**Assignment: Estimating Photometric Redshift of a Galaxy Cluster Using Red Sequence**

# Objective

The goal of this assignment is to estimate the photometric redshift of a galaxy cluster from DES Y6 (proprietary) griz data by exploiting the red sequence, a prominent feature in color-magnitude diagrams of galaxy clusters. Students will first identify the red sequence and calibrate its evolution in g-r vs r, r- i vs i, and i-z vs z as a function of redshift. This involves calibrating the slope, normalization, and intrinsic scatter characterizing the Color-Magnitude Relation (CMR).

# Tasks

## Data Preparation:

* + Download (e.g., from VizieR) and load the photometric data for galaxies from the COSMOS survey (Weaver et al. 2022). (Either “classic” and “farmer” are fine, I used "farmer”).
  + Apply selection criteria to filter the galaxies based on redshift, morphology, and stellar mass:
    - Redshift *z <* 1*.*5
    - Exclude point sources (["SolModel"] != "PointSource")
    - Exclude flagged models (["FModel"] == 0)
    - Stellar mass *>* 108*.*5*M*⊙

## Passive Galaxy Selection:

* + Implement a function to identify passive (red) galaxies in the COSMOS data using UVJ color-color cuts from Whitaker et al. (2011).
    - Note that in the COSMOS survey, the flux for example in the rest-frame U band is provided as "EZrestU". You will need to convert this flux to a AB magnitude using, e.g.:  
      U=−2.5×log10(EZrestU)+23.9
  + Verify the existence of bimodality (passive/quiescent) in the data.

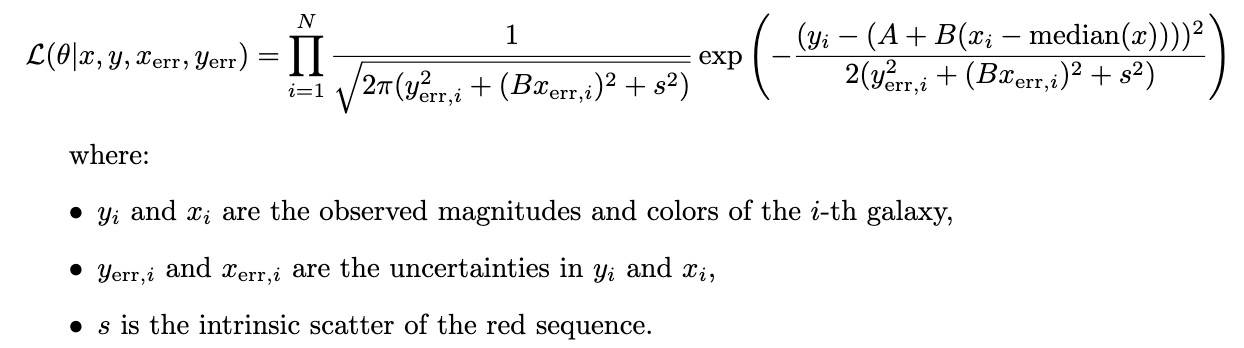
## Cross-Matching with DES Data:

* + Cross-match COSMOS passive galaxies with [DES data](https://drive.google.com/file/d/1OdG4sm6yUjuNeWdZ_9uKoPIuEOaWc9CA/view?usp=sharing) to obtain DES photometry for each COSMOS passive galaxy.
  + What is a good distance for cross-matching the catalogs?

## Color-Magnitude Diagrams:

* + Create color-magnitude diagrams (CMDs) for the selected galaxies using DES photometry.
  + Plot the g-r vs r, r-i vs i, and i-z vs z diagrams to visualize the red sequence at different redshifts.

## Red Sequence Calibration:

* + Divide the sample into redshift bins using COSMOS calibrated photometric redshifts.
  + For each redshift bin, fit a linear model to the three CMD of DES photometry.
    - First further clean the sample, for example using 3 sigma-clipping
    - Then, fit a linear model with normal distribution *y* = *A* + *B ·* (*x −* median(*x*)) ± s
    - Store the maximum posterior value of the normalization *A*, slope *B*, and intrinsic scatter *s* as a function of redshift.
  + Create a function that interpolates these parameters as a function of redshift for the three CMDs.
  + Visualize the evolution of these parameters with redshift.
  + Comment on the 4000 ˚A break transition as a function of redshift for the different bands.

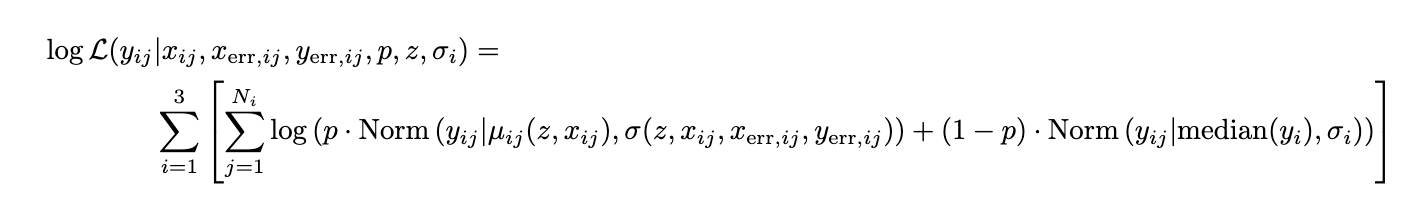
## Photometric Redshift Estimation:

* + Load the DES [Y6 data](https://drive.google.com/drive/folders/1socY9YcZ4FXR4428QC9kF_DZJuAeY2HG?usp=sharing) cutouts around the positions of SPT cluster candidates.
  + Explore the DES Y6 data and apply the following cuts:
    - Quality flags (e.g., FLAGS FOOTPRINT == 1, FLAGS GOLD == 0,FLAGSTR == "ok", FLAGS\_FOREGROUND == 0, FITVD\_FLAGS == 0, MASK\_FLAGS == 0, SPREADERR\_MODEL\_G <0.1 
    - Magnitude limits (e.g., G *<* 24*.*5, R *<* 24,

I *<* 23*.*4, Z *<* 22*.*75), MAGERR in the different bands >0

# Bayesian Analysis

For Bayesian inference, define the following likelihood function:



Assume that for each galaxy *i* in the cluster field, there’s a probability *p* of belonging to the cluster at redshift z (and thus with colors *yj* defined by the function previously calibrated that returns the CMD as a function of redshift), and a probability *1-p* that the galaxy is described by an interloper population which you can assume is normally distributed in each of the three color with mean equal to the median of the observed colors and variance described by σi (i=1-3).

# Steps to follow

## Initialize parameters and minimize the negative log-likelihood:

* + Initialize parameters and use a minimization algorithm (e.g., Powell or Nelder-Mead) to find the best starting point for the MCMC chains.
  + Iterate over a range of initial values of redshift to ensure the global minimum is found.
  + Store the best parameters as the starting point for MCMC.

## Run the MCMC chains:

* + Use the emcee library to run the MCMC chains and sample the posterior distribution of the model parameters (*p,z,σ1,σ2,σ3*)
  + Initialize the walkers around the best parameters found in the previous step.
  + Run the MCMC sampler for a sufficient number of steps to ensure convergence.

# Photometric redshifts

* What is the marginalized posterior on the photometric redshift?
* How does it compare the photometric redshift estimates you obtained with values you could retrieve from Vizier.