Problem 46

Overview

Chromatin folding is process in biology. We assume that chromatin polymers follow a random walk model. In this model, the two-dimensional distance between two chromatin fibers, R, follows a Raleigh distribution, with density

$$f(r|\theta) = \frac{r}{\theta^2} e^{\frac{-r^2}{2\theta^2}}$$

where $r \geq 0$ and $\theta > 0$.

Estimating Theta

Maximum Likelihood

Derivation of the Estimator

The likelihood function is

$$lik(\theta) = \frac{1}{\theta^{2n}} e^{-\frac{1}{2\theta^2}\sum_{i=1}^{n} r_i^2} \prod_{i=1}^{n} r_i$$

The log likelihood is thus

$$l(\theta) = -2n\ln(\theta) + \sum_{i=1}^{n} \ln(r_i) - \frac{1}{2\theta^2} \sum_{i=1}^{n} r_i^2$$

The first-order condition for maximizing the likelihood satisfies

$$0 = \frac{-2n}{\hat{\theta}} + \frac{1}{\hat{\theta}^3} (\sum_{i=1}^{n} r_i^2)$$

Which, after rearranging, gives the maximum likelihood estimator for θ, $\hat{\theta}_{MLE}$

$$\hat{\theta}_{MLE} = \frac{1}{\sqrt{2}} \sqrt{\frac{1}{n} \sum_{i=1}^{n} r_i^2}$$

Asymptotic Variance

As usual, the asymptotic variance of a maximum likelihood estimator is roughly

$$Var(\hat{\theta}_{MLE}) \approx \frac{1}{nI(\theta)}$$

where under sufficient smoothness conditions that at the level of this textbook we will assume hold,

$$I(\theta) = E\left[\frac{\partial}{\partial\theta}\log(f(x|\theta))\right]^2 = -E\left[\frac{\partial^2}{\partial\theta^2}\log(f(x|\theta))\right]$$

$f(r|\theta) = \frac{r}{\theta^2}e^{\frac{-r^2}{2\theta^2}}$, so

$$\begin{aligned}
\frac{\partial}{\partial\theta}\log f(r|\theta) &= -\frac{2}{\theta} + \frac{r^2}{\theta^3}, \\
\frac{\partial^2}{\partial\theta^2}\log f(r|\theta) &= \frac{2}{\theta^2} - 3\frac{r^2}{\theta^4}, \\
I(\theta) &= -\frac{2}{\theta^2} + 3\frac{E(R^2)}{\theta^4}
\end{aligned}$$

The expectation of R^2 is

$$E(R^2) = \frac{1}{\theta^2}\int_0^\infty r^3 e^{\frac{-r^2}{2\theta^2}}\,dr = 4\theta^2\int_0^\infty x^3e^{-x^2}\,dx = 2\theta^2\int_0^\infty ue^{-u}du$$

by making the substitutions $x = \frac{r}{\sqrt{2}\theta}$ and $u = x^2$. Integrating by parts,

$$E(R^2) = 2\theta^2\left[-ue^{-u}|_0^\infty + \int_0^\infty e^{-u}du\right] = -2\theta^2(e^{-u})|_0^\infty = 2\theta^2$$

Thus

$$\begin{aligned}
I(\theta) &= -\frac{4}{\theta^2}, \\
Var(\hat{\theta}_{MLE}) &\to \frac{\theta^2}{4n}
\end{aligned}$$

Method of Moments

Derivation of the Estimator

The expectation of R is

$$E(R) = \frac{1}{\theta^2}\int_0^\infty r^2e^{\frac{-r^2}{2\theta^2}}\,dr = (2\sqrt{2})\theta\int_0^\infty x^2e^{-x^2}dx$$

by making the substitution $x = \frac{r}{\sqrt{2}\theta}$. Integrating by parts,

$$E(R) = (2\sqrt{2})\theta\left[\frac{1}{2}xe^{-x^2}|_0^\infty + \frac{1}{2}\int_0^\infty e^{-x^2}dx\right] = \sqrt{2}\theta\int_0^\infty e^{-x^2}dx = \theta\sqrt{\frac{\pi}{2}}$$

since $\int_0^\infty e^{-x^2} dx = \sqrt{\pi}$, and the integrand is an even function. Thus

$$\hat{\theta}_{MoM} = \bar{X}\sqrt{\frac{2}{\pi}}$$

Asymptotic Variance

By the Central Limit Theorem, $\bar{X}$ converges to a normal random variable with mean $E(R)$ and variance $\frac{Var(R)}{n}$ as n approaches infinity. From previous results.

$$Var(R) = E(R^2) - E(R)^2 = 2\theta^2 - \frac{\pi}{2}\theta^2 = \frac{4-\pi}{2}\theta^2$$

and the asymptotic variance of the method of moments estimator for θ is

$$Var(\hat{\theta}_{MoM}) \to \left(\frac{4}{\pi} - 1\right)\frac{\theta^2}{n}$$

in which $\frac{4}{\pi} - 1 \approx .273295...$

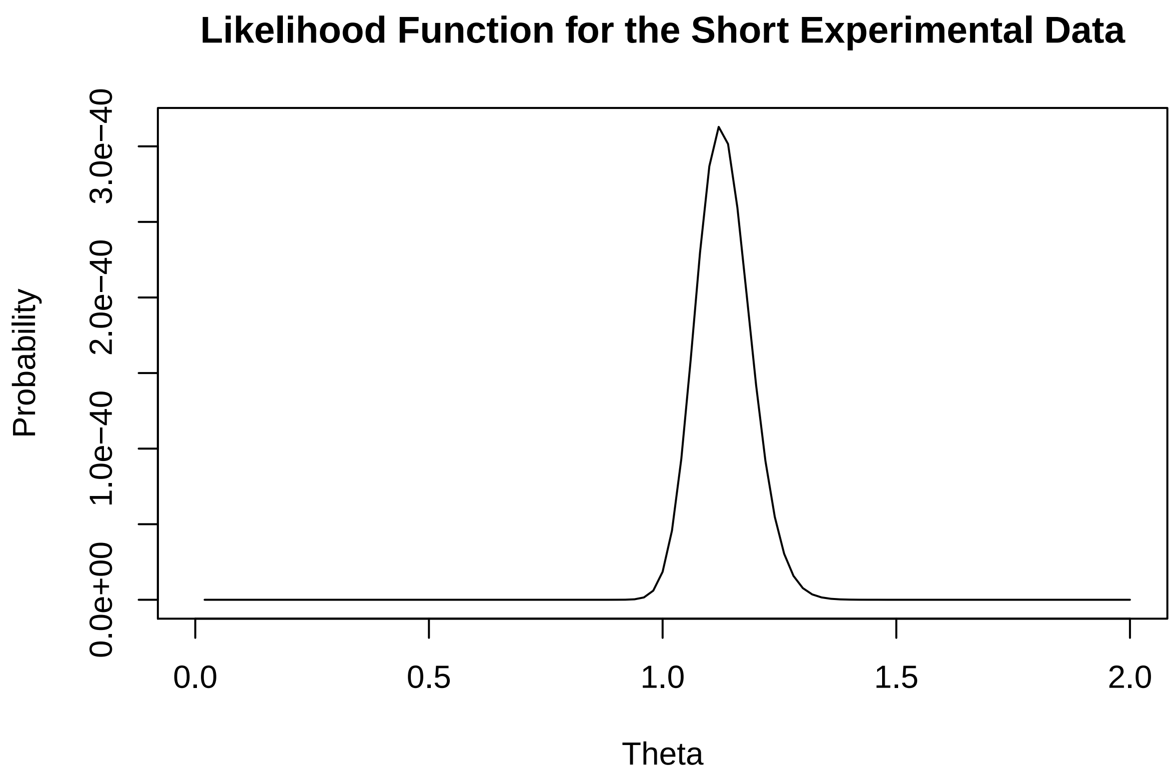
Data Analysis

```
# Returns the likelihood function for a Rayleigh
# distribution with the given data (formatted as a
# vector).
likeFunc <- function(data) {
    n = length(data)
    sumLnRi <- sum(log(data))
    sumRiSquared <- sum(data^2)
    output <- function(theta) {
        # Do the exp-log of the likelihood to avoid overflow
        # errors
        exp(-2 * n * log(theta) + sumLnRi - (sumRiSquared/(2 *
            theta^2)))
    }
    return(output)
}

# Returns the maximum likelihood estimate and estimated
# variance. output$mle is the estimate, output$var is
# the estimated variance of the MLE estimator
mle <- function(data) {
    n <- length(data)
    output <- list()
    output$mle <- sqrt(mean(data^2)/2)
    output$var <- output$mle/(4 * n)
    return(output)
}
```

Short Experiment

```
short_mle <- mle(short_data)
short_likeFunc <- likeFunc(short_data)
plot(short_likeFunc, from = 0, to = 2, n = 101, type = "l",
    main = "Likelihood Function for the Short Experimental Data",
    xlab = "Theta", ylab = "Probability")
```



The MLE estimate is $\hat{\theta}_{MLE} = 1.1237776$, with an estimated variance of 0.0029265.