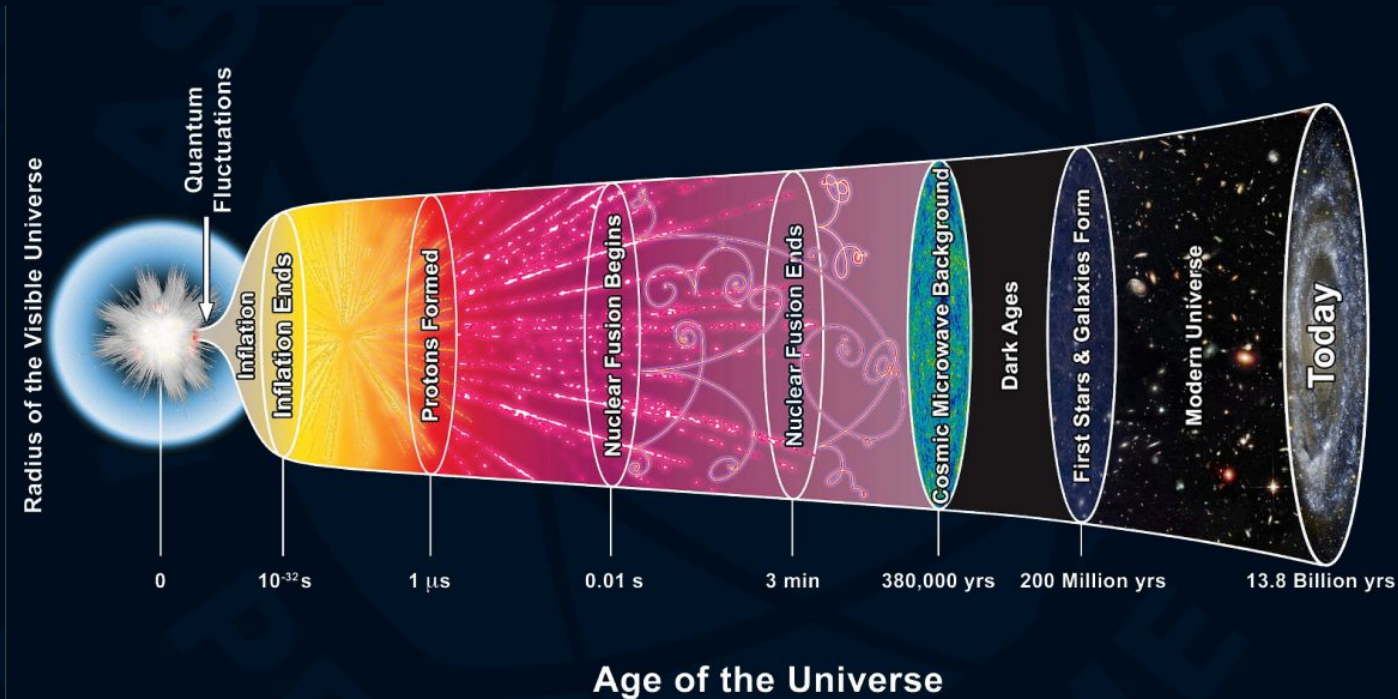


Cosmology Big Data

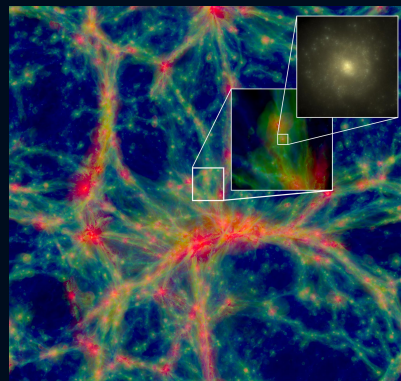
Rens Verkade

Cosmology in a nutshell



Cosmological model

- Current 'standard' Λ CDM model
- A universe with Cold Dark Matter and a cosmological constant Λ
- Cold Dark matter allows for bottom-up clustering
- The Cosmological constant plays an important role in the expansion of the universe



EAGLE simulation

Standard candles

- Type I A Supernovae
- White dwarf in a binary system (usually with a red giant)



Nasa

Standard candles

- Type I A Supernovae
- White dwarf in a binary system (usually with a red giant)
- Accretion until $M_{WD} \simeq 1.43M_{\odot}$



Nasa

Standard candles

- Type I A Supernovae
- White dwarf in a binary system (usually with a red giant)
- Accretion until $M_{WD} \simeq 1.43M_{\odot}$
- The Chandrasekhar limit where $R_{WD} = 0$
- White dwarf contracts heavily and a runaway thermo-nuclear reaction produces the supernova

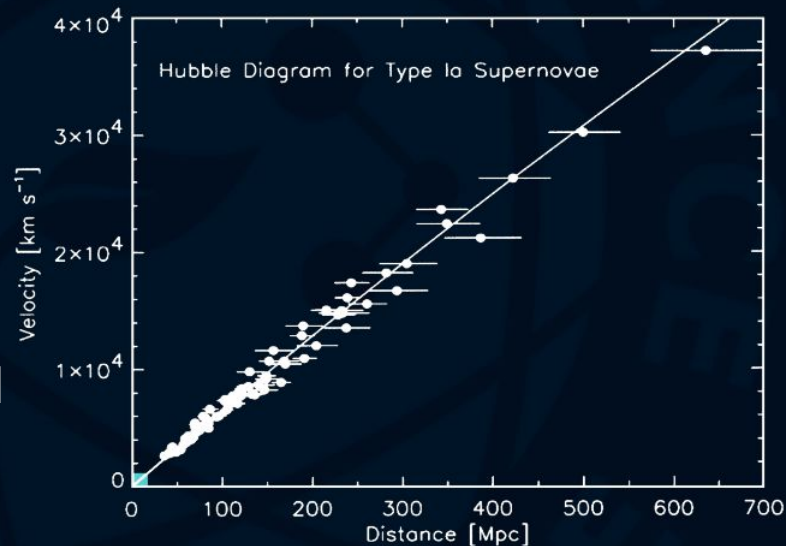


Nasa

An expanding universe

- Type Ia Supernovae have the same peak absolute magnitude
- Difference to observed magnitude gives distance
- Velocity from redshift of peak and spectral lines
- Hubble finds a linear relation!

$$v = H_0 d$$



E. Hubble, PNAS March, 1929, 15 (3) 168-174

Hubble parameter

- The Hubble rate parameterises the growth/shrinkage of the universe

$$H(t) = \frac{\dot{a}}{a} \quad a \text{ Scale factor,} \quad a(0) = 1$$

Hubble parameter

- The Hubble rate parameterises the growth/shrinkage of the universe

$$H(t) = \frac{\dot{a}}{a} \quad a \text{ Scale factor,} \quad a(0) = 1$$

- Little h is an often used unitless version $h \equiv \frac{H_0}{100 \text{ km/s/Mpc}}$

Hubble parameter

- The Hubble rate parameterises the growth/shrinkage of the universe

$$H(t) = \frac{\dot{a}}{a} \quad a \text{ Scale factor,} \quad a(0) = 1$$

- Little h is an often used unitless version $h \equiv \frac{H_0}{100 \text{ km/s/Mpc}}$
- Depends on the cosmological contents

$$H(z) = H_0 \sqrt{\Omega_\Lambda + \Omega_k(1+z)^2 + \Omega_m(1+z)^3 + \Omega_r(1+z)^4}$$

$$\Omega_i = \frac{\rho_i}{\rho_{crit}} \quad \Omega_k = 1 - \Omega_m - \Omega_r - \Omega_\Lambda$$

Some statistical tools

- Probability of B, given A $P(B|A)$

Some statistical tools

- Probability of B, given A $P(B|A)$
- Our goal is how likely the values of some parameters θ are, given how well they reproduce the data D.
- Bayes' theorem
$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Some statistical tools

- Probability of B, given A $P(B|A)$
- Our goal is how likely the values of some parameters θ are, given how well they reproduce the data D.
- Bayes' theorem
$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$
- Posterior $P(\theta|D)$
- Likelihood $P(D|\theta)$

Some statistical tools

- Probability of B, given A $P(B|A)$
- Our goal is how likely the values of some parameters θ are, given how well they reproduce the data D.
- Bayes' theorem $P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$
- Posterior $P(\theta|D)$
- Likelihood $P(D|\theta)$
- Prior $P(\theta)$
- Evidence $P(D)$

Some statistical tools

- The statistic we are interested in depends on what you want to know
- Posterior tells you how likely your theory model is given the data $P(\theta|D)$

Some statistical tools

- The statistic we are interested in depends on what you want to know
- Posterior tells you how likely your theory model is given the data $P(\theta|D)$
- In discoveries we are more interested in the likelihood $P(D|\theta)$
- Given your hypothesised model, probability of finding the data you have measured. Often we first predict and then measure. (Higgs!)

Bayes' Theorem in practice

- Let's say you get tested for some disease
- It's rare so 1 in 100 people with the symptoms actually have the disease

*Example from Megan Bell
[https://tomrocksmaths.com/
2021/08/31/bayes-theorem-
and-disease-testing/](https://tomrocksmaths.com/2021/08/31/bayes-theorem-and-disease-testing/)*

Bayes' Theorem in practice

- Let's say you get tested for some disease
- It's rare so 1 in 100 people with the symptoms actually have the disease
- The test is positive, what is the probability you have the disease

*Example from Megan Bell
[https://tomrocksmaths.com/
2021/08/31/bayes-theorem-
and-disease-testing/](https://tomrocksmaths.com/2021/08/31/bayes-theorem-and-disease-testing/)*

Bayes' Theorem in practice

- Let's say you get tested for some disease
- It's rare so 1 in 100 people with the symptoms actually have the disease
- The test is positive, what is the probability you have the disease
- Not $1/100$, we have additional info, the test

Example from Megan Bell
[https://tomrocksmaths.com/
2021/08/31/bayes-theorem-
and-disease-testing/](https://tomrocksmaths.com/2021/08/31/bayes-theorem-and-disease-testing/)

Bayes' Theorem in practice

- We need to first know the test's performance

Bayes' Theorem in practice

- We need to first know the test's performance
- Test is 99% sensitive, if you have the disease and get tested, 99% of the tests are positive (True positive)

Bayes' Theorem in practice

- We need to first know the test's performance
- Test is 99% sensitive, if you have the disease and get tested, 99% of the tests are positive (True positive)
- Test is 94% specific, if you don't have the disease and get tested, 94% of tests are negative (True Negative)

Bayes' Theorem in practice

- We need to first know the test's performance
- Test is 99% sensitive, if you have the disease and get tested, 99% of the tests are positive (True positive)
- Test is 94% specific, if you don't have the disease and get tested, 94% of tests are negative (True Negative)

	Have the disease	Don't have the disease
Test positive	99%	6%
Test negative	1%	94%

Bayes' Theorem in practice

- If 100 people get tested, we will have 1 true positive and 6 false positive. So probability to have the disease with a positive test $1/7$

Bayes' Theorem in practice

- If 100 people get tested, we will have 1 true positive and 6 false positive. So probability to have the disease with a positive test $1/7$
- Check using Bayes' theorem
$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Bayes' Theorem in practice

- If 100 people get tested, we will have 1 true positive and 6 false positive. So probability to have the disease with a positive test 1/7

- Check using Bayes' theorem
$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

$$P(\text{sick} | t = +) = \frac{P(t = + | \text{sick})P(\text{sick})}{P(t = +)}$$

Bayes' Theorem in practice

- If 100 people get tested, we will have 1 true positive and 6 false positive. So probability to have the disease with a positive test 1/7

- Check using Bayes' theorem $P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$

$$P(\text{sick} | t = +) = \frac{P(t = + | \text{sick})P(\text{sick})}{P(t = +)}$$

$$P(\text{sick} | t = +) = \frac{P(t = + | \text{sick})P(\text{sick})}{P(t = + | \text{sick})P(\text{sick}) + P(t = + | \text{not sick})P(\text{not sick})}$$

Bayes' Theorem in practice

$$P(\text{sick} \mid t = +) = \frac{P(t = + \mid \text{sick})P(\text{sick})}{P(t = +)}$$

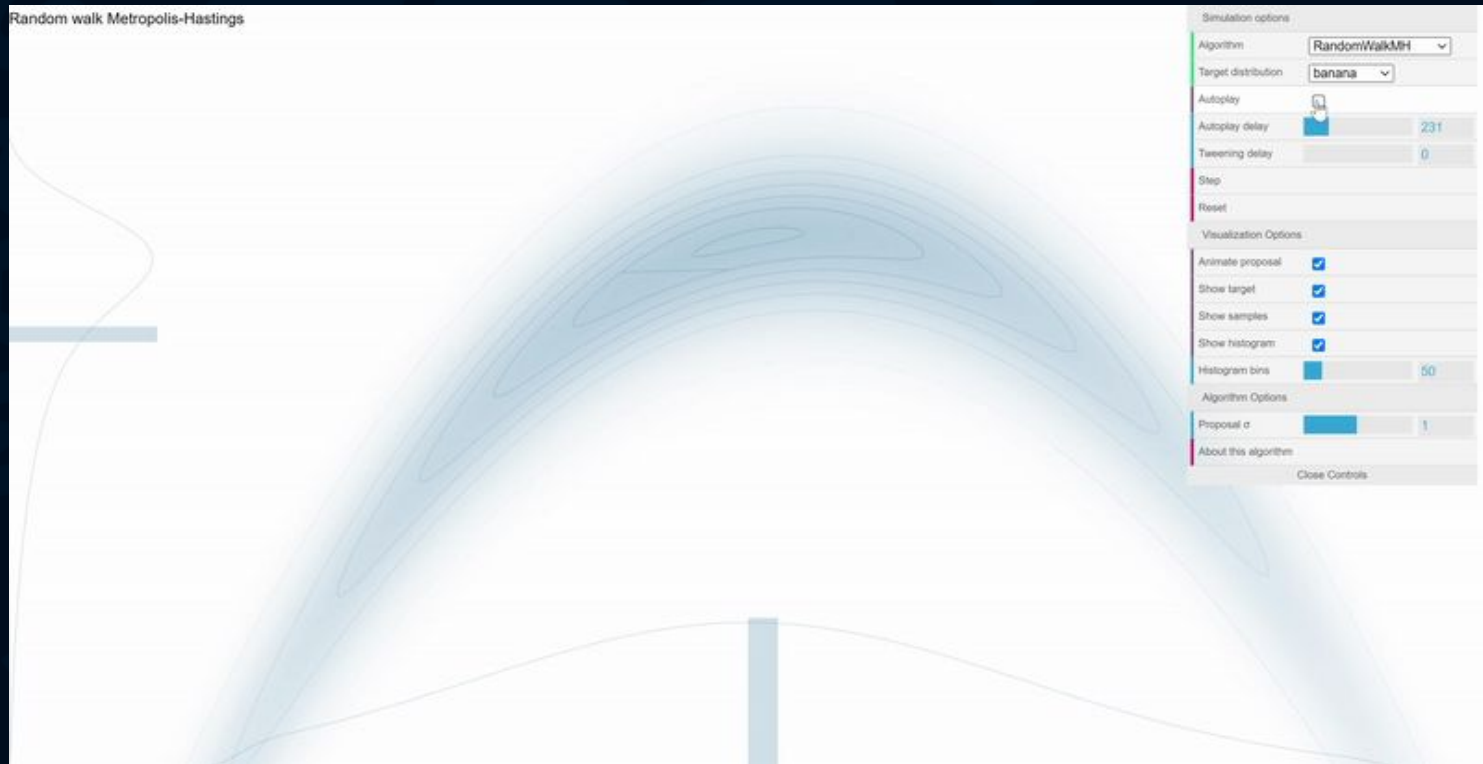
$$P(\text{sick} \mid t = +) = \frac{P(t = + \mid \text{sick})P(\text{sick})}{P(t = + \mid \text{sick})P(\text{sick}) + P(t = + \mid \text{not sick})P(\text{not sick})}$$

$$P(\text{sick} \mid t = +) = \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.06 \times 0.99} = \frac{1}{7}$$

Monte Carlo Markov Chain (MCMC)

- MCMC is one way to sample the probability distribution $P(\theta|D)$
- The algorithm steps randomly and checks if the new Θ is better (χ^2 - test)
- If the step is better we keep the new location, if not try again

Monte Carlo Markov Chain (MCMC)



<https://chi-feng.github.io/mcmc-demo/>

Monte Carlo Markov Chain (MCMC)

- MCMC is one way to sample the probability distribution $P(\theta|D)$
- The algorithm steps randomly and checks if the new Θ is better (χ^2 - test)
- If the step is better we keep the new location, if not try again
- Stop after n steps, or preferably after some convergence test is met

Hand-in

- Work on the cluster <https://jupyterhubroute-pra3024-2025.apps.dsri2.unimaas.nl/>
- Notebook is on Canvas and due DATE at 00:00
 - needs to be compiled (so output visible)
 - needs to be runnable (without extra installations)
 - Have comments AND Markdown (graded)
 - Use of sources is fine but give proper references otherwise it's plagiarism and you fail the exercise
 - Use of AI is allowed but understand the code so explain properly what you do in the markdown cells!
 - Check rubric on canvas to see how you are graded
- If questions arise email ON TIME
 - rens.verkade@maastrichtuniversity.nl
 - george.scriven@maastrichtuniversity.nl





$$P(\text{have disease}|\text{positive test}) = \frac{P(\text{positive test}|\text{have disease})P(\text{have disease})}{P(\text{positive test})}$$

$$P(\text{have disease}|\text{positive test}) = \frac{P(\text{positive test}|\text{have disease})P(\text{have disease})}{P(\text{positive test}|\text{have disease})P(\text{have disease}) + P(\text{positive test}|\text{no disease})P(\text{no disease})}$$