) = \frac{\psi_2(x, z_i) \sum_{y \in \{-1,1\}} \psi_1(x, y)}{(\psi_2(x, z_i) + \psi_2(x, z_j)) \sum_{y \in \{-1,1\}} \psi_1(x, y)} = \frac{\psi_2(x, z_i)}{\psi_2(x, z_i) + \psi_2(x, z_j)}$

And Z is not dependent on Y when given X

So $Y \perp Z|X$

b)

Using the same principle as above, since we have $w_{23} \neq 0$, then we know an edge must exist between Y and Z, and thus they are not conditionally indepenedent

$P(X = x, Y = y, Z = z) = \alpha \exp[\eta_1 x + \eta_2 y + \eta_3 z - w_{12}xy - w_{13}xz - w_{23}yz]$

$P(Y = y_i|X) = \alpha \sum_{z\in\{-1,1\}} \exp[\eta_1 x + \eta_2 y_i + \eta_3 z - w_{12}xy_i - w_{13}xz - w_{23}y_i z] = \alpha \sum_{z\in\{-1,1\}} \exp[\eta_2 y_i - w_{12}xy_i - w_{23}y_i z] \exp[\eta_1 x + \eta_3 z - w_{13}xz]$

And we clearly can't get a Y term free of Z due to $w_{23}yz$, thus Y will not be conditionally independent of Z when given X. The equality of $P(Y|X \cap Z) = P(Y|X)$ will not hold, thus no conditional independence

Similarly for Z:

$P(Z = z_i|X) = \alpha \sum_{y\in\{-1,1\}} \exp[\eta_1 x + \eta_2 y + \eta_3 z_1 - w_{12}xy - w_{13}xz_i - w_{23}yz_i] = \alpha \sum_{y\in\{-1,1\}} \exp[\eta_3 z_i - w_{13}xz_i - w_{23}yz_i] \exp[\eta_1 x + \eta_2 y - w_{12}xy]$

And we can't get Z free of Y

So even with X, Y and Z are not conditionally independent

$P(X = x, Y = y, Z = z) = \alpha \exp[\eta_1 x + \eta_2 y + \eta_3 z - w_{12}xy - w_{13}xz - w_{23}yz - w_{123}xyz]$

$P(Y = y_i|X) = \alpha \sum_{z\in\{-1,1\}} \exp[\eta_1 x + \eta_2 y_i + \eta_3 z - w_{12}xy_i - w_{13}xz - w_{23}y_i z - w_{123}xy_i z] = \alpha \sum_{z\in\{-1,1\}} \exp[\eta_2 y_i - w_{12}xy_i - w_{23}y_i z - w_{123}xy_i z] \exp[\eta_1 x + \eta_3 z - w_{13}xz]$

$P(Z = z_i|X) = \alpha \sum_{y\in\{-1,1\}} \exp[\eta_1 x + \eta_2 y + \eta_3 z_i - w_{12}xy - w_{13}xz_i - w_{23}yz_i - w_{123}xyz_i] = \alpha \sum_{y\in\{-1,1\}} \exp[\eta_3 z_i - w_{13}xz_i - w_{23}yz_i - w_{123}xyz_i] \exp[\eta_1 x + \eta_2 y - w_{12}xy]$

c)

$R(\omega', \omega) = R(\omega, \omega') = 1/3$ for all ω, ω', since there is a 1/3 chance of picking the coordinate to flip between eah state. So the MH ratio becomes:

$\frac{P(\omega')}{P(\omega)} = \frac{\alpha \exp(\eta_1 x' + \eta_2 y' + \eta_3 z' - w_{12}x'y' - w_{13}x'z')}{\alpha \exp(\eta_1 x + \eta_2 y + \eta_3 z - w_{12}xy - w_{13}xz)} = \frac{\exp(\eta_1 x' + \eta_2 y' + \eta_3 z' - w_{12}x'y' - w_{13}x'z')}{\exp(\eta_1 x + \eta_2 y + \eta_3 z - w_{12}xy - w_{13}xz)} = \frac{\exp(1/2(x'+y'+z') - x'y' + x'z')}{\exp(1/2(x+y+z) - xy + xz)}$

```r
eta <- 0.5
w_12 <- 1
w_13 <- -1


MH_prob <- function(SS){
  x <- SS[,"X"]
  y <- SS[,"Y"]
  z <- SS[,"Z"]
  exp(eta * x + eta * y + eta * z - w_12 * x * y - w_13 * x * z)
  }

stateSpace <- matrix(data = 1, ncol = 3, dimnames = list(c(), c("X", "Y", "Z")))

SShistory <- stateSpace

set.seed(123)
tictoc::tic()
for (i in 1:2500000) {

  flip <- sample(x = c("X", "Y", "Z"), 1)
```

```
  newStateSpace = stateSpace
  newStateSpace[,flip] = -newStateSpace[,flip]



  u <- runif(1)

  if(u < min(1, MH_prob(newStateSpace) / MH_prob(stateSpace))){
    stateSpace <- newStateSpace
  }

  if(i %% 1000 == 0 & i >= 10000){
    #print(i)
    SShistory <- rbind(SShistory, stateSpace)
  }


}
tictoc::toc()
```

```
## 44.287 sec elapsed
```

```
SShistory <- SShistory[-1, ]
cor(SShistory[,"X"], SShistory[,"Y"])
```

```
## [1] -0.6135023
```