

STAT 135 - Lab 3

GSI: Yutong Wang
Feb 7, 2020

To-do Today

1. Review MLE Problems
 - a. MLE for the Poisson distribution
 - b. MLE for the uniform p.d.f.
2. Data cleaning in R
3. Quiz time (45min)

Review: Maximum Likelihood Estimator

Suppose R.V.s X_1, X_2, \dots, X_n have a joint density $f(X_1, \dots, X_n | \theta)$.

Given observed values $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ which are fixed numbers.

1. **Write down the likelihood function of parameter θ .**

The likelihood function is $L(\theta) = f(x_1, x_2, \dots, x_n | \theta)$

If X_i are independent and identically distributed (i.i.d.), we have $L(\theta) = \prod_{i=1}^n f(x_i | \theta)$

2. **Optimize the likelihood function over θ .**

$\operatorname{argmax}_{\theta \in \Theta} (L(\theta; X))$ represents the parameter θ that maximizes the likelihood of observing our data.

We optimize by (a) setting the derivative to be zero; (b) taking the endpoint or other approaches.

Note: it is always easier to differentiate w.r.t. θ if we take logarithm of the function first.

MLE for Poisson distribution

Consider n i.i.d. random variables, each of them $\text{Pois}(\lambda)$, with unknown parameter λ . Suppose we measure the values $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$. The likelihood function is $\ell(\lambda) = f(x_1|\lambda) \cdots f(x_n|\lambda)$, where $f(k|\lambda) = e^{-\lambda} \lambda^k / k!$, with $k = 0, 1, 2, 3, \dots$. Find the value of λ that maximizes the probability of the data, i.e., the maximum likelihood estimate $\hat{\lambda}_{\text{ML}}$.

MLE for Uniform p.d.f.

Consider n i.i.d. continuous random variables, each of them $\text{Unif}(0, \theta)$, with unknown parameter θ . Suppose we measure the values $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ (here $x_i > 0$ for all i). Compute the likelihood function for such data sample, and compute $\hat{\theta}_{\text{ML}}$.

Data Cleaning in R

- 1) Go to bCourses/Pages/Lab, and download *beeswas.txt*
- 2) Load *beeswas* into R as a dataframe called *bees*
- 3) Examine the structure of the bees using command *str(bees)*. Notice that the data is factor data with the numbers as strings. This is terrible to work with. Best to convert to a numeric vector.

Data Cleaning in R

- 4) Give the column names: meltingpoint and hydrocarbon
- 5) Switch the order of the columns
- 6) Keep just meltingpoints > 63 and hydrocarbons > 14 using *filter()*
- 7) Use *mutate()* to multiply every value of hydrocarbon by 2

QUIZ TIME

