GALMORPH
PYTHON v. 1.0

1 Important Assumptions

- Galaxy images are assumed to be standard FITS images, located in HDU 1.
- The galaxy of interest is assumed to be in the center of the image (i.e., it is assumed that galaxies are each in individual, properly centered postage stamps that may have been extracted from a larger mosaic).
- It is assumed that the images are background-subtracted images.

If the second and third assumptions are violated, the segmentation map algorithm will fail.

If you need these assumptions relaxed, contact `pfreeman@cmu.edu` and I'll see what I can do.

2 Segmentation Map Construction

The construction of the segmentation map in this version of galmorph has been simplified (relative to the `R` version of the code base), and can be described in a single sentence: the map consists of connected pixels that overlap the center of the postage stamp galaxy image and have intensity $> n\hat{\sigma}$, where n is a user-controlled variable (default 5). $\hat{\sigma}$ is the estimate of the standard deviation of the background, computed using only those pixels with intensity < 0.

No attempt is made to separate blended galaxies. This is a conscious choice, as robust deblending is hard (and could significantly slow down computations). For now, it is up to the user in post-processing to attempt to identify blended galaxies using the full set of statistics, and to remove them from the sample if appropriate. (For instance, blended galaxies can have systematically larger Gini values.)

3 Image Statistics

All image statistics are computed within the segmentation map.

1. **Multimode** (M) statistic (based on Freeman et al. 2013)
   The M statistic identifies galaxies with disturbed morphologies. Consider an intensity quantile q_l such that a proportion l of the pixel intensities i_{mn} are smaller than q_l. (Here m and n denote pixel coordinates.) For a given q_l, we define a new indicator variable j_{mn} such that

$$j_{mn} = \begin{cases} 1 & \text{if } i_{mn} \geq q_l \\ 0 & \text{otherwise} \end{cases}$$

   Within the image j_{mn}, we obtain the areas of largest and second-largest groups of contiguous pixels, which we denote $A_{l,(1)}$ and $A_{l,(2)}$ respectively. We define the area ratio as

$$R_l = \frac{A_{l,(2)}^2}{A_{l,(1)} n_{seg}} ,$$

   where n_{seg} is the number of pixels in the segmentation map, i.e., the mask used to define the extent of the galaxy within the image. This formulation imposes a strict upper limit on R_1

of 1/2 that is achieved if $A_{l,(1)} = A_{l,(2)} = n_{seg}/2$. The M statistic is the maximum observed value of R_l over all quantiles l:

$$M = \max_l R_l\,.$$

2. **Intensity** (I) statistic (Freeman et al. 2013)
One of the shortcomings of the M statistic is that it does not consider the summed intensity within contiguous pixel groups. For instance, a contiguous group with a large number of pixels may have a smaller summed intensity than other, smaller groups of pixels. To mitigate this shortcoming we utilize the I statistic. We associate each pixel (m, n) with a local maximum (m', n') by following the maximum gradient ascent path. All pixels that are associated with a given pixel (m', n') are grouped together, and within each group, we sum the pixel intensities i_{mn}. (Note that in a data pre-processing step, we smooth the image data with a symmetric Gaussian kernel with $\sigma \sim 1$ pixel, to decrease the effect that pixel noise has on the construction of pixel groups.) We rank the summed intensities in descending order and use the first and second sorted values to compute the I statistic:

$$I = \frac{I_{(2)}}{I_{(1)}}\,.$$

3. **Deviation** (D) statistic (Freeman et al. 2013)
The deviation D statistic is used to capture evidence of galaxy asymmetry. It is the distance from the local maximum associated with $I_{(1)}$ to the galaxy's "center of mass":

$$(m_{\rm cen}, n_{\rm cen}) = \left(\frac{1}{n_{seg}} \sum_m \sum_n m i_{mn}, \frac{1}{n_{seg}} \sum_m \sum_n n i_{mn} \right)\,, \tag{1}$$

where the summation is over the n_{seg} pixels within the segmentation map. The D statistic is:

$$D = \sqrt{(m_{\rm cen} - m_{I_{(1)}})^2 + (n_{\rm cen} - n_{I_{(1)}})^2} \Big/ \sqrt{n_{seg}/\pi}\,.$$

where the normalizing factor $\sqrt{n_{seg}/\pi}$ is a galaxy radius estimate achieved by assuming that the segmentation map is circular.

4. **Gini** (G) statistic (Lotz et al. 2004)
The Gini coefficient measures the relative distribution of pixel intensities within the segmentation map: $G = 0$ means that the intensities are uniform across the galaxy, while $G = 1$ means that all of a galaxy's light falls into a single pixel. The G statistic is defined as

$$G = \frac{1}{\bar{i} n_{\rm seg}(n_{\rm seg} - 1)} \sum_k (2k - n_{\rm seg} - 1) i_{m_{(k)} n_{(k)}}\,,$$

where $\bar{i}$ is the sample mean of all intensities within the segmentation map and $m_{(k)} n_{(k)}$ denotes the coordinates of the pixel with the $k^{\rm th}$-smallest intensity value.

5. $\mathbf{M_{20}}$ statistic (Lotz et al. 2004)
M_{20} describes the spatial distribution of pixel intensities. First, we compute a total second-order moment:

$$M_{\rm tot} = \sum_m \sum_n i_{mn} \left[(m - m_{\rm cen})^2 + (n - n_{\rm cen})^2 \right]$$

where $m_{\rm cen}$ and $n_{\rm cen}$ are the coordinates of the galaxy's center of mass (equation 1) and the summation in done over all pixels within the segmentation map. We then repeat the summation done above using only the brightest 20% of the pixels; we call this sum $M_{\rm bright}$. M_{20} is then defined as

$$M_{20} = \log_{10}\left(\frac{M_{\rm bright}}{M_{\rm tot}}\right) .$$

6. **Concentration** (C) statistic (Conselice 2003)
   The concentration statistic encapsulates the area over which the bulk of a galaxy's summed intensity lies. Its calculation assumes circular symmetry. At a given radius r from the galaxy's center, we define two quantities: the summed intensity within the annulus defined by r and $r + dr$, and the overall average summed intensity:

$$\begin{aligned} \mu(r) &= \frac{\int_0^{2\pi}\int_{r-\delta r}^{r+\delta r} i(r',\theta) r' dr' d\theta}{\int_0^{2\pi}\int_{r-\delta r}^{r+\delta r} r' dr' d\theta} \\ \bar{\mu}(r) &= \frac{\int_0^{2\pi}\int_0^{r+\delta r} i(r',\theta) r' dr' d\theta}{\int_0^{2\pi}\int_0^{r+\delta r} r' dr' d\theta} . \end{aligned}$$

   (We show the calculations as integrals for conceptual clarity, but the actual calculations are done as sums over image pixels.) r' is the solution of the equation $\mu(r)/\bar{\mu}(r) = \epsilon$, where ϵ is commonly chosen to be 0.2. We compute the total summed intensity within the radius r', then determine the smaller radii within which there are 20% and 80% of that total summed intensity. The C statistic is:

$$C = 5 \times \log\left(r_{80\%}/r_{20\%}\right) .$$

   The smaller $r_{20\%}$ is relative to $r_{80\%}$, the higher the value of C, as the galaxy will appear "more concentrated."

7. **Asymmetry** (A) statistic (Conselice 2003)
   The A statistic is a measure of how asymmetric a galaxy is after its image is rotated 180° the central pixel and then subtracted from the original image. For an asymmetric galaxy, the difference image will exhibit significant residual structures, leading the A statistic to differ significantly from zero. The A statistic is defined as

$$A = \frac{\sum_m \sum_n |i_{mn} - i_{180,mn}|}{\sum_m \sum_n |i_{mn}|} - B_{180} ,$$

   where i and i_{180} are the pixel intensities in the original and rotated images respectively and B_{180} is the average background asymmetry, defined using the intensities of pixels lying outside the segmentation map.

8. **Generalized Concentration** (GC)
   Underlying the concentration statistic calculation is the assumption of circular apertures. The generalized concentration statistic does away with this assumption. We simply determine the smallest groups of pixels that contain 20% and 80% of the total intensity within the segmentation map:

$$GC = 5 \times \log\left(n_{80\%}/n_{20\%}\right) ,$$

where $n_{20\%}$ and $n_{80\%}$ are the numbers of pixels in each group, respectively.

GC is observed to be highly correlated with C, but outliers on the plot of GC vs. C may be observed when the segmentation map contains two (or more) fairly well-separated but still blended nuclei.

9. **Radius** (R)
   This is simply the radius of the circle (in arc-seconds) whose area matches that of the segmentation map, i.e.

$$R = P\sqrt{n_{\text{seg}}/\pi}$$

   where P is the plate-scale in arc-seconds/pixel.

10. **Signal-to-Noise** (SN)
    The median value of intensity over $\hat{\sigma}$ within the segmentation map.

11. **Sigma** ($sigma$)
    The estimate of $\hat{\sigma}$ made during segmentation map construction.

12. **Segmentation Map Size** ($segpix$)
    The number of pixels in the segmentation map. This quantity may be useful in identifying blended galaxies.

13. **Principal Axes** ($axmin$,$axmax$,$angle$)
    Standard estimates of semi-major and minor axes and rotation angle.

14. **Error Flags** ($errFlag$)

    - 0: no issue with statistic computation
    - 1: blank image
    - 2: segmentation map has width and/or height 1 pixel (which can occur for faint galaxies whose most intense pixels just peak above $n\hat{\sigma}$ in segmentation map construction)
    - 3: an exception was thrown by the code