# Probability and Statistics II

Changmin Yu

February 2020

## 1 Maximum Likelihood Estimation

- Given data $y \in \mathbb{R}^n$ as realisations of a random variable $Y$, specify its density $f(y;\theta)$ up to some unknown vector of parameters $\theta \in \Theta \subset \mathbb{R}^d$, where $\Theta$ is the parameter space.
- Define the **likelihood function** as a function of the parameters $\theta$:

$$L(\theta) = L(\theta; y) = c(y) f(y;\theta) \tag{1}$$

  where $c(y)$ is some unknown constant for normalisation

- The maximum likelihood estimator (MLE) of $\theta$, $\hat{\theta}$, is the value of parameters such that $\hat{\theta}$ maximises $L(\theta)$.

$$\hat{\theta} = \arg\max_{\theta} L(\theta; y) \tag{2}$$

- It is often to work with log-likelihood function.

## 2 Maximum A Posteriori Estimation

- Bayes' theorem:

$$f(x|y) = \frac{f(y|x)f(x)}{f(y)} = \frac{f(y|x)f(x)}{\int f(y|x)f(x)dx} \tag{3}$$

- Given data $y$ and parametric density function $f(y;\theta)$, the maximum a posteriori (MAP) estimator of $\theta$ given its prior belief $g(\theta)$ is

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} \frac{f(y;\theta)g(\theta)}{\int f(y;\theta)g(\theta)d\theta} \tag{4}$$

# 3 Ordinary Least Squares

- $Y = X\beta + \epsilon$
- OLS estimator

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\arg\min} ||Y - X\beta||^2 = (X^T X)^{-1} X^T Y \tag{5}$$

note that this is also an orthogonal projection onto the column space of $X$.