Master's thesis
Astronomy

Your Title Here

Anni Järvenpää

July 7, 2017

Tutor: Associate Professor Peter Johansson
Dr. Till Sawala

Censors: prof. Smith
doc. Smythe

UNIVERSITY OF HELSINKI
Department of Physics

PL 64 (Gustaf Hällströmin katu 2a)
00014 University of Helsinki

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty			Laitos — Institution — Department
Faculty of Science			Department of Physics

Tekijä — Författare — Author
Anni Järvenpää
Työn nimi — Arbetets titel — Title
Your Title Here
Oppiaine — Läroämne — Subject
Astronomy

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's thesis	July 7, 2017	0 pages

Tiivistelmä — Referat — Abstract

Abstract goes here.

Avainsanat — Nyckelord — Keywords

Your keywords here

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — övriga uppgifter — Additional information

Contents

1. Introduction

1.1 TL;DR version of prerequisite information

1. galaxies form
    - Why?
    - When?
    - How?
    - Where?
2. galaxies form in groups
3. our local group is one of these
4. something about large scale distribution of galaxies

1.2 History of Local Group Research

LG objects visible with naked eye -> realization they are something outside our galaxy -> realization they are something very much like our galaxy

First determining distance was difficult, now mass is more interesting question

1.3 Aim of This Thesis

Whatever the main results end up being, presented in somewhat coherent manner and hopefully sugar-coated enough to sound Important and Exciting.

2. Theoretical Background

Think whether LG or LCDM first

2.1 Local Group

Definition of galaxy group, our local group is one of these.

Mass estimate (Li, Yang masses for the LG and MW)

Maybe something about scale of things in our universe, what are galaxy groups made of, what do you get if you go one distance scale up, what's different in galaxy clusters

2.1.1 Structure

Galaxies that are part of LG, distribution of smaller ones around bigger ones

Current mass estimates (at least timing argument, hubble flow and maybe satellites)

2.1.2 Evolution

How have we ended up in a situation described earlier? What will happen in future?

2.2 Expanding universe

2.2.1 Discovery

Make maths, add cosmological constant, make observations, remove cosmological constant

Enough cosmology here or in other sections to make other parts of thesis to make sense and to suffice as master's thesis = basic textbook cosmology and galaxy formation theory

2.2.2 ΛCDM Cosmology

2.2.3 Hubble flow

What is, where seen, what means, how to measure, hotness/coldness

Plot: observations with fitted hubble flow

3. general simulation thingies

Data used here from EAGLE which uses modified GADGET-2 which is a tree-code that uses leapfrog

3.1 N-body simulations

3.1.1 Hierarchical Tree Algorithm

3.1.2 Numerical Integrators

3.1.3 Halo Finding with Subfind

3.2 Description of actual simulations used

Volume, number of particles, compare to other simulations, where better and where maybe worse

Resimulation of interesting regions

Simulation has same parameters as EAGLE 800 Mpc volume used schaye 2015 paper DM-only parts: Volker-Springer Gadget and Gadget 2 papers 1999 and 2005 or something, gravity part is more interesting than SPH Zooms can use multiple meshes, only one is used here gravitational softening

4. Mathematical and statistical methods

TODO: merkkaa heti alkuun kenen lähestymistapaa aiot seurailla

täällä tarvittavat esitiedot ja önnönnöö, listaa mm. mitä aiot kertoa kunhan tiedät itsekään

4.1 Statistical Background

vähän parempi tässä kuin aiemman otsikon alla

Precision of the used equipment limits accuracy of all data gathered from physical experiments, simulations or observations. Therefore the results are affected by the measurement process and the results have to be presented as estimates with some error, magnitude of which is affected by both number of data points and accuracy of the measurement equipment. (Bohm and Zech, 2010)

Estimating errors for measured quantities offers a way to test hypotheses and compare different experiments. This is done using different statistical methods, of which the most relevant for this thesis are covered here. The methods used in this work are shortly introduced in the following sections together with basic statistical concepts that are necessary to understand the methods. (Bohm and Zech, 2010)

4.1.1 Hypothesis testing and p-values

A common situation in scientific research is that one has to compare a sample of data points to either a model or another sample in order to derive a conclusion from the dataset. In statistics, this is known as hypothesis testing. For example, this can mean testing hypotheses such as ”these two variables are not correlated” or ”this sample is from a population with a mean of 1.0”. (J. V. Wall, 2003) Next paragraphs shortly introduce the basic concept of hypothesis testing and methods that can used to test the hypothesis ”these two samples are drawn from the same distribution”.

Typically the process of hypothesis testing begins with forming of a null hypothesis H_0 that is formatted such that the aim for the next steps is to either reject it or deduce that it cannot be rejected with a chosen significance level. Negation of the null hypothesis is often called research hypothesis or alternative hypothesis and denoted as H_1. For example, this can lead to H_0 ”this dataset is sampled from a normal distribution” and H_1 ”this dataset is not sampled from a normal distribution”. Choosing the hypothesis in this manner is done because often the research hypothesis is difficult to define otherwise. (Bohm and Zech, 2010; J. V. Wall, 2003)

After setting the hypothesis one must choose an appropriate test statistic. Ideally this is chosen such that the difference between cases H_0 and H_1 is as large as possible. Then one must choose the significance level α which corresponds to the probability of rejecting H_0 in the case where H_0 actually is true. This fixes the critical region i.e. the values of test statistic that lead to the rejection of the H_0. (Bohm and Zech, 2010; J. V. Wall, 2003)

miten α valitaan, millaiset arvot ovat tyypillisiä? miten α -> critical value -> rejection/no rejection. otoskoko = N mukaan tänne

It is crucial not to look at the test results before choosing α in order to avoid intentional or unintentional fiddling with the data or changing the criterion of acceptance or rejectance to give desired results. Only after these steps should the test statistic be calculated. If the test statistic falls within the critical region, H_0 should be rejected and otherwise stated that H_0 cannot be rejected at this significance level.

(Bohm and Zech, 2010; J. V. Wall, 2003)

This kind of probability based decision making is always prone to error. It is easy to see that α corresponds to the chance of H_0 being rejected when it is true. This is known as error of the first kind. However, this is not the only kind of error possible. It might also occur that H_0 is false but it does not get rejected, which is known as error of the second kind. (Bohm and Zech, 2010)

Despite statistical tests having a binary outcome "H_0 rejected" or "H_0 not rejected", a continuous output is often desired. This is what p-values are used for. The name p-value hints towards probability, but despite it's name p-value is not equal to the probability that the null hypothesis is true. These p-values are functions of a test statistic and the p-value for a certain value t_{obs} of a test statistic gives the probability that under the condition that H_0 is true, the value of a test statistics for a randomly drawn sample is at least as extreme as t_{obs}. Therefore if p-value is smaller than α, H_0 is to be rejected. (Bohm and Zech, 2010)

4.1.2 Distribution functions

ei hyvä, harkitse esim http://puppulause-generaattori.fi/?ava-insana=jakauma-funktio

Some statistical tests such as the Kolmogorov-Smirnov test and the Anderson-Darling test make use of distribution functions such as cumulative density function (CDF) and empirical distribution function (EDF) in determining the distribution from which a sample is drawn.

PDF määritelmä vaikea ymmärtää

To understand CDF and EDF, one must first be familiar with probability density function (PDF). As the name suggests, PDF is a function the value of which at some point x represents the likelihood that the value of the random variable would equal x. This is often denoted $f(x)$. Naturally for continuous functions the probability of drawing any single value from the distribution is zero, so these values should be interpreted as depicting relative likelihoods of different values. For example if $f(a) = 0.3$ and $f(b) = 0.6$ we can say that drawing value b is twice as likely as drawing value a. (Heino et al., 2012)

Another way to use the PDF is to integrate it over semi-closed interval from negative infinity to some value a to obtain the CDF, often denoted with $F(x)$:

$$F(x) = \int_{-\infty}^{x} f(x')\, dx'. \tag{4.1}$$

This gives the probability of a random value drawn from the distribution having value that is smaller than x. Relation between the PDF and the CDF is illustrated in figure 4.1, where PDFs and CDFs are shown for three different distributions. It is easy to see the integral relation between PDF and CDF and how wider distributions have wider CDFs. (Heino et al., 2012)

esittelet nyt nolosti EDF:n nimeltä kahdesti, mieti ratkaisu

Both the PDF and the CDF apply to whole population or the set of all possible outcomes of a measurement. In reality the sample is almost always smaller than this. Therefore one cannot measure the actual CDF. Nevertheless, it is possible to calculate a similar measure of how big a fraction of measurements falls under a given value. This empirical counterpart of the CDF is known as empirical distribution function (EDF), often denoted $\hat{F}(x)$, and for a dataset $X_1, X_2, ..., X_n$ containing n samples it is defined to be

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I[X_i \leq x] \tag{4.2}$$

where I is the indicator function, value of which is 1 if the condition in brackets is true, otherwise 0. (Feigelson and Babu, 2012)

Due to the EDF being a result of random sampling, it may deviate from the underlying CDF considerably as can be seen by comparing CDFs in figure 4.1 and corresponding EDFs in figure 4.2. The latter figure also has EDFs corresponding to two random samples drawn from the distribution of the green curve in the first figure to further illustrate the differences that can arise from random sampling. This randomness also makes determining whether two samples are drawn from the same distribution difficult.

lisää johonkin selitys normaalijakauman parametreille

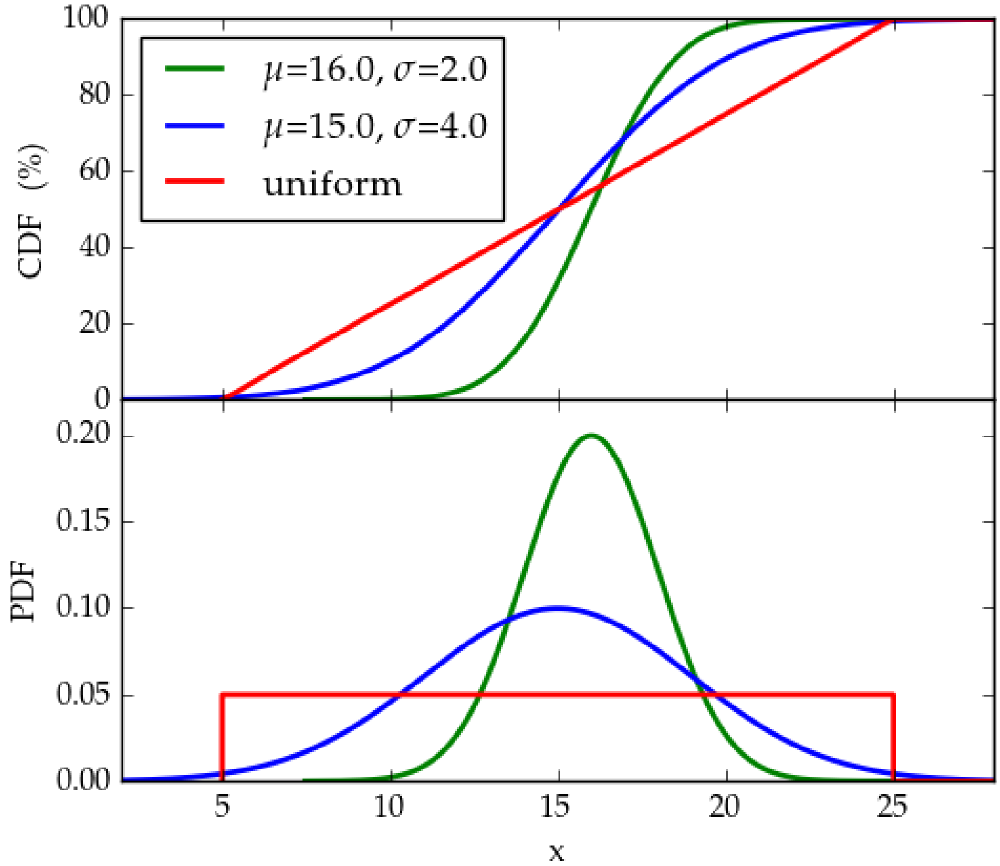

Figure 4.1: Cumulative distribution function (top panel) for three random samples (PDFs shown in the bottom panel) drawn from different distributions, two of which are normal and one is uniform.

kerro, miten sample size suhteutuu oikeaan käsiteltyyn dataan

4.2 Regression Analysis

line fitting and other trivial things

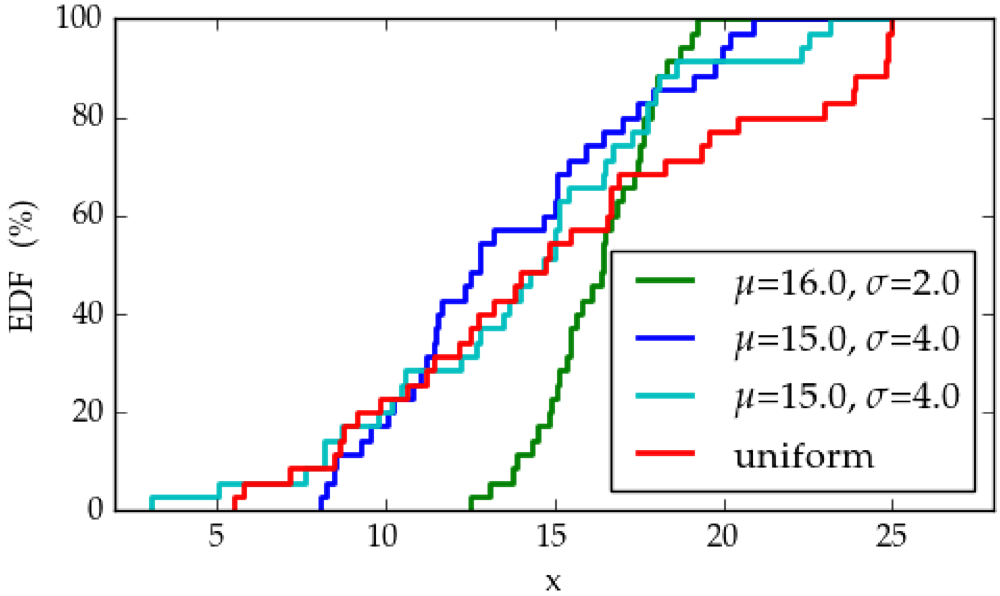

Figure 4.2: Empirical distribution function for four random samples (N=35) drawn from the same distributions as in figure 4.1. Note that both the blue and the cyan data are drawn from the same distribution.

4.3 Error analysis

TODO: oispa parempi otsikko. mieti, onko tämä muutenkaan hyvä nyt kun on siirretty yksi otsikkotaso ylöspäin

4.4 Comparing two samples drawn from unknown distributions

A common question in multiple fields of science is whether two or more samples are drawn from the same distribution. This can occur for example when comparing effectiveness of two procedures, determining if the instrument has changed over time or whether observed data is compatible with simulations. There are multiple two-sample tests that can address this kind of questions, e.g. χ^2, Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling tests. (Bohm and Zech, 2010; Feigelson and Babu, 2012)

selitä categorical data

In addition to comparing two samples, these tests can be used as one-sample tests to determine whether it is expected that the sample is from a particular distribution. However, some restrictions apply when using the one-sample variants. Some of these tests use categorical data, for example ”number of galaxies that are active” or ”number of data points between values 1.5 and 1.6” and compares numbers of samples in different categories, whereas the others are applied to numerical data and compare empirical distribution functions (EDF) of the datasets. (Feigelson and Babu, 2012)

4.4.1 χ^2 test

keksi paremmat esimerkit koko kappaleeseen, jotain relevanttia myöhempää tutkimusta ajatellen. katso kommentit paperista sen jälkeen, kaikkia ei täällä vielä

Astronomical data often involves classifying objects into categories such as ”stars with exoplanets” and ”stars without exoplanets” or the spectral classes of stars. One tool for analyzing such categorical data is χ^2 test. It can be used both to determine whether a sample can be drawn from a certain distribution and to test whether two samples can originate from a single distribution. (Corder, 2014; Feigelson and Babu, 2012)

For one-sample test, the χ^2 test uses the number of measurements in each bin together with a theoretical estimate calculated from the null hypothesis. For example one might have observed exoplanets and tabulated the number of planet-hosting stars of different spectral class as is shown in table 4.1 and now wants to test the observations against null hypothesis ”Distribution of stellar classes for observed exoplanet-hosting stars is equal to that of main sequence stars in solar neighbourhood as given by Ledrew (2001)” using significance level $\alpha = 0.01$. The data is categorical, so now χ^2 test is a justified choice. (Corder, 2014)

In this case the first step would be to calculate the expected observation counts for each bin according to the null hypothesis. Table 4.2 contains these expected counts (f_e) together with the observations (f_o). These observed and expected values

Stellar class	Number of observed planetary systems
A	6
F	38
G	39
K	134

Table 4.1: Example of categorical data.

Stellar class	Observations (f_o)	Theory (f_e)
A	6	6
F	38	28
G	39	71
K	134	112
total	217	217

Table 4.2: Data of table 4.1 together with expected values if null hypothesis was true.

are then used to calculate the χ^2 test statistic, defined as

$$\chi^2 = \sum_i \frac{(f_o - f_e)^2}{f_e}. \tag{4.3}$$

mistä kriittisiä arvoja voi lukea ja mistä sain omani, onhan selitetty (tässä tai) aiemmin

In this case, one gets $\chi^2 \approx 23.6$. The data has four bins, so the degree of freedom is $4-1=3$. Next one can compare the calculated χ^2 value to a tabulated critical value for our significance level $\alpha = 0.01$. These tabulated values can be widely found in statistics textbooks and books specifically dedicated to statistical tables. (Corder, 2014)

In this case according to Corder (2014) the critical value is 11.34, which means that one can reject the null hypothesis and conclude that at 1% significance level the distribution of stellar classes for observed exoplanet-hosting stars is not equal to that of main sequence stars in solar neighbourhood. This of course can either be due to exoplanets being more numerous around some stellar classes than others or arise from some observational effect such as the observer observing more of the later type stars and thus arbitrarily skewing the distribution of the exoplanet finds.

The χ^2 test can also be used to test for independence of two or more samples. The data is again tabulated and now the χ^2 test statistic is calculated as

$$\chi^2 = \sum_i \sum_j \frac{(f_{oij} - f_{eij})^2}{f_{eij}} \tag{4.4}$$

where f_{oij} denotes the observed frequency in cell (i, j) and f_{eij} is the expected frequency for that cell. The expected frequency can be calculated using the following formula

$$f_{eij} = \frac{R_i C_i}{N} \tag{4.5}$$

where R_i is the number of samples in row i, C_j is the number of samples in column j and N is the total sample size. (Corder, 2014)

According to Corder (2014), the degrees of freedom is $(R-1)(C-1)$ where R is the number of rows and C is the number of columns in tabulated data. This is true in many if not most cases, but the way of collecting data can affect the degrees of freedom in both one-sample and multi-sample cases. For example, if the one-sample model is not renormalized to fit the total number of observed events or in two-sample case the sample sizes differ the degrees of freedom equal to number of bins N_b instead of $N_b - 1$. (Press et al., 2007).

Before performing χ^2 test on a dataset, it is important to confirm that the data meets the assumptions for χ^2 test. First of all, the data has to consist of counts i.e. not for example percentages or fractions. These counts should be independent of each other and there has to be enough of them, generally > 50 is sufficient. Bins should also be chosen such that all bins have at least five counts according to the null hypothesis. If the last condition is not met, one can can consider combining bins. (Bock et al., 2014; Heino et al., 2012)

TODO: ei hyvä, pearsonista olisi hyvä sanoa jo ehkä alussa, ehkä vähän pidempi tuosta mitä muita vaihtoehtoja on -osasta

The method described above is sometimes referred to as Pearson's χ^2 test due to existence of other tests where χ^2 distribution is used. In some cases, such as with small 2×2 contingency tables and when expected cell counts are small, other variants of χ^2 test should be used. For example the Yates's χ^2 test or the Fisher exact test work better in these cases than the χ^2 test. (Corder, 2014)

4.4.2 Kolmogorov-Smirnov test

For astronomers one of the most well-known statistical test is the Kolmogorov-Smirnov test, also known as the KS test. It is computationally inexpensive to calculate, easy to understand and does not require binning of data. It is also a nonparametric test i.e. the data does not have to be drawn from a particular distribution. (Feigelson and Babu, 2012)

In the astrophysical context this is often important because astrophysical models usually do not fix a specific statistical distribution for observables and it is common to carry out calculations with logarithms of observables, after which the originally possibly normally distributed residuals will no longer follow a normal distribution. When using the KS test, the values on the x-axis can be freely reparametrized: for example using $2x$ or $\log x$ on x-axis will result in same value of the test statistic as using just x. (Feigelson and Babu, 2012; Press et al., 2007)

The test can be used as either one-sample or two-sample test, both of which are very similar. For two-sample variate the test statistic for the KS test is calculated based on empirical distribution functions $\hat{F}_1$ and $\hat{F}_2$ derived from two samples and the test statistic

$$D = \sup_x |\hat{F}_1(x) - \hat{F}_2(x)| \tag{4.6}$$

uses the maximum vertical distance of the EDFs. This test statistic is then used to

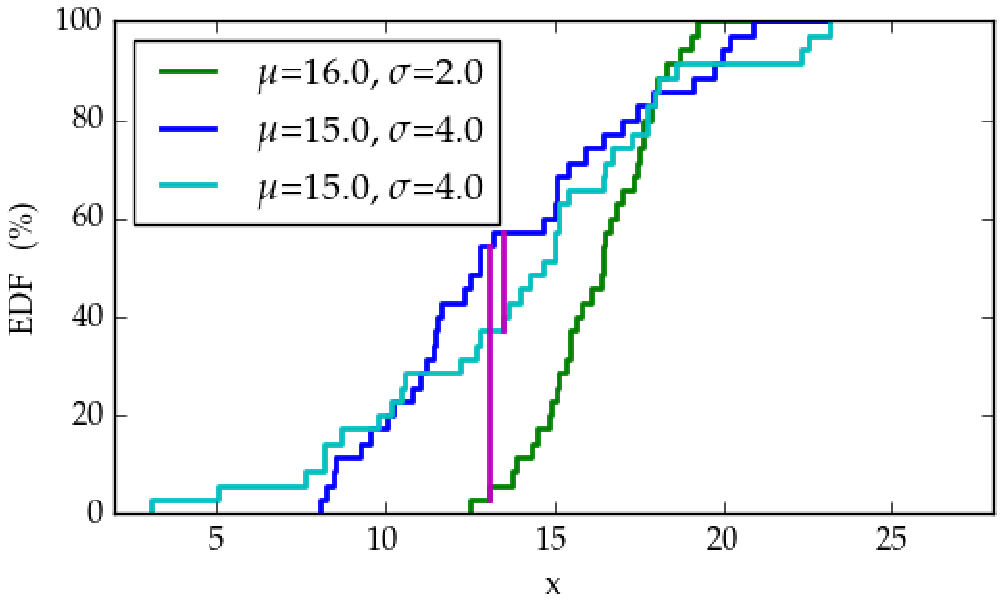

Figure 4.3: KS test parameter values (magenta vertical lines) shown graphically for three samples from figure 4.2.

determine the p-value and thus decide whether the null hypothesis can be rejected. For one-sample variate the procedure is similar, but EDF $\hat{F}_2$ is substituted with the CDF that corresponds to the null hypothesis. (Bohm and Zech, 2010; Feigelson and Babu, 2012)

As an example, let's consider two pairs of samples from figure 4.2: green and blue (two samples drawn from different normal distributions) and blue and cyan (two samples drawn from same normal distribution). We can formulate the test and null hypotheses for both pairs as H_0="the two samples are drawn from the same distribution" and H_1="the two samples are not drawn from the same distribution" and choose a significance level of for example $\alpha = 0.05$ or $\alpha = 0.01$.

mistä p-value saadaan, kerro taas aiemmin (tai täällä)

The test statistic is then calculated and for these samples we get $D = 0.51$ for the green-blue pair and $D = 0.20$ for the blue-cyan pair. Test statistics are illustrated in figure 4.3 where the test statistics D are shown as vertical magenta lines. These values of D correspond to p-values 9.9×10^{-5} and 0.44 respectively,

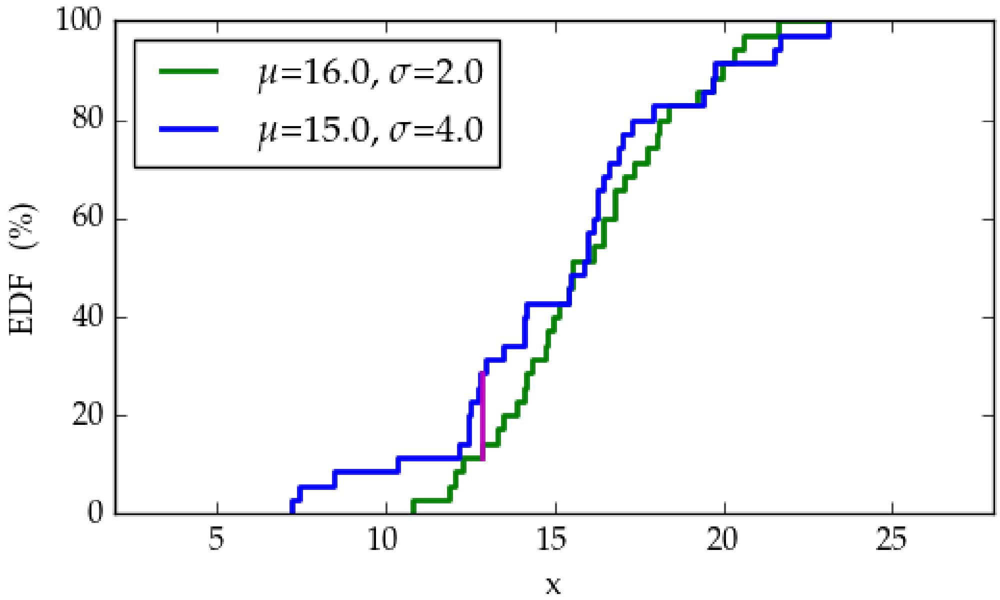

Figure 4.4: KS test ran on another pair of samples drawn from blue and green distributions in figure 4.1.

which means that the null hypothesis ”green and blue samples are drawn from the same distribution” is rejected at both 0.05 and 0.01 significance levels but the null hypothesis ”blue and cyan samples are drawn from the same distribution” cannot be rejected.

In this case the KS test produced result that matches the actual distributions from which the samples were drawn. Using a different random realization might have resulted in a different conclusion, for example one shown in figure 4.4 results in $D = 0.17$ that corresponds to a p-value of 0.64 i.e. null hypothesis could not have been rejected using the α specified earlier. In a similar manner there can be cases where two samples from one distribution are erroneously determined not to come from the same distribution if the samples differ from each other enough due to random effects.

The latter example case also illustrates one major shortcoming of the KS test: it is not very sensitive to small-scale differences near the tails of the distribution.

For example in figure 4.4 the blue sample goes much further left, but because EDF is always zero at the lowest allowed value and one at the highest one the vertical distances near the tails are small and the test is most sensitive to differences near the median value of the distribution. On the other hand, the test performs quite well when the samples differ globally or have different means. (Feigelson and Babu, 2012)

The KS test is also subject to some limitations and it is important to be aware of them in order to avoid misusing it. First of all, the KS test is not distribution free "explain better" if the model parameters, e.g. mean and standard deviation for normal distribution, are estimated from the dataset that is tested. Thus the tabulated critical values can be used only if model parameters are determined from some other source such as a simulation, theoretical model or another dataset. (Feigelson and Babu, 2012)

Another severe limitation of KS test is that it is only applicable to one-dimensional data. If the dataset has two or more dimensions, there is no unique way of ordering the points to plot EDF and therefore if KS test is used, it is no longer distribution free. Some variants that can handle two or more dimensions have been invented, such as ones by Peacock (1983) and Fasano and Franceschini (1987), but the authors do not provide formal proof of validity of these tests. Despite this, the authors claim that Monte Carlo simulations suggest that the methods work adequately well for most applications. (Press et al., 2007)

4.4.3 Other tests based on EDFs

ehkä vähän lyhyenpuoleisia kappaleita

Unsatisfactory sensitivity of the KS test motivates the use of other more complex tests. Such tests are for example the Cramér-von Mises test (CvM) and Anderson-Darling (AD) test, both of which have their strengths. Similar to KS test, both of these can be used as one-sample or two-sample variants. (Bohm and Zech, 2010; Feigelson and Babu, 2012)

First of these tests integrates over the squared difference between the EDF of the sample and CDF from the model or two EDFs in case of two-sample test. The test statistic W^2 for one-sample case can be expressed formally as

$$W^2 = \int_{-\infty}^{\infty} [\hat{F}_1(x) - F_0(x)]^2 \, dF_0(x) \tag{4.7}$$

For two-sample version, the theoretical CDF F_0 has to be replaced with another empirical distribution function $\hat{F}_2$. (Bohm and Zech, 2010; Feigelson and Babu, 2012)

Due to integration, the CvM test is able to differentiate distributions based on both local and global differences, which causes it to often perform better than the KS test. Similar to the KS test, the CvM test also suffers from EDFs or an EDF and a CDF being equal at the ends of the data range, which again makes the test less sensitive to differences near the tails of the distribution. (Feigelson and Babu, 2012)

In order to achieve constant sensitivity over the entire range of values, the statistic has to be weighted according to the proximity of the ends of the distribution. The AD test does this with its test statistic defined as

$$A^2 = N \int_{-\infty}^{\infty} \frac{[\hat{F}_1(x) - F_0(x)]^2}{F_0(x)[1 - F_0(x)]} \, dF_0(x) \tag{4.8}$$

where N is the number of data points in sample. This weighing makes the test more powerful than the KS and CvM tests in many cases. (Bohm and Zech, 2010; Feigelson and Babu, 2012)

hnnngh Also other more specific tests exist, such as the Kuiper test which is well suited for cyclic measurements. The test should always be chosen to match the dataset such that it best differentiates between the null and research hypotheses. (Bohm and Zech, 2010; Feigelson and Babu, 2012)

4.5 Cluster Analysis

DBSCAN

5. Findings from DMO Halo Catalogue Analysis

5.1 Selection of Local Group analogues

criteria, how many found, what are like (some plots maybe? distributions of masses, separations, velocity components, number of subhaloes within some radius or correlations between two of those?). Some of this might be part of previous chapter too (relevant to resimulation)?

5.2 Hubble Flow Measurements

HF, local H_0, H_0 within shells, zero-point, are previous consistent with what went into the simulation

5.3 Anisotropy of Hubble flow

huom obs nb: isotropy + randomness or anisotropy? esittele konsepti. plots: see notebook last pages

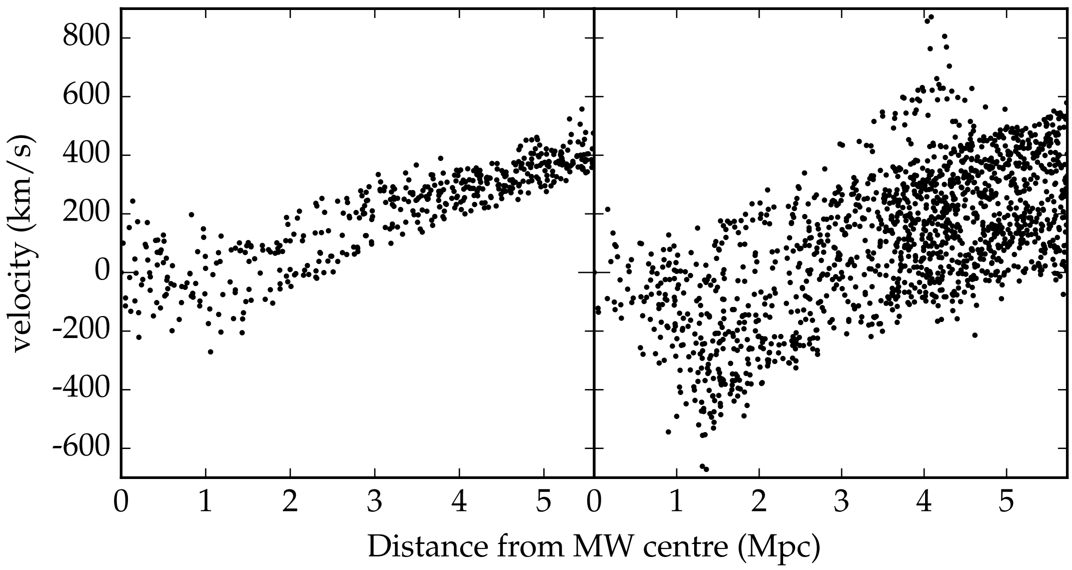

Figure 5.1: Hubble Flows around Milky Way in two simulations.

5.4 Statistical Estimate of the Local Group Mass

Analysis similar to Fattahi et al 2016 paper

6. Conclusions

Bibliography

D. Bock, P. Velleman, and R De Veaux. *Stats: Modeling the World*. Pearson, third edition edition, 2014.

G. Bohm and G. Zech. *Introduction to statistics and data analysis for physicists*. DESY, 2010. ISBN 9783935702416. URL http://www-library.desy.de/preparch/books/vstatmp_engl.pdf.

Gregory W. Corder. *Nonparametric statistics : a step-by-step approach*. Wiley, Hoboken, New Jersey, second edition edition, 2014. URL http://login.libproxy.helsinki.fi/login?url=http://site.ebrary.com/lib/helsinki/Doc?id=10885010.

G. Fasano and A. Franceschini. A multidimensional version of the Kolmogorov-Smirnov test. *Monthly Notices of the Royal Astronomical Society*, 225:155–170, March 1987. doi: 10.1093/mnras/225.1.155.

Eric D. Feigelson and G. Jogesh Babu. *Modern Statistical Methods for Astronomy: With R Applications*. Cambridge University Press, 2012. doi: 10.1017/CBO9781139015653.

R. Heino, K. Ruosteenoja, and J. Räisänen. *Havaintojen tilastollinen käsittely*. Department of Physics (University of Helsinki), 2012.

C. R. Jenkins J. V. Wall. *Practical Statistics for Astronomers*. Cambridge Observing

Handbooks for Research Astronomers. Cambridge University Press, illustrated edition edition, 2003. ISBN 9780521454162,0521454166.

G. Ledrew. The Real Starry Sky. *Journal of the Royal Astronomical Society of Canada*, 95:32, February 2001.

J. A. Peacock. Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices of the Royal Astronomical Society*, 202:615–627, February 1983. doi: 10.1093/mnras/202.3.615.

William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, editors. *Numerical recipes: The art of scientific computing*. Cambridge University Press, New York, third edition edition, 2007.