STATISTICS ON ITK

Based on “The ITK Software Guide”

24 September 2024

SAMARASEKERA A.M.P.S. 200558U

ITK STATISTICS

- Primary purpose is to provide general capabilities for statistical pattern classification.
- Not limited for classification.
- Also used to
  - Perform statistical analysis
  - Preprocess image data for other tasks.

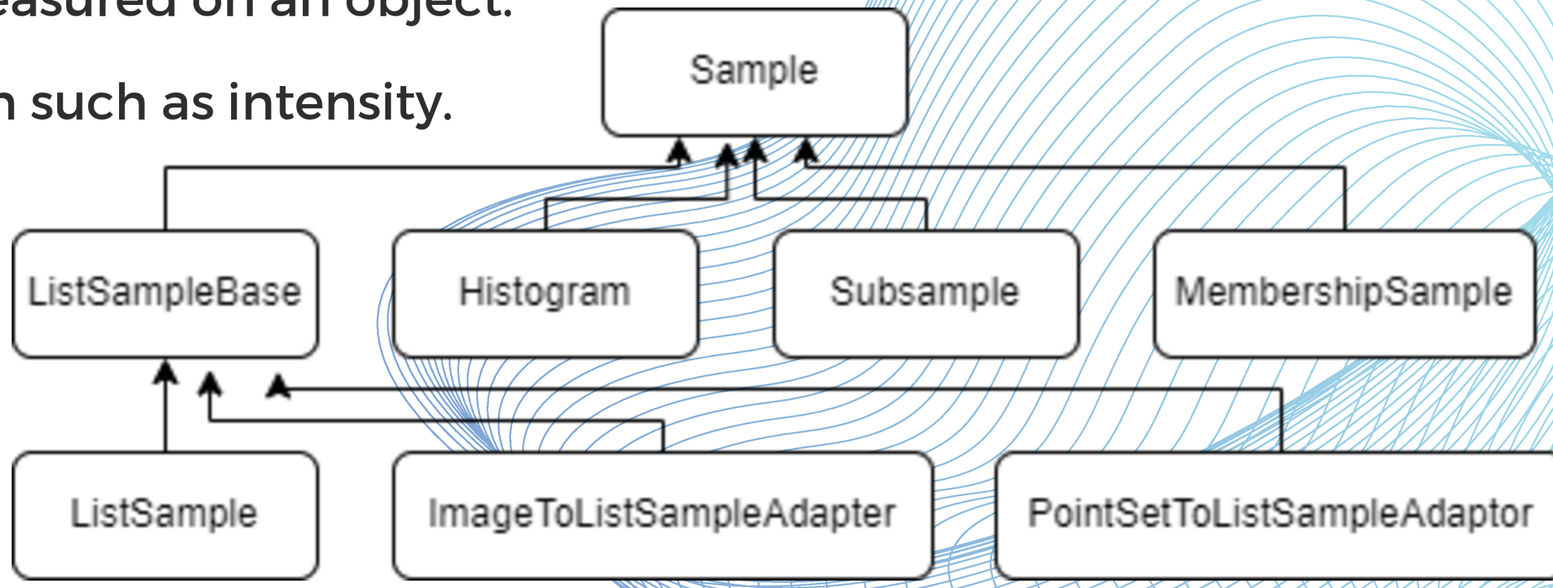
DATA CONTAINERS

itk::Statistics::Sample - Data container of elements called measurement vectors.

Measurement Vectors?

Array of values measured on an object.

Stores information such as intensity.



SAMPLE INTERFACE

Two methods for storing vectors.

PushBack Method : Add a vector at the end

Resize Method : Adjust the size using ‘SetMeasurementVector’

SAMPLE ADAPTORS

Treating other data structures images, and point sets as samples.

- **ImageToListSampleAdaptor:** Treats an image as a list of measurement vectors.
- **PointSetToListSampleAdaptor:** Treats points from a PointSet as measurement vectors.
- **JointDomainImageToListSampleAdaptor:** Adapts both pixel values and their physical coordinates for statistical analysis.

HISTOGRAM

Class that stores measurement vectors in bins.

SUBSAMPLE

Class that creates a subset of an existing Sample by referencing elements rather than copying them.

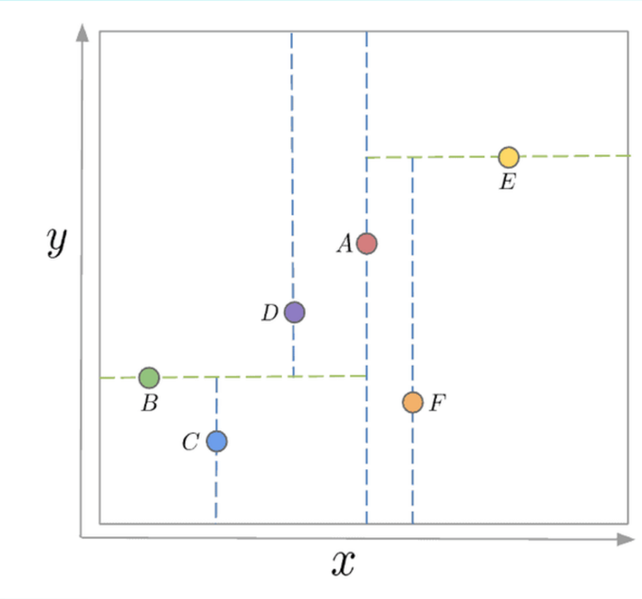
MEMBERSHIP SAMPLE

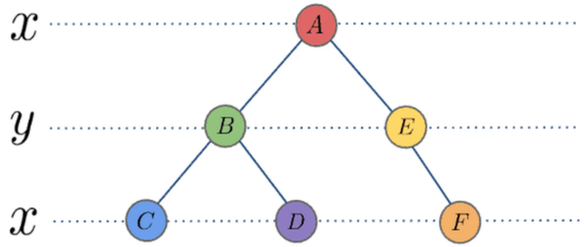
Same as subsample but adds classification information.

For labelled data, ‘MembershipSampleGenerator’ is used to create Membership Samples.

K-D TREE

- Capable of organising data points in a k-dimensional space.
- Is a binary tree.
- To add a new point, compare its coordinates with the current node's coordinates to decide whether to go left (smaller) or right (larger), alternating dimensions as you go down the tree.
- Used for tasks such as search and nearest-neighbour queries.





ALGORITHMS AND FUNCTIONS

SAMPLE STATISTICS

- Mean - ‘MeanSampleFilter’ to compute the mean vector of a sample.
- Covariance - ‘CovarianceSampleFilter’ computes the covariance matrix.

- Weighted Mean - ‘WeightedMeanSampleFilter’ to compute the weighted mean vector of a sample.
- Weighted Covariance - ‘WeightedCovarianceSampleFilter’ computes the weighted covariance matrix.

Used for processes such as expectation-maximization parameter estimation.

SAMPLE GENERATION

SampleToHistogramFilter: Converts a sample into a histogram.
NeighborhoodSampler: Extracts samples from a spatial neighborhood within an image, used for statistical analysis with proximity or spatial consideration.

DISTANCE METRIC

Computes the Euclidean distance between two measurement vectors,

Used in clustering and nearest-neighbor searches.

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

DECISION RULES

- Used in classification
- Returns the index of one data element in a vector of data elements.
- A set of membership functions (e.g. probability density functions, distance metrics) are compared.

Maximum Decision Rule: Selecting the class with the highest value

Minimum Decision Rule: Chooses the class with the smallest value.

Maximum Ratio Decision Rule: Uses ratios to decide classification.

RANDOM VARIABLE GENERATION

Generates random variables that follow a Gaussian distribution,
Used in simulations and statistical modeling.

STATISTICS APPLIED TO IMAGES

IMAGE HISTOGRAMS

- A graphical representation of the distribution of pixel intensities in an image.
- Shows the frequency of occurrence of intensity range (bin).

Scalar Image Histogram with Adaptor

1. Use **ImageToListSampleAdaptor** to treat an image as a list of samples.
2. Use **SampleToHistogramFilter** to compute the histogram.

STATISTICS APPLIED TO IMAGES

IMAGE INFORMATION THEORY

IMAGE ENTROPY

- Measures the uncertainty in an image.
- The formula is based on the probability distribution of pixel intensities derived from the histogram.

$$H = -\sum_i p_i \cdot \log_2 (p_i)$$

STATISTICS APPLIED TO IMAGES

MUTUAL INFORMATION

- Mutual information measures how much information two images share.
- Commonly used in **image registration** tasks.

Three formats used.

1. Mutual Information = Entropy_1 + Entropy_2 - Joint_Entropy
2. Normalized Mutual Information = 2 x Mutual_information / (Entropy_1 + Entropy_2)
3. Mutual Information = (Entropy_1 + Entropy_2) / Joint_Entropy

Where joint entropy :

$$\mathrm{H}(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log_2[P(x,y)]$$

THANK YOU