MA2252 Introduction to computing

lectures 13-14

Basic statistics, reading and copying files, plotting

Matias Ruiz

October 2023

Basic statistics

Statistical quantities on an array summarise global information of the elements of the array.

a = $[a_1, a_2, a_3, \cdots a_n]$

- Mean:

$$\bar{a} = \frac{1}{n} \sum_{k=1}^{n} a_k$$

  Careful, the `mean` is very sensitive to outlayers. The median is robust to outlayers: `median` = middle element of the sorted array

- Harmonic mean:

$$\frac{n}{\sum_{k=1}^{n} \frac{1}{a_k}}$$

- Geometric mean:

$$(\Pi_{k=1}^{n} a_k)^{1/n}$$

Basic statistics

Statistical quantities on an array summarise global information of the elements of the array.

a = $[a_1, a_2, a_3, \cdots a_n]$

- Variance:

$$v = \frac{1}{n-1}\sum_{k=1}^{n}(a_k - \bar{a})^2$$

  Sometimes it's useful to standardise your data so a' has mean 0 and variance 1:

$$a' = \frac{a - \bar{a}}{\sqrt{v}}$$

Basic statistics

If we have two arrays we can estimate how much the correlate

$\mathtt{a} = [a_1, a_2, a_3, \cdots a_n]$ and $\mathtt{b} = [b_1, b_2, b_3, \cdots b_n]$

- Covariance:

$$c = \frac{1}{n-1} \sum_{k=1}^{n} (a_k - \bar{a})(a_k - \bar{a})$$

- Correlation

$$c' = \frac{1}{n-1} \sum_{k=1}^{n} a'_k b'_k$$

Basic statistics: reminder of random variables

Matlab has already many built-in functions for computing basic statistical quantities:

- Mean: `mean(a)`
- Harmonic mean: `harmmean(a)`
- Geometric mean: `geomean(a)`
- Median: `med(a)`
- Variance: `var(a)`
- Covariance: `cov(a,b)`
- Correlation: `corrcoef(a,b)`

Reading and copying files

- Often the data will come from exterior sources.
- With `MATLAB` you can read and write files (xls, test, csv, etc) to and from your work file.
- **Excel**:
  `[numeric, txt, raw] = xlsread('file_name', options)`
  and
  `[success, message] = write('file_name', options)`.
- **text**: `fopen`, `fclose`, `fprintf`.
- **csv**: `csvread`, `csvwrite`.

Plotting

- **2D Plotting**: `plot`, `imagesc`
- **3D Plotting**: `surf` (typically with the help of `mershgrid`)