Department of Computer Science
UNIVERSITY OF COLORADO **BOULDER**

Maximum Likelihood

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

SEPTEMBER 29, 2016

Maximum Likelihood

- Going from data to parameters
- Deriving the things we just told you on faith
- Using in HW2

Exponential Distribution

You observe $x_1, x_2, \ldots x_N$. What is the MLE for the parameter θ?

$$f_\theta(x) = \lambda \exp\{-\lambda x\} \mathbb{1}[x > 0] \tag{1}$$

$(\lambda > 0)$

Exponential Distribution

You observe $x_1, x_2, \dots x_N$. What is the MLE for the parameter θ?

$$f_\theta(x) = \lambda \exp\{-\lambda x\} \mathbb{1}[x > 0] \tag{1}$$

$(\lambda > 0)$

$$\ell = N \log \lambda - \sum_i \lambda x_i \tag{2}$$

Exponential Distribution

You observe $x_1, x_2, \ldots x_N$. What is the MLE for the parameter θ?

$$f_\theta(x) = \lambda \exp\{-\lambda x\} \mathbb{1}[x > 0] \tag{1}$$

$(\lambda > 0)$

$$\ell = N \log \lambda - \sum_i \lambda x_i \tag{2}$$

$$0 = \frac{N}{\lambda} - \sum_i x_i \tag{3}$$

$$\lambda = \frac{N}{\sum_i x_i} \tag{4}$$

$$\tag{5}$$

Uniform Distribution

You observe $x_1, x_2, \ldots x_N$ from a distribution uniform in $[0, \theta]$. What is the MLE for the parameter θ? Real problem: tanks from serial numbers.

Uniform Distribution

You observe $x_1, x_2, \dots x_N$ from a distribution uniform in $[0, \theta]$. What is the MLE for the parameter θ? Real problem: tanks from serial numbers.

Probability of observing N observations from uniform distribution

$$f_\theta(\vec{x}) = \prod_i \frac{1}{\theta} = \frac{1}{\theta}^N \mathbb{1}\left[0 \leq x_i \leq \theta\right] \tag{6}$$

Uniform Distribution

You observe $x_1, x_2, \dots x_N$ from a distribution uniform in $[0, \theta]$. What is the MLE for the parameter θ? Real problem: tanks from serial numbers. Probability of observing N observations from uniform distribution

$$f_\theta(\vec{x}) = \prod_i \frac{1}{\theta} = \frac{1}{\theta}^N \mathbb{1}\left[0 \le x_i \le \theta\right] \quad (6)$$

$$\ell = \begin{cases} -N \log \theta & \text{if } \theta > \max x_i \\ -\infty & \text{otherwise} \end{cases} \quad (7)$$

Uniform Distribution

You observe $x_1, x_2, \dots x_N$ from a distribution uniform in $[0, \theta]$. What is the MLE for the parameter θ? Real problem: tanks from serial numbers. Probability of observing N observations from uniform distribution

$$f_\theta(\vec{x}) = \prod_i \frac{1}{\theta} = \frac{1}{\theta}^N \mathbb{1}\left[0 \leq x_i \leq \theta\right] \quad (6)$$

$$\ell = \begin{cases} -N \log \theta & \text{if } \theta > \max x_i \\ -\infty & \text{otherwise} \end{cases} \quad (7)$$

Maximum at $\theta = \max x_i$. (But biased down: needs to be adjusted up.)

More complicated: Poisson

- We have the following data

Number Marriagies	Age
0	12
0	50
2	30
2	36
6	92

Assuming a model

Let's assume that the number of marriages comes from a Poisson distribution whose parameter is a function of age

$$\lambda_i = \lambda_0 \text{age}_i \tag{8}$$

- Likelihood
- Log-likelihood
- Gradient λ_0
- MLE

Assuming a model

Let's assume that the number of marriages comes from a Poisson distribution whose parameter is a function of age

$$\lambda_i = \lambda_0 \text{age}_i \tag{8}$$

- Likelihood

$$p(x_i) = \frac{\exp\{\lambda_0 \text{age}\} (\lambda_0 \text{age})^{x_i}}{x_i!} \tag{9}$$

- Log-likelihood
- Gradient λ_0
- MLE

Assuming a model

Let's assume that the number of marriages comes from a Poisson distribution whose parameter is a function of age

$$\lambda_i = \lambda_0 \text{age}_i \tag{8}$$

- Likelihood

$$p(x_i) = \frac{\exp\{\lambda_0 \text{age}\}(\lambda_0 \text{age})^{x_i}}{x_i!} \tag{9}$$

- Log-likelihood

$$\ell = \sum_i \log\left[\frac{\exp\{\lambda_0 \text{age}_i\}(\lambda_0 \text{age}_i)^{x_i}}{x_i!}\right] \tag{10}$$

$$= \sum_i \log\left[\exp\{-\lambda_0 \text{age}_i\}(\lambda_0 \text{age}_i)_i^x\right] - \log x_i! \tag{11}$$

$$= -\lambda_0 \sum_i \text{age}_i + \sum_i x_i \log(\lambda_0 \text{age}_i) - \sum_i \log x_i! \tag{12}$$

Assuming a model

Let's assume that the number of marriages comes from a Poisson distribution whose parameter is a function of age

$$\lambda_i = \lambda_0 \text{age}_i \tag{8}$$

- Likelihood

$$p(x_i) = \frac{\exp\{\lambda_0 \text{age}\}\,(\lambda_0 \text{age})^{x_i}}{x_i!} \tag{9}$$

- Log-likelihood

$$\ell = -\lambda_0 \sum_i \text{age}_i + \sum_i x_i \log(\lambda_0 \text{age}_i) - \sum_i \log x_i! \tag{10}$$

- Gradient λ_0
- MLE

Age of Marriage

- Gradient λ_0
- λ_0 MLE

Age of Marriage

- Gradient λ_0

$$\frac{\partial \ell}{\partial \lambda_0} = -\sum_i \text{age}_i + \sum_i x_i \frac{\partial \log \lambda_0 \text{age}_i}{\partial \lambda_0} \tag{11}$$

$$= \sum_i \text{age}_i + \sum_i x_i \frac{\text{age}_i}{\lambda_0 \text{age}_i} \tag{12}$$

$$= \sum_i \text{age}_i + \frac{1}{\lambda_0} \sum_i x_i \tag{13}$$

- λ_0 MLE

Age of Marriage

- Gradient λ_0

$$= \sum_i \text{age}_i + \frac{1}{\lambda_0} \sum_i \text{age}_i \tag{11}$$

- λ_0 MLE

Age of Marriage

- Gradient λ_0

$$=\sum_i \text{age}_i + \frac{1}{\lambda_0}\sum_i \text{age}_i \tag{11}$$

- λ_0 MLE

$$0 = -\sum_i \text{age}_i + \frac{1}{\lambda_0}\sum_i x_i \tag{12}$$

$$\sum_i \text{age}_i = \frac{1}{\lambda_0}\sum_i x_i \tag{13}$$

$$\lambda_0 = \frac{\sum_i x_i}{\sum_i \text{age}_i} \tag{14}$$

Our dataset

Number Marriagies	Age
0	12
0	50
2	30
2	36
6	92

- λ_0?
- Expected number of marriages for someone 22 years old? Most likely number of marriages?

Our dataset

Number Marriagies	Age
0	12
0	50
2	30
2	36
6	92

- λ_0: 0.044
- Expected number of marriages for someone 22 years old? Most likely number of marriages?

Our dataset

Number Marriagies	Age
0	12
0	50
2	30
2	36
6	92

- λ_0: 0.044
- Expected number of marriages for someone 22 years old? Most likely number of marriages?
    - $\mathbb{E}_{\lambda_0}[X] = 1.0$

Our dataset

Number Marriagies	Age
0	12
0	50
2	30
2	36
6	92

- λ_0: 0.044
- Expected number of marriages for someone 22 years old? Most likely number of marriages?
  - $\mathbb{E}_{\lambda_0}[X] = 1.0$
  - Two modes: 0, 1

Now we're getting somewhere!

- Data Science = Reverse of Probabilities
- Building models from data
- Making predictions, refining models

Quiz

- Poisson distribution
$$f(x) = \frac{\exp\{\lambda\}(\lambda)^x}{x!} \tag{15}$$
- Normal distribution
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$