IDL Assignment II

How many stars do we need to observe to distinguish between different model predictions of the metallicity distribution function?
Or more generally …
How many objects do we need to observe to distinguish between different probability distributions? (redshifts or other galaxy properties, black hole masses, exoplanets etc.)

Our clever and friendly neighbourhood theorists have models which predict the so-called "metallicity distribution function" (MDF), i.e., the number of stars at a given metallicity.

Here is one example from Salvadori (http://adsabs.harvard.edu/abs/2007MNRAS.381..647S)

658 *S. Salvadori, R. Schneider and A. Ferrara*

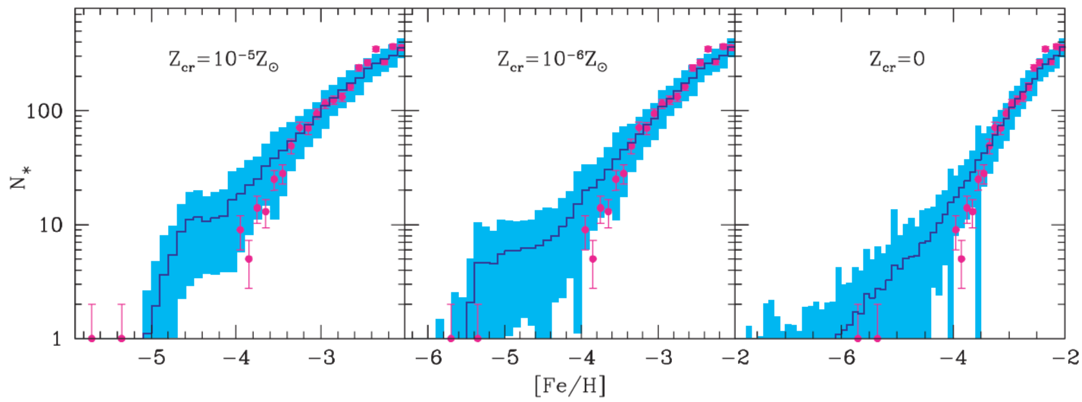
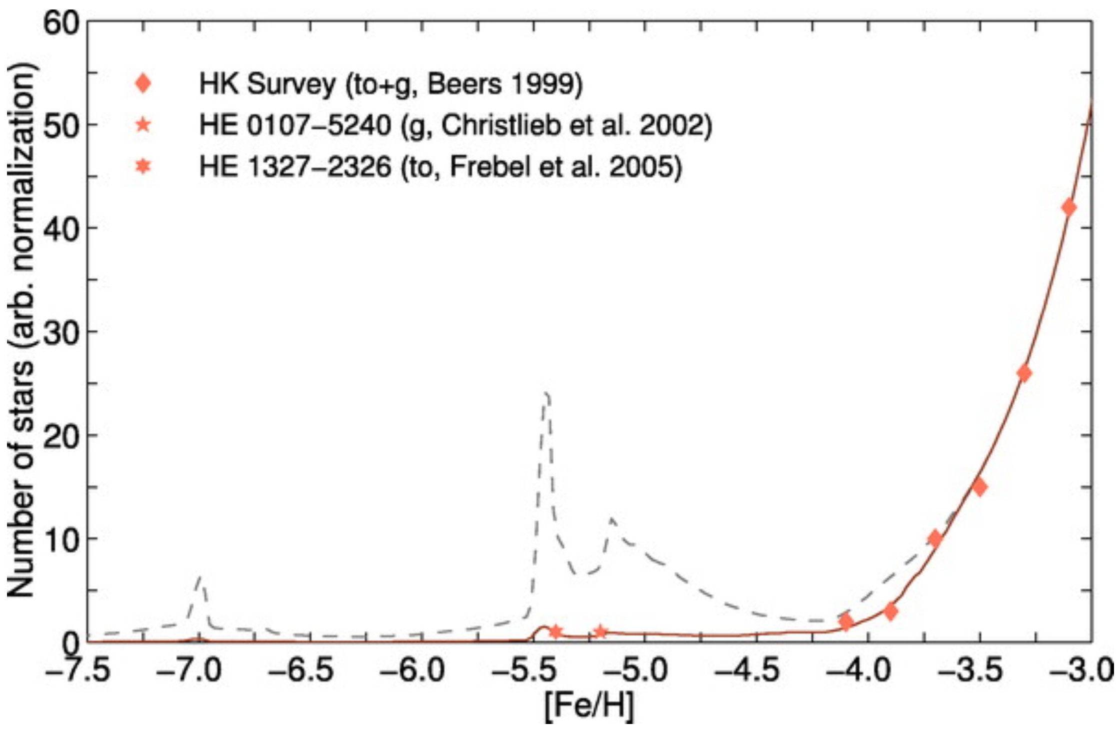


Figure 7. The same as the left-hand panel of Fig. 6 but for values of $Z_{cr} = 10^{-5}$, 10^{-6} and 0 $Z_{\odot}$.

Here is another example from Karlsson (http://adsabs.harvard.edu/abs/2006ApJ...641L..41K)



The input physics and assumptions in the models control the shape of the predicted metallicity distribution function. In particular, the so-called "critical metallicity" is a key quantity; it is the metallicity below which low-mass star formation is not possible and we would not expect to observe any long-lived low-mass stars today below that value (if you care about this, here is a review http://adsabs.harvard.edu/abs/2004ARA%26A..42...79B). Look again at that figure from Salvadori and notice how the critical metallicity affects the tail of the metallicity distribution function.

As simple-minded observers, we would like to know how many stars do we need to observe in order to discriminate between these models. Indeed, quantifying the necessary sample size to test a hypothesis (discriminating between model predictions in this case) is a crucial ingredient for successful proposals (observing/funding/jobs). Recall that 6-10 metre class optical/IR facilities (e.g., Magellan, Keck, Gemini, Subaru, VLT) cost ~AUD$1 per second(!) to operate and that the competition for telescope time is fierce.

Starting IDL
Type "idl" from the command line to start

Here is a possibly useful document
https://www.atmos.colostate.edu/programming/IDL/idl_week1.pdf

You can search/find IDL routines here
http://www.harrisgeospatial.com/docs/funclisting.html

In the dropbox link below, there is a **very** simple example (dy_example.pro)
To run this example, type ".r dy_example" from command line within IDL

The assignment

1. Here are the metallicity distribution function predictions in ascii format.
https://www.dropbox.com/sh/jd0y3msb6bvjl6i/AACNAv9Y4-lbTrIRgQKggGAta?dl=0
For the Salvadori models, the columns are
column 1 : [Fe/H]
column 2 : N*([Fe/H])
column 3 : SigmaN*([Fe/H])
For the Karlsson model, the columns are [Fe/H] and #.

Here is a relatively easy way to read ascii files in IDL
http://idlastro.gsfc.nasa.gov/ftp/pro/misc/readcol.pro
IDL array definitions
http://www.harrisgeospatial.com/docs/Creating_Arrays.html

2. Randomly draw 10 stars from each of these distributions
in the metallicity regime [Fe/H] ≤ −4.0
Note that these are arbitrary distributions (i.e., neither Gaussian nor uniform). Look at this document if you don't know how to generate random numbers from an arbitrary distribution.
www.ece.virginia.edu/mv/edu/prob/stat/random-number-generation.pdf

Generating random numbers in IDL
http://www.harrisgeospatial.com/docs/RANDOMN.html
http://www.harrisgeospatial.com/docs/RANDOMU.html
Are random numbers random?
http://www.idlcoyote.com/code_tips/randomnumbers.html

3. Compare any two distributions (e.g., Karlsson vs. Salvadori Z_crit=0) and quantify the likelihood that the data are drawn from the same distribution. Use your favourite statistical test; a well-known one is Kolmogorov-Smirnov (KS) test.
http://idlastro.gsfc.nasa.gov/ftp/pro/math/kstwo.pro

This might be a useful reference
sparky.rice.edu/astr360/kstest.pdf

What are the probabilities? Do this for all combinations of model distributions.

4. How reliable are those results? (If you are unsure, google "shot noise" or the Poisson distribution.) Repeat the above exercise 100,000 times and plot the distribution of the probabilities.

For what fraction of the realisations can you reject the null hypothesis at the 90% confidence level?

Writing loops in IDL
http://www.harrisgeospatial.com/docs/FOR.html

Plotting in IDL
http://www.harrisgeospatial.com/docs/PLOT_Procedure.html
http://idlastro.gsfc.nasa.gov/ftp/pro/plot/plothist.pro

5. Repeat the above steps but randomly drawing 100, 200, 300, 400 etc., stars from each distribution. Keep going (in steps of 100 stars) until you can reject the null hypothesis at the 95% confidence level in 95% of realisations.

Why are we doing this? My collaborators want to know how many stars do we need to observe in the metallicity regime [Fe/H] ≤ −4 in order to discriminate between the Salvadori and Karlsson models, and, between the various Salvadori models with different critical metallicities?