

Problem 46

Overview

Chromatin folding is a poorly understood process in biology. We assume that the 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

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

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