(a) Show that the MLE of μ satisfies

$$\cos(\widehat{\mu})\sum_{i=1}^{n}\sin(D_i) - \sin(\widehat{\mu})\sum_{i=1}^{n}\cos(D_i) = 0$$

and give an explicit formula for $\widehat{\mu}$ in terms of $\sum_{i=1}^{n}\sin(D_i)$ and $\sum_{i=1}^{n}\cos(D_i)$. (Hint: Use the formula $\sin(x-y) = \sin(x)\cos(y) - \cos(x)\sin(y)$. Also verify that your estimator actually maximizes the likelihood function.)

We have the PDF so we can compute MLE estimate.

First, $P(D_i : k, \mu) = \prod_{i=1}^{n} \frac{1}{2\pi I_0(k)} \exp(k\cos(D_i - \mu))$

Then, $L(D_i, k, \mu) = \ln P(D_i : k, \mu)$

$$= \sum_{i=1}^{n} -\ln(2\pi I_0(k)) + k\cos(D_i - \mu)$$

Derive WRT μ and equal to 0

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{n} \sin(D_i - \mu) = 0$$

Note that since $\sin(x-y) = -\sin(y-x)$

Then, $\sum_{i=1}^{n} \sin(D_i - \mu) = -\sum_{i=1}^{n} \sin(\mu - D_i)$

$$= \sum_{i=1}^{n} [\cos(\mu)\sin(D_i) - \sin(\mu)\cos(D_i)]$$

$$= \cos(\mu)\sum_{i=1}^{n}\sin(D_i) - \sin(\mu)\sum_{i=1}^{n}\cos(D_i)$$

$$= 0$$

Then for MLE $\hat{\mu}$ satisfies

$$\cos(\hat{\mu})\sum_{i=1}^{n}\sin(D_i) - \sin(\hat{\mu})\sum_{i=1}^{n}\cos(D_i) = 0$$

Since $\cos(\hat{\mu})\sum_{i=1}^{n}\sin(D_i) - \sin(\hat{\mu})\sum_{i=1}^{n}\cos(D_i) = 0$

Then $\cos(\hat{\mu})\sum_{i=1}^{n}\sin(D_i) = \sin(\hat{\mu})\sum_{i=1}^{n}\cos(D_i)$

So, $\frac{\sin(\hat{\mu})}{\cos(\hat{\mu})} = \frac{\sum_{i=1}^{n}\cos(D_i)}{\sum_{i=1}^{n}\sin(D_i)}$

$$\tan(\hat{\mu}) = \frac{\sum_{i=1}^{n}\cos(D_i)}{\sum_{i=1}^{n}\sin(D_i)}$$

Then $\hat{\mu} = \tan^{-1}\left(\frac{\sum_{i=1}^{n}\cos(D_i)}{\sum_{i=1}^{n}\sin(D_i)}\right)$

(b) Suppose that we put the following prior density on (κ, μ):

$$\pi(\kappa, \mu) = \frac{\lambda}{2\pi} \exp(-\lambda\kappa) \quad \text{for } \kappa > 0 \text{ and } 0 \le \mu < 2\pi.$$

Show that the posterior (marginal) density of κ is

$$\pi(\kappa|d_1, \cdots, d_n) = c(d_1, \cdots, d_n) \frac{\exp(-\lambda\kappa) I_0(r\kappa)}{[I_0(\kappa)]^n}$$

where

$$r = \left\{ \left(\sum_{i=1}^n \cos(d_i) \right)^2 + \left(\sum_{i=1}^n \sin(d_i) \right)^2 \right\}^{1/2}$$

(Hint: To get the marginal posterior density of κ, you need to integrate the joint posterior over μ (from 0 to 2π). The trick is to write

$$\sum_{i=1}^n \cos(d_i - \mu) = r \cos(\theta - \mu)$$

for some θ.)

Given : $f(\theta ; k, \mu) = \frac{1}{2\pi I_0(k)} \exp(k \cos(\theta - \mu))$

$I_0(k) = \frac{1}{2\pi} \int_0^{2\pi} \exp(k \cos(\theta)) \, d\theta$

The posterior is the following:

$$\pi(k, \mu | d_1, \cdots, d_n) = c(d_1, \cdots, d_n) \, \pi(k, \mu) \, L(d_1, \cdots, d_n)$$

$$= c(d_1, \cdots, d_n) \frac{\lambda}{2\pi} \exp(-\lambda k) \prod_{i=1}^n \frac{\exp(k \cos(d_i - \mu))}{(2\pi) \, I_0(k)}$$

Since $L(d_1, \cdots, d_n) = \prod_{i=1}^n \frac{\exp(k \cos(d_i - \mu))}{(2\pi) \, I_0(k)}$

$$= \frac{\exp(k \sum_{i=1}^n \cos(d_i - \mu))}{(2\pi^n) \, I_0(k)^n}$$

$$= \frac{\exp(k r \cos(\theta - \mu))}{(2\pi)^n \, I_0(k)^n}$$ for some θ, $r = \{(\sum_{i=1}^n \cos(d_i))^2 + (\sum_{i=1}^n \sin(d_i))^2\}^{1/2}$

Then, simplifying posterior:

$$\pi(k, \mu | d_1, \cdots, d_n) = c(d_1, \cdots, d_n) \frac{\lambda}{2\pi} \exp(-\lambda k) \frac{\exp(k r \cos(\theta - \mu))}{(2\pi)^n \, I_0(k)^n}$$

$$= \frac{c(d_1, \cdots, d_n)}{(2\pi)^n} \frac{\exp(-\lambda k)}{I_0(k)^n} \frac{\lambda}{2\pi} \exp(k r \cos(\theta - \mu))$$

$$= c(d_1, \cdots, d_n) \frac{\exp(-\lambda k)}{I_0(k)^n} \frac{\lambda}{2\pi} \exp(k r \cos(\theta - \mu))$$ since $(2\pi)^n$ is just a constant

Note: Since $I_0(k) = \frac{1}{2\pi} \int_0^{2\pi} \exp(k \cos(\theta)) \, d\theta$

Integrate posterior wrt μ:

$$\pi(k | d_1, \cdots, d_n) = \int_0^{2\pi} c(d_1, \cdots, d_n) \frac{\exp(-\lambda k)}{I_0(k)^n} \frac{\lambda}{2\pi} \exp(k r \cos(\theta - \mu)) \, d\mu$$

$$= c(d_1, \cdots, d_n) \frac{\exp(-\lambda k)}{I_0(k)^n} \frac{\lambda}{2\pi} \int_0^{2\pi} \frac{\lambda}{2\pi} \exp(k r \cos(\theta - \mu)) \, d\mu$$

using substitution $x = \theta - \mu \Rightarrow \frac{dx}{d\mu} = -1$

$$dx = -d\mu$$

$$\int_0^{2\pi} \frac{\lambda}{2\pi} \exp(k r \cos(\theta-\mu))\, d\mu = -\int_{\theta}^{\theta-2\pi} \exp(k r \cos(x))\, dx$$

$$= \int_{\theta-2\pi}^{\theta} \exp(k r \cos(x))\, dx$$

Then, $\pi(k \mid d_1, \cdots, d_n) = c(d_1, \cdots, d_n) \frac{\exp(-\lambda k)}{I_0(k)^n} \frac{\lambda}{2\pi} \int_{\theta-2\pi}^{\theta} \exp(k r \cos(x))\, dx$

Since we know $\exp(\cos(x))$ is periodic then,

$$\int_{\theta-2\pi}^{\theta} \exp(k \cos(x))\, dx = \int_0^{2\pi} \exp(k \cos(x))\, dx$$

So, $\pi(k \mid d_1, \cdots, d_n) = c(d_1, \cdots, d_n) \frac{\exp(-\lambda k)}{I_0(k)^n} \frac{\lambda}{2\pi} \int_0^{2\pi} \exp(k r \cos(x))\, dx$

$$= \lambda c(d_1, \cdots, d_n) \frac{\exp(-\lambda k)}{I_0(k)^n} I_0(rk)$$

$$= C(d_1, \cdots, d_n) \frac{\exp(-\lambda k)\, I_0(rk)}{[I_0(k)]^n}$$

λ is just constant

■

2) (a) Suppose we have data $(x_1, y_1), \cdots, (x_n, y_n)$ and we define

$$z_i = a + bx_i + y_i \quad (i = 1, \cdots, n)$$

for some constants a and b. Suppose that $\widehat{\beta}_0$ and $\widehat{\beta}_1$ minimize

$$\text{median}\{(y_i - \beta_0 - \beta_1 x_i)^2 : i = 1, \cdots, n\}$$

and $\widetilde{\beta}_0$ and $\widetilde{\beta}_1$ minimize

$$\text{median}\{(z_i - \beta_0 - \beta_1 x_i)^2 : i = 1, \cdots, n\}.$$

What is the relationship between $(\widehat{\beta}_0, \widehat{\beta}_1)$ and $(\widetilde{\beta}_0, \widetilde{\beta}_1)$?

Consider this :

Let $\hat{\beta}_0$ and $\hat{\beta}_1$ minimize median$\{(y_i - \beta_0 - \beta_1 x_i)^2 : i = 1, \cdots, n\}$

Now since $z_i = a + bx_i + y_i$

we can rewrite the following

$$\text{median}\{(z_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i)^2 : i = 1, \cdots, n\}$$ $\tilde{\beta}_0, \tilde{\beta}_1$ minimizes this

$$= \text{median}\{(a + bx_i + y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i)^2 : i = 1, \cdots, n\}$$

$$= \text{median}\{(y_i - (\tilde{\beta}_0 - a) x_i - (\tilde{\beta}_1 - b) x_i)^2 : i = 1, \cdots, n\}$$

This looks like the first median equation

We know $\hat{\beta}_0$ and $\hat{\beta}_1$ minimizes this

So $\hat{\beta}_0 = \tilde{\beta}_0 - a$

$\hat{\beta}_1 = \tilde{\beta}_1 - b$

Then the relationship b/w $(\hat{\beta}_0, \hat{\beta}_1)$ and $(\tilde{\beta}_0, \tilde{\beta}_1)$

is just constant (a, b) apart. ■

(b) Show that if $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ for $i = 1, \cdots, n$ then the bias and variance of the LMS estimators does not depend on β_0 and β_1; in other words, they depend only on $\{x_i\}$ and $\{\varepsilon_i\}$. (Hint: Use the result of part (a).)

Assume that $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

From part a) we saw that $Z_i = a + bX_i + Y_i$ for a, b constant.

Then Z is just some $a + bX_i$ deviation from Y_i

Take $\hat{\beta}$ and $\tilde{\beta}$ as defined in part a).

It's sufficient enough to show $Bias(\hat{\beta}) = Bias(\tilde{\beta})$

Since $Bias(\hat{\beta}) = Bias(\tilde{\beta} - a)$

$= Bias(\tilde{\beta})$ since a is constant

Here we are computing for two different estimators

$\text{median}\{(Y_i - \beta_0 - \beta_1 X_i)^2\}$

$\text{median}\{(Z_i - \beta_0 - \beta_1 X_i)^2\}$

However, we have $Bias(\tilde{\beta}) = Bias(\hat{\beta})$ this means that two different estimators do not depend on true values but only $\{X_i\}$, and $\{e_i\}$.

■

(c) Consider the following model:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (i = 1, \cdots, n)$$

where $x_i = i/n$ and $\{\varepsilon_i\}$ are independent $\mathcal{N}(0,1)$ random variables. For least squares estimation, the variance of $\widehat{\beta}_1$ tends to 0 like constant$/n$ as $n \to \infty$. For LMS estimation, $\text{Var}(\widehat{\beta}_1) \approx \gamma/n^\alpha$ for some $\gamma > 0$ and $\alpha > 0$. In lecture, we claimed that $\alpha = 2/3$. The theoretical proof of this is very technical; however, it is possible to estimate α via simulation.

Using linear regression to fit $\ln(Var(\beta_1)) = \ln(\gamma) - \alpha \ln(n) + \text{noise}$

with $n = (50, 100, 500, 1000, 5000)$

We get $\hat{\alpha} = 0.6911$

This is approximately $2/3$

R code attached.

(d) Repeat part (c) using Cauchy errors. (For Cauchy errors, the variance of the least squares estimator does not tend to 0 as $n \to \infty$.) Do you get a similar value of α?

Using cauchy errors, the estimated value of $\alpha = 0.7272$

This value is similar to α estimated using normal errors.

However, it is a little bit different from the theoretical value of $\alpha = 2/3$

STA355 A3

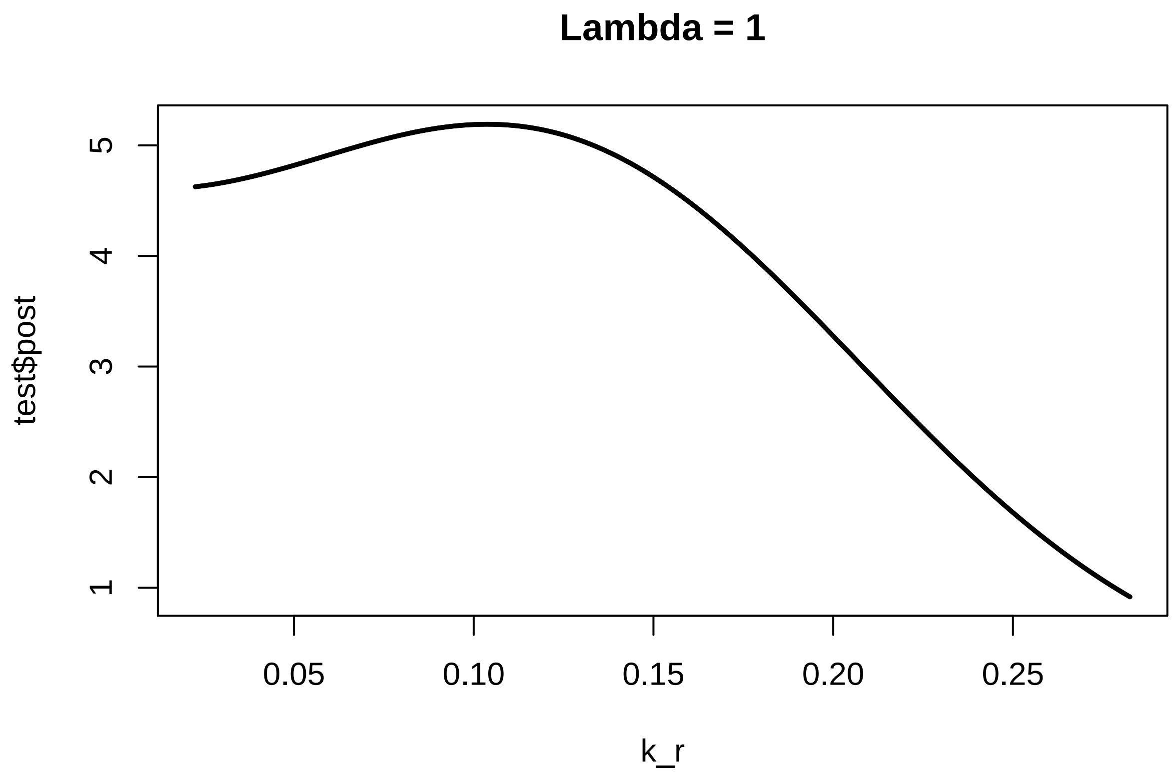
Harold Hyun Woo Lee

3/9/2021

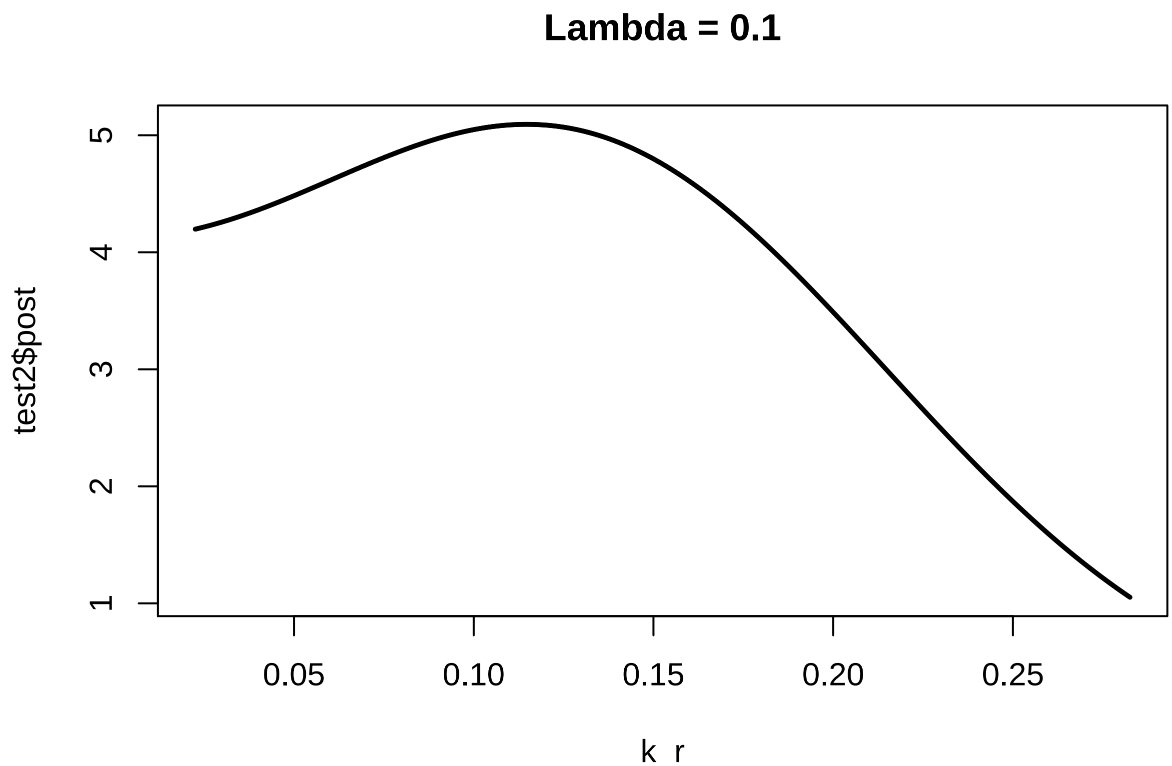
Q1c)

```
dat <- scan("bees.txt")

# Convert the data from degrees to radians
dat <- dat*pi/180
n <- length(dat)

r <- ((sum(cos(dat)))^2+(sum(sin(dat)))^2)^(1/2)
k_hat <- ((r/n)*(2-(r^2/n^2)))/(1-(r^2/n^2))

# Helper function to get valid range of K
k_range <- function(x, N, delta = c(0.05, 0.2)){

  # Calculate h
  h <- diff(delta)/N

  # Get the sequence
  thetas <- seq(from = delta[1], to = delta[2], by = h)

  return(list("h" = h, "theta" = thetas))
}

k_range2 <- function(x, N, del){

  # Calculate h
  h <- del*2/N

  # Get the sequence
  thetas <- seq(from = k_hat - del, to = k_hat + del, by = h)

  return(list("h" = h, "theta" = thetas))
}






U <- function(k, lambda, r, n){
```

```
  before <- (exp(-lambda*k)*besselI(r*k, 0))/(besselI(k,0))^n
  prenorm <- before/max(before)

  prenorm

}

denom <- function(x, lambda, k, N, h){
    n <- length(x)
    r <- ((sum(cos(x)))^2+(sum(sin(x)))^2)^(1/2)

    post <- U(k, lambda, r, n)

    mult <- c(1/2,rep(1,N-1),1/2) # multipliers for trapezoidal rule
    norm <- sum(mult*post)/(1/h) # integral evaluated using trapezoidal rule
    post <- post/norm # normalized posterior
    post.cdf <- cumsum(mult*post)/(1/h) # compute the posterior cdf
    lower <- max(k[post.cdf<0.025])
    upper <- min(k[post.cdf<0.975])

    return(list("post" = post, "post.cdf" = post.cdf,
                "lower" = lower, "upper" = upper))

}


k_r <- k_range2(dat, N=1000, 0.13)$theta
h <- k_range2(dat, N=1000, 0.13)$h

test <- denom(dat, lambda = 1, k_r, N = 1000, h)

plot(k_r, test$post, main="Lambda = 1", cex=0.2)
```




```
test2 <- denom(dat, lambda = 0.1, k_r, N = 1000, h)

plot(k_r, test2$post, main="Lambda = 0.1", cex=0.2)
```



The posterior densities for two lambda values are shown above in the plots. The difference between the two lambdas is that for lambda = 1, the tail on the left side is slightly bigger than lambda = 0.1 Otherwise, they are very much similar to each other. As for the possibility that k = 0, I believe we can't rule it out. If you look at the posterior density graph, at k = 0 we see the posterior density is almost 5. If the probability was 0 then posterior density wouldn't be so high.

Q2c)

```
library(MASS)

set.seed(1)
var_norm <- function(n){

  # M is 500
  nrep <- 500

  # Declaring x values
  x <- c(1:n)/n

  beta <- NULL

  for (i in 1:nrep) {
    y <- rnorm(n)
    r <- lmsreg(y~x)
    beta <- c(beta,r$coef[2])
```

```r
    }
  var <- var(beta)
  return(var)
}

B1_norm_var <- c(var_norm(50), var_norm(100), var_norm(500), var_norm(1000), var_norm(5000))

log_B1_norm_var <- log(B1_norm_var)

x <- c(log(50), log(100), log(500), log(1000), log(5000))

fit <- lm(log_B1_norm_var~x)
alpha = fit$coefficients[2]
-alpha
```

```
##         x
## 0.6911227
```

Q2d)

```r
set.seed(1)
var_cauchy <- function(n){

  # M is 500
  nrep <- 500

  # Declaring x values
  x <- c(1:n)/n

  beta <- NULL

  for (i in 1:nrep) {
    y <- rcauchy(n)
    r <- lmsreg(y~x)
    beta <- c(beta,r$coef[2])
    }
  var <- var(beta)
  return(var)
}

B1_cau_var <- c(var_cauchy(50), var_cauchy(100), var_cauchy(500), var_cauchy(1000), var_cauchy(5000))

log_B1_cau_var <- log(B1_cau_var)

x <- c(log(50), log(100), log(500), log(1000), log(5000))

fit2 <- lm(log_B1_cau_var~x)
alpha2 = fit2$coefficients[2]
-alpha2
```

```
##         x
## 0.7272641
```