Physics 361 - Machine Learning in Physics

Lecture 3 – Background

Jan. 30th 2024

Moritz Münchmeyer

Unit 1: Background

1.2 Classical Statistics and Data Analysis Background

Sources:
- Cowan - Statistical data analysis
- also mostly covered in deeplearningbook.org

Likelihood, Posterior, Prior

- The Likelihood is the probability of measuring data $\vec{d}$ given a model M with parameters $\vec{\lambda}$.

$$\mathcal{L}(\vec{d} \mid M, \vec{\lambda})$$

often not explicitly written

often used as the loss function in macine learning.

- The Posterior is the probability of parameters $\vec{\lambda}$ given observed data $\vec{d}$.

$$P(\vec{\lambda}, M \mid \vec{d})$$

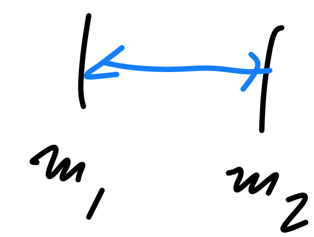
e.g M: standard model
$\vec{\lambda}$: Higgs mass
$\vec{d}$: collision data

- We can use Bayes theorem to get P from L

$$P(\vec{\lambda} \mid \vec{d}) = \frac{\mathcal{L}(\vec{d} \mid \vec{\lambda}) \, P(\vec{\lambda})}{P(\vec{d})}$$

- The prior $P(\vec{\lambda})$ is the probability of $\vec{\lambda}$ before we perform the measurement.

  E.g. • flat in some interval



  • Gaussian around some prior measurement.

  Prior is important if the data is not very constraining.

- Finally we have the evidence

  $P(\vec{d})$ is the probability of the data under model M for ANY parameters $\vec{\lambda}$.

$$P(\vec{d}) = \int \mathcal{L}(d|\vec{\lambda})\, P(\vec{\lambda})\, d\vec{\lambda}$$

Often used in model comparison

Maximum Likelihood and Maximum Aposteriori

- Combining the concepts of estimators and likelihoods we define the

  Maximum Likelihood estimator

$$\hat{\lambda}_{ML} = \underset{\lambda}{\mathrm{argmax}}\ \mathcal{L}(\vec{d} \mid \vec{\lambda})$$

  sometimes we can do it analytically:

$$\left.\frac{\partial \mathcal{L}(\vec{d} \mid \lambda)}{\partial \lambda}\right|_{\lambda = \hat{\lambda}} = 0$$

  If not analytically tractable use an optimizer.

- We define the maximum aposteriori estimator (MAP)

$$\hat{\lambda}_{MAP} = \underset{\lambda}{\mathrm{argmax}}\ P(\lambda \mid \vec{d})$$

Gaussian Likelihood

Gaussian Likelihoods are often good approx. of data in physics.

Example: We measure a person's weight w. To get an uncertainty we measure 100 times.

We assume that the data is described by $d_i = w + n_i$

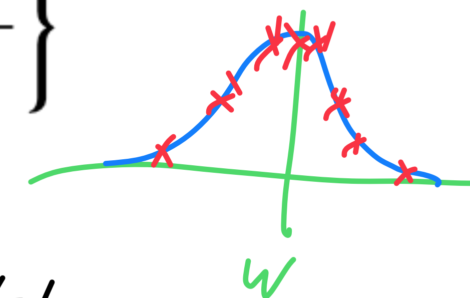
truth

Gaussian noise with variance σ_w^2

We want to measure the parameters w and σ_w^2.

For a single observation the likelihood is:

$$\mathcal{L}(d|w,\sigma_w) \equiv P(d|w,\sigma_w) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left\{-\frac{(d-w)^2}{2\sigma_w^2}\right\}$$



For m independent measurements $\mathcal{L}$ is the product:

$$\mathcal{L}\left(\{d_i\}_{i=1}^{m} \middle| w, \sigma_w\right) = \frac{1}{(2\pi\sigma_w^2)^{m/2}} \exp\left\{-\frac{\sum_{i=1}^{m}(d_i - w)^2}{2\sigma_w^2}\right\}$$

We get the posterior from Bayes theorem:

$$P(w,\sigma_w | \{d_i\}) = \frac{\mathcal{L}(\{d_i\}|w,\sigma_w)\, P(\lambda)}{P(\{d_i\})}$$

If the prior is flat we can work with the maximum likelihood estimator.

The max. likelihood estimator for w is:

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{\overset{=0}{\sum_{j=1}^{m}(d_j - w)}}{\sigma_w^2 (2\pi\sigma_w^2)^{m/2}} \overbrace{\exp\left\{-\frac{\sum_{i=1}^{m}(d_i - w)^2}{2\sigma_w^2}\right\}}^{\neq 0}$$

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \quad \Leftrightarrow \quad \sum_{j=1}^{m}(d_j - w) = 0$$

solve for w: $\boxed{w = \hat{w} = \frac{1}{m}\sum_{i=1}^{m} d_i}$

This is the mean as expected.

We could also do $\frac{\partial \mathcal{L}}{\partial \sigma_w}$ to find $\hat{\sigma}_w$.

While here the result is „obvious", this is the general procedure to find estimators given a model (= Likelihood).

Sampling the posterior : MCMC

- We have learned how to get the posterior for parameters $\vec{\lambda}$ given data $\vec{d}$.
- It is given by $$P(\lambda|d) = \frac{P(d|\lambda)\,P(\lambda)}{P(d)}$$

  Often we only need the unnormalized posterior
  $$P(\lambda|d) \propto P(d|\lambda)\,P(\lambda)$$
  The evidence does not depend on $\vec{\lambda}$ and can thus be ignored when finding λ.
- There is still a problem to deal with: computational limits

Example: Assume $P(\lambda|\vec{d})$ depends on 20 parameters.

How do we work with such a high dimensional posterior?

If we evaluate it on a grid with 100 points per dimension we'd need

$$(100)^{20} \text{ evaluations.}$$

This cannot be done.

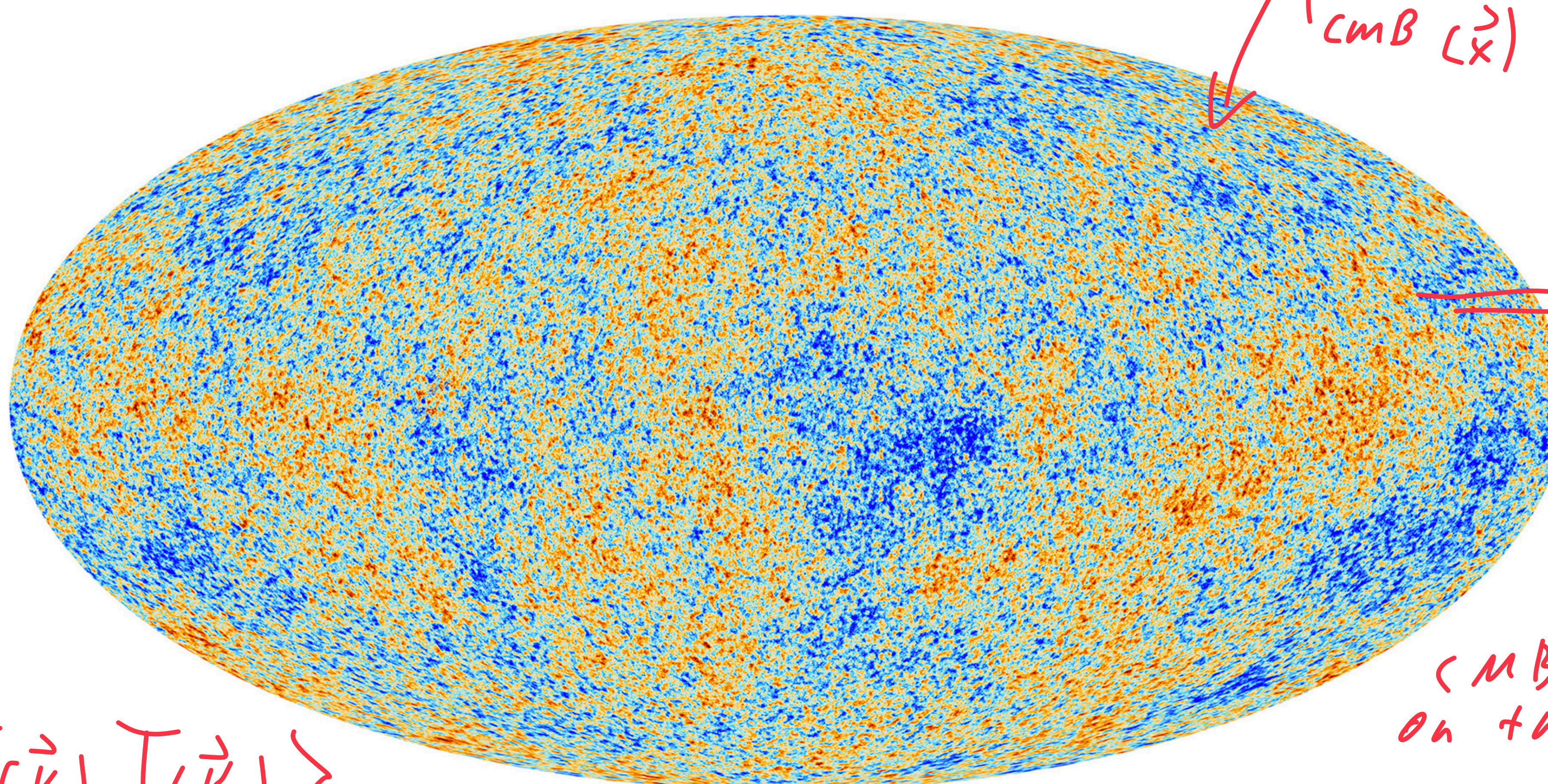
Instead one works with such posteriors by sampling from them using a method called Markov Chain Monte Carlo (MCMC)

A physics application of these concepts: The CMB power spectrum

- I want to show you an application of these ideas to real physics. The CMB power spectrum analysis is one of the jewels of physics, telling us much of what we know about the history of the universe.
- Of course this is a complicated topic and I can only give you a brief idea.
- The Cosmic Microwave Background is a radiation that permeates the universe. It has a temperature of about 3K and has been measure extremely precisely.

$T_{CMB}(\vec{x})$

From $T(\vec{x})$ we can get the Fourier transform $T(\vec{k})$



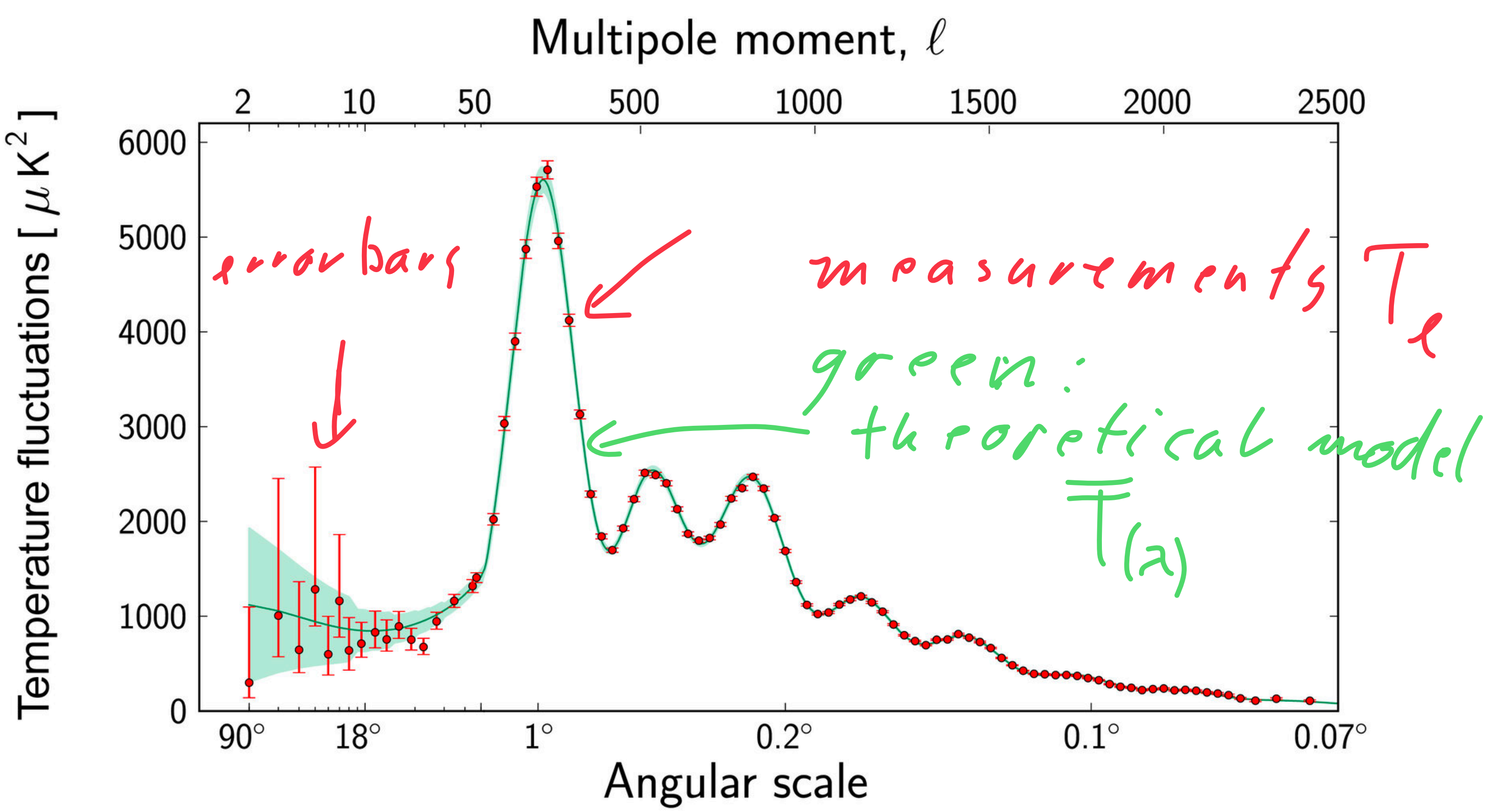
$\Longrightarrow P(k)$ power spectrum

CMB temperat. on the sky 3K

$$P(k) = \langle T(\vec{k})\, T(\vec{k}) \rangle$$

The dataset

- Our measurements will be the power spectrum as a function of scale. It can be extracted from the previous map by taking its Fourier transform and squaring the mode amplitudes.
- Our data points are:



data set: $\{T_\ell\}$ with some error bars

The likelihood

- We will make a Gaussian likelihood.
- Our theory model is that the mean curve (green) is a known function of cosmological parameters, which we want to measure.

Theoretical model (green curve):

$\bar{T}_\ell(\vec{\lambda})$ depends on cosmological parameters, e.g:

Ω_c : Dark matter density

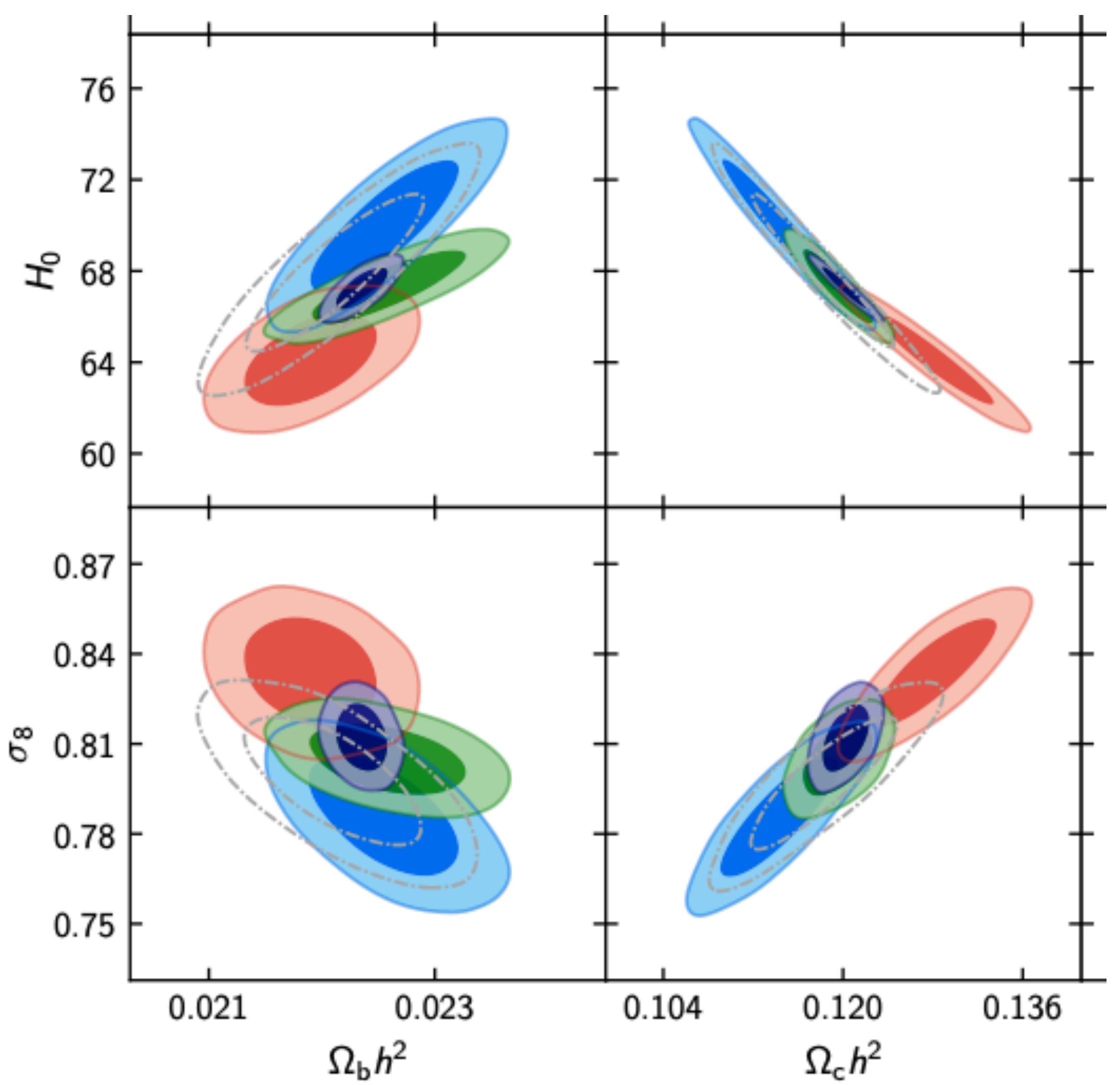
Ω_b : baryon density

It can be calculated from the laws of nature (e.g GR)

$\Rightarrow$ Gaussian Likelihood $\quad \ln \mathcal{L}(\{T_\ell\}|\lambda) \propto \frac{1}{2} \sum_{\ell=1}^{n} \frac{(T_\ell - \bar{T}_\ell(\vec{\lambda}))^2}{2\sigma_\ell^2}$

Sampling the posterior

- Since we have a likelihood, we can now use Bayes theorem and get the posterior.
- Then, one can sample from the posterior, to get the likely values of the desired cosmological parameters.
- The result looks are Monte Carlo plots like this:



H_0 : Hubble parameter

Ω_c : dark matter

Ω_b : baryons

Course logistics

- **Reading for this lecture:**
  - **For example:** Deeplearningbook.org chapter 3 and 5.

- **Problem set**: First problem set next week.