Matlab 9: Maximum Likelihood Estimate

Cheng-Hsin Hsu

National Tsing Hua University

Department of Computer Science

Slides are based on the materials from Prof. Roger Jang

What is Maximum Likelihood Estimate

- MLE
  - Maximum likelihood estimate
- Goal:
  - Given a dataset with no labels, how can we find the best statistical model with the optimum parameters to describe the data?
- Applications
  - Prediction
  - Analysis

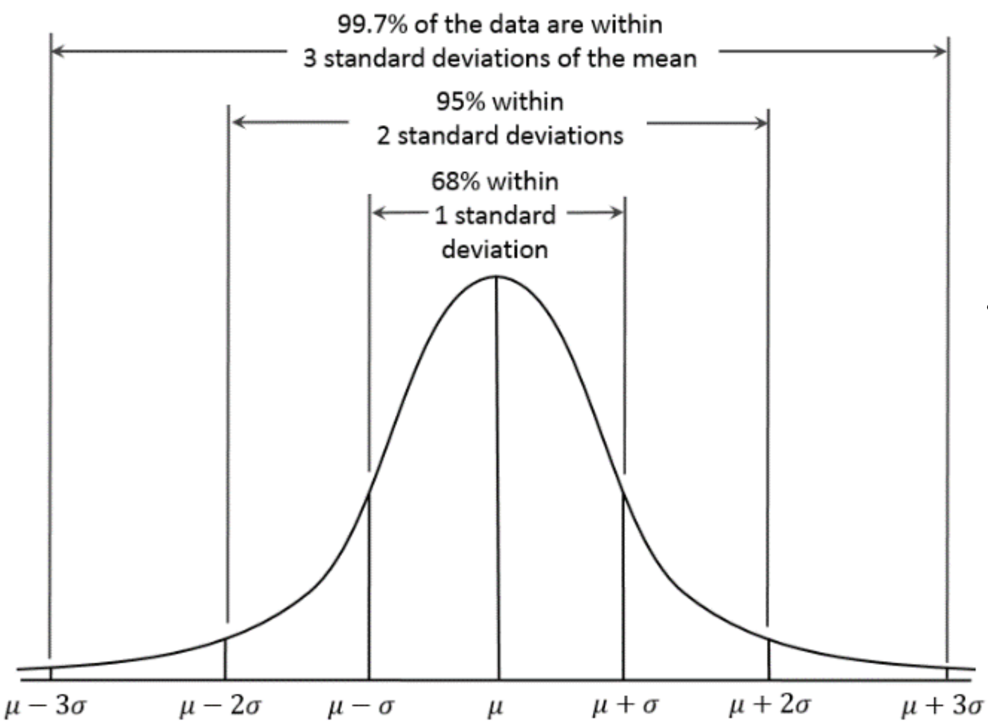
What Are Statistical Models?

- Statistical models are used to describe the probabilities of random variables
  - Discrete variables ➔ Probability (mass) functions (PMF)
  - Continuous variables ➔ Probability density functions (PDF)
- Examples
  - Discrete variables
    - The outcome of tossing a coin or a die

  - Continuous variables
    - The distance to the bull eye when throwing a dart
    - The time needed to run 100-m dash
    - The heights of kids in a kindergarten

More about Models

- Discrete variables
  - Outcome of tossing a coin ➔ Pr{head}=1/2, Pr{tail}=1/2
- Continuous variables
  - Distance to the bull's eye when throwing a dart ➔ A PDF of Gaussian or normal distribution



$$g\left(x;\mu,\sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$\Pr\{x \in [4,6]\} = \int_4^6 g(x;\mu,\sigma^2)dx$$

Basic Steps in MLE

- Steps
  1. Perform a experiments to collect data
  2. Choose a parametric model of the data, with certain tunable parameters
  3. Formulate the likelihood as an objective function to be maximized
  4. Maximize the objective function and derive the parameters of the model
- Examples
  - Flip a coin → To find the probabilities of head and tail
  - Throw a dart → To find your PDF of distance to the bull eye
  - Height of people in a day care

Probability Mass Functions for Discrete Variables

- Flip an unfair coin 5 times to get 3 heads and 2 tails
  - By intuition: Pr(head)=3/5, Pr(tail)=2/5
  - By MLE
    - Assume these 5 tosses are independent events to have the overall probability

$$J(p,q) = p^3 q^2, \; with \;\; p + q = 1, p \geq 0, q \geq 0$$

$$\Rightarrow J(p) = p^3 (1-p)^2$$

$$\Rightarrow \frac{dJ(p)}{dp} = 0$$

$$\Rightarrow p = 3/5, q = 2/5$$

Arithmetic Mean >= Geometric Mean

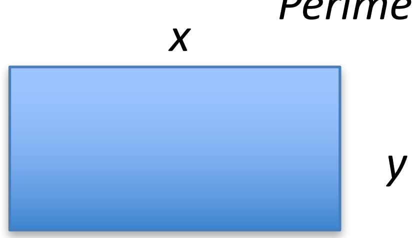
- AM-GM inequality

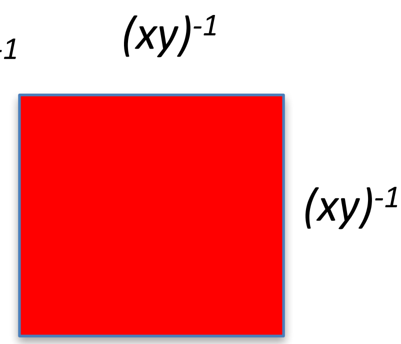
$$\frac{\sum_{i=1}^{n} x_i}{n} \geq \left(\prod_{i=1}^{n} x_i \right)^{1/n}, \ with \ x_i \geq 0, \forall i$$

The equality holds only when $x_1 = x_2 = \cdots = x_n$.

- Algebraic intuition:
  - (x-y) ^{2} = x^2 – 2xy + y^2 = (x+y) ^{2} – 4xy >=0 ➔ (x+y)^2 >= 4xy
- Geometrical intuition:

Perimeter: x+y >= 2 (xy) $^{-1}$





Use AM-GM Inequality for MLE Problems

$$\frac{\sum_{i=1}^{n} x_i}{n} \geq \left(\prod_{i=1}^{n} x_i\right)^{1/n}, \textit{ with } x_i \geq 0, \forall i$$

The equality holds only when $x_1 = x_2 = \cdots = x_n$.

- Goal is to get rid of p and q on the left-hand side, remember that *p + q = 1*

$$\frac{\frac{p}{3}+\frac{p}{3}+\frac{p}{3}+\frac{q}{2}+\frac{q}{2}}{5} \geq \left(\left(\frac{p}{3}\right)^3\left(\frac{q}{2}\right)^2\right)^{1/5}$$

$$\Rightarrow p^3 q^2 \text{ achieves its maximum when } \frac{p}{3} = \frac{q}{2} \Rightarrow p = \frac{3}{5}, q = \frac{2}{5}$$

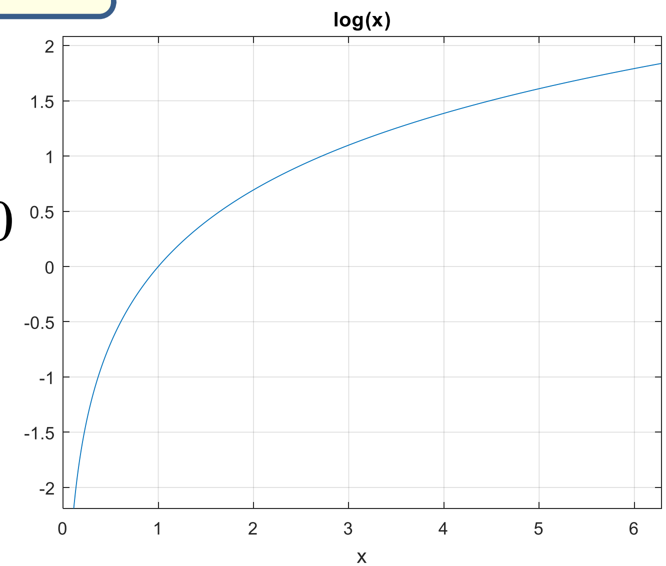
How to Prove AM-GM Inequality?

- Basic inequality

Jensen's inequality

$y = \ln x$ is a concave function $\Rightarrow$

$$\ln\left(\frac{mp+nq}{m+n}\right) \geq \frac{m\ln p + n\ln q}{m+n}, \ with \ \ p,q>0$$



- Proof by induction

$$\ln\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) \geq \left(\frac{\sum_{i=1}^{n} \ln x_i}{n}\right), \ with \ x_i > 0, \ \forall i$$

Proof Sketch by Induction

$$n=1 \Rightarrow x_1 \ge x_1$$

$$n=2 \Rightarrow \ln\left(\frac{x_1+x_2}{2}\right) \ge \frac{\ln x_1 + \ln x_2}{2}. \text{(Or you can start with } \left(\sqrt{x_1}-\sqrt{x_2}\right) \ge 0\text{)}$$

$$n=3 \Rightarrow \ln\left(\frac{x_1+x_2+x_3}{3}\right) = \ln\left(\frac{2\left(\frac{x_1+x_2}{2}\right)+x_3}{3}\right) \ge \frac{2\ln\left(\frac{x_1+x_2}{2}\right)+\ln x_3}{3} \ge \frac{\ln x_1+\ln x_2+\ln x_3}{3}$$

$$n=k \text{ holds by assumption} \Rightarrow \ln\left(\frac{\sum_{i=1}^{k} x_i}{k}\right) \ge \left(\frac{\sum_{i=1}^{k} \ln x_i}{k}\right)$$

$$n=k+1 \Rightarrow \ln\left(\frac{\sum_{i=1}^{k} x_i + x_{k+1}}{k+1}\right) = \ln\left(\frac{k\frac{\sum_{i=1}^{k} x_i}{k}+x_{k+1}}{k+1}\right) \ge \frac{k\ln\left(\frac{\sum_{i=1}^{k} x_i}{k}\right)+\ln x_{k+1}}{k+1} \ge \frac{k\left(\frac{\sum_{i=1}^{k} \ln x_i}{k}\right)+\ln x_{k+1}}{k+1} = \frac{\sum_{i=1}^{k+1} \ln x_i}{k+1}$$

Another Probability Mass Function

- Toss a 3-side die for many times and obtain n_1 of side 1, n_2 of side 2, and n_3 of side 3, then what is the most likely probabilities for sides 1, 2, and 3, respectively?
  - Our intuition…
  - By MLE…

$$J(p,q,r) = p^{n_1} q^{n_2} r^{n_3}, \; with \; p+q+r=1, p \geq 0, q \geq 0, r \geq 0$$

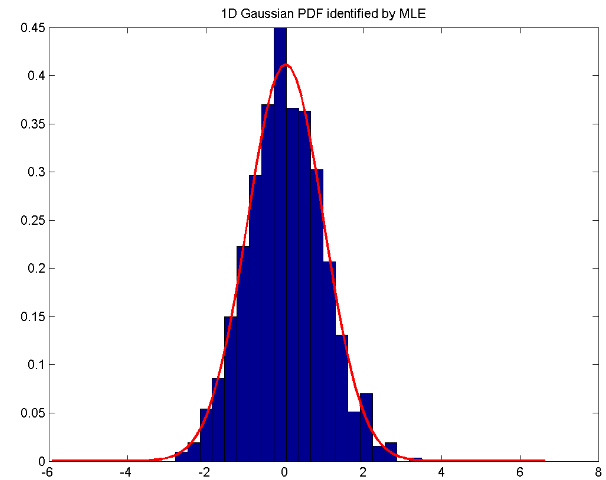
MLE for PDF of Continuous Variables of 1D

$$g\left(x;\mu,\sigma^2\right)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

PDF

$$X=\{x_1,x_2\cdots x_n\}$$
$$\Rightarrow p\left(X;\mu,\sigma^2\right)=\prod_{i=1}^{n} g\left(x_i;\mu,\sigma^2\right)$$

Overall PDF, or likelihood



$$\begin{aligned}J(\mu,\sigma^2)&=\ln p\left(X;\mu,\sigma^2\right)\\&=\ln\left[\prod_{i=1}^{n} g\left(x_i;\mu,\sigma^2\right)\right]\\&=\sum_{i=1}^{n}\ln g\left(x_i;\mu,\sigma^2\right)\\&=\sum_{i=1}^{n}\left[-\frac{1}{2}\ln(2\pi)-\ln\sigma-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]\\&=-\frac{n}{2}\ln(2\pi)-n\ln\sigma-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2\end{aligned}$$

Log likelihood

MLE

$$\frac{\partial J(\mu,\sigma^2)}{\partial\mu}=\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)=0\Rightarrow\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\frac{\partial J(\mu,\sigma^2)}{\partial\sigma}=-\frac{n}{\sigma}-\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)\left(-\frac{x_i-\mu}{\sigma^2}\right)=0\Rightarrow\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\hat{\mu}\right)^2$$

MLE for PDF of Continuous Variables of ND

$$g(x;\mu,\Sigma)=\frac{1}{\sqrt{(2\pi)^d|\Sigma|}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

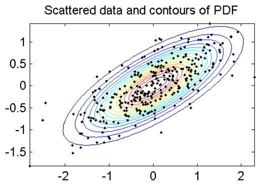
PDF

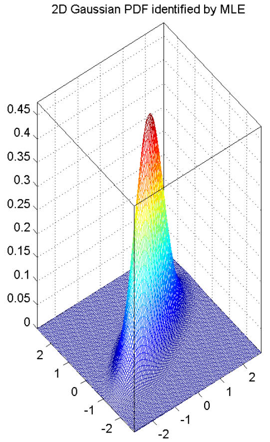
$$X=\{x_1,x_2\cdots x_n\}$$
$$\Rightarrow p(X;\mu,\Sigma)=\prod_{i=1}^{n}g(x_i;\mu,\Sigma)$$

Overall PDF, or likelihood

Log likelihood

$$\begin{aligned}J(\mu,\Sigma)&=\ln p(X;\mu,\Sigma)\\&=\ln\left[\prod_{i=1}^{n}g(x_i;\mu,\Sigma)\right]\\&=\sum_{i=1}^{n}\ln g(x_i;\mu,\Sigma)\\&=\sum_{i=1}^{n}\left[-\frac{d}{2}\ln(2\pi)-\frac{1}{2}\ln|\Sigma|-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right]\\&=-\frac{nd}{2}\ln(2\pi)-\frac{n}{2}\ln|\Sigma|-\frac{1}{2}\sum_{i=1}^{n}\left[(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right]\end{aligned}$$





$$\begin{aligned}\nabla_\mu J(\mu,\Sigma)&=-\frac{1}{2}\sum_{i=1}^{n}\left[-2\Sigma^{-1}(x_i-\mu)\right]\\&=\Sigma^{-1}\left(\sum_{i=1}^{n}x_i-n\mu\right)\end{aligned}$$

$$\nabla_\mu J(\mu,\Sigma)=0\Rightarrow\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

MLE

$$\hat{\Sigma}=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\hat{\mu}\right)\left(x_i-\hat{\mu}\right)^T$$

Discussions

- Can we choose other probability density function instead of Gaussian/normal distributions? ➔ Yes!
- What are the other available PDF? → Check matlab’s mle function to see what distributions are supported!

Matlab #8 Homework (M8)

1. (2%) Learn how to use the mle(.) function of Matlab. Please submit a short PDF file, and provide all the details in it.
    1. (0.5%) Generate 10000 random data samples following the gamma distribution with arbitrary parameters (you get to pick them). Plot the histogram.
    2. (1%) Apply the mle(.) function to estimate the parameters. Show your code in the report.
    3. (0.5%) Compare the parameters selected by you and the estimated parameters. What are your observations?

Questions?