Practical Statistics for Physics and Astrophysics

Prof: Ben Metcalf

Why take a course on statistics?

Statistics is what allows us to connect experimental and observational data to knowledge about the physical world in the presents of uncertainty.

Astronomy / Astrophysics is particularly dependent on statistics

Data is noisy, systematic errors and biases are high. Selection of samples is important.

The quantities of interest are often statistical in nature -
velocity dispersion, luminosity functions, correlation functions, power spectra

How do individual measurements relate to statistical quantities?

Physical quantities are often related to individual measurements in complex ways.

How do we go from individual measurements to the physical qualities of interest?
How well do we know the these physical quantities?

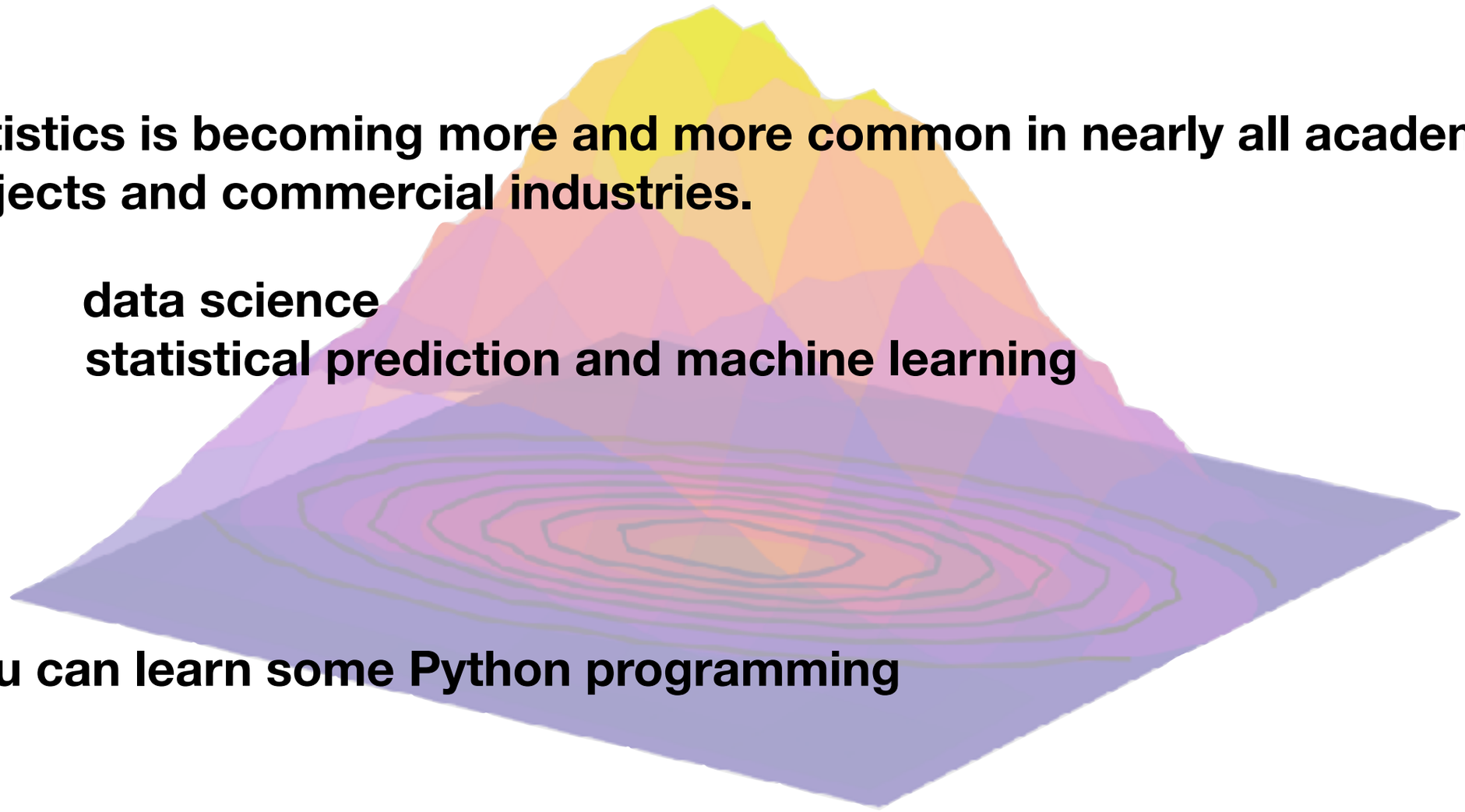
Big Data - some data sets are becoming very large and require careful statistical analysis

Practical Statistics for Physics and Astrophysics

Prof: Ben Metcalf

Why take a course on statistics?

Statistics is becoming more and more common in nearly all academic subjects and commercial industries.



data science
statistical prediction and machine learning

You can learn some Python programming

Practical Statistics for Physics and Astrophysics

Prof: Ben Metcalf

Why take a course on statistics?

examples:

binary stars
or extra-solar planets

$$m_1 + m_2 = \frac{P}{2\pi G}\frac{(v_1^r + v_2^r)^2}{\sin^3 i}$$

- errors in period and radial velocities
- inclination is unknown but could be averaged over
- selection bias related to radial velocities

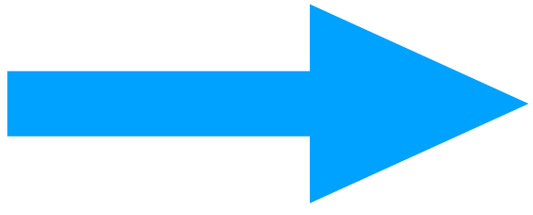
cosmology

data:

galaxies' angular positions

galaxies' redshifts uncertainties

cosmic variance
selection effects



cosmological theory

cosmological parameters

$\Omega_{\rm matter}$ Ω_R $H(z)$ Ω_Λ

Practical Statistics for Physics and Astrophysics

Prof: Ben Metcalf

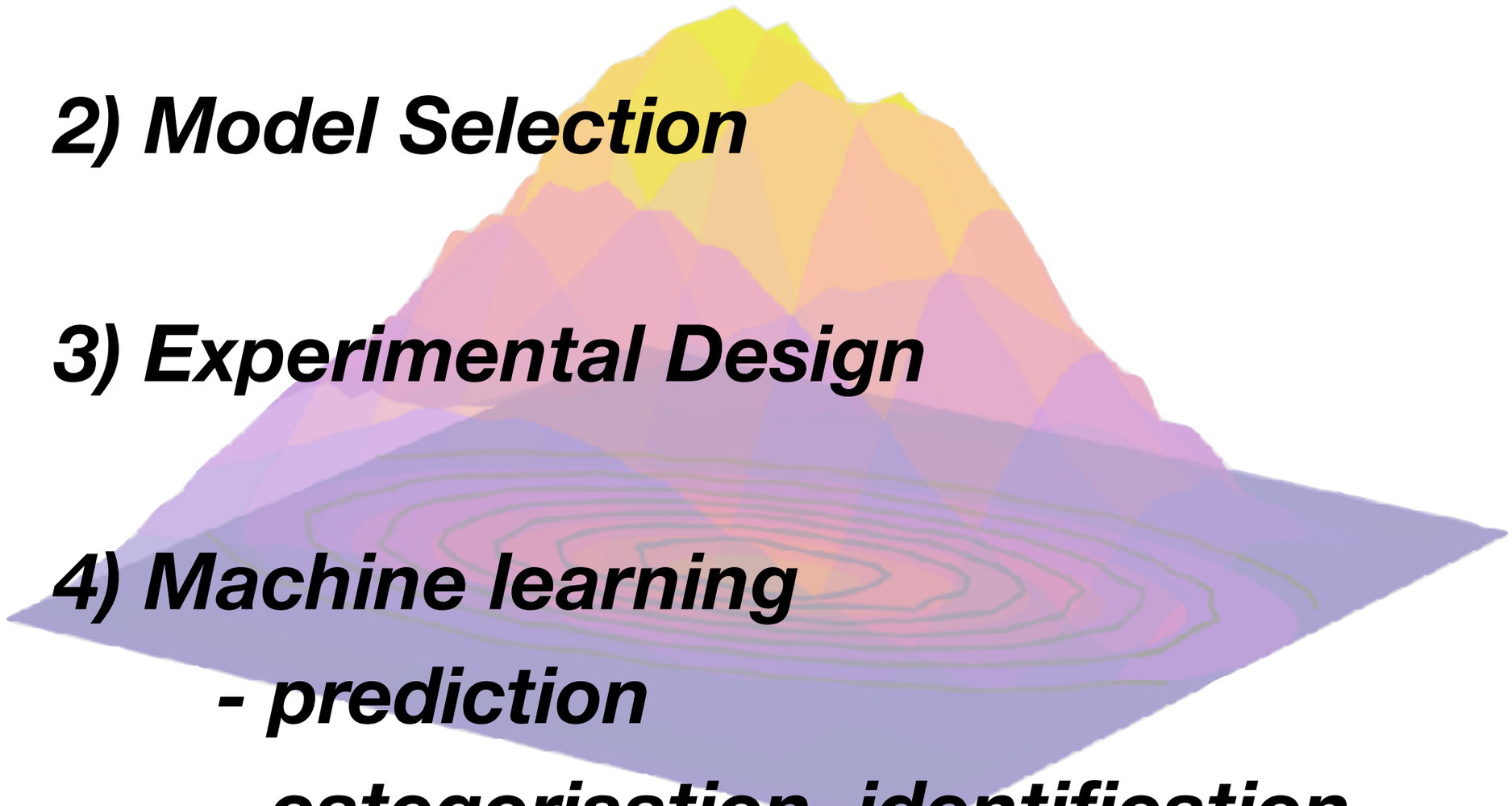
1) Statistical Inference

2) Model Selection

3) Experimental Design

4) Machine learning

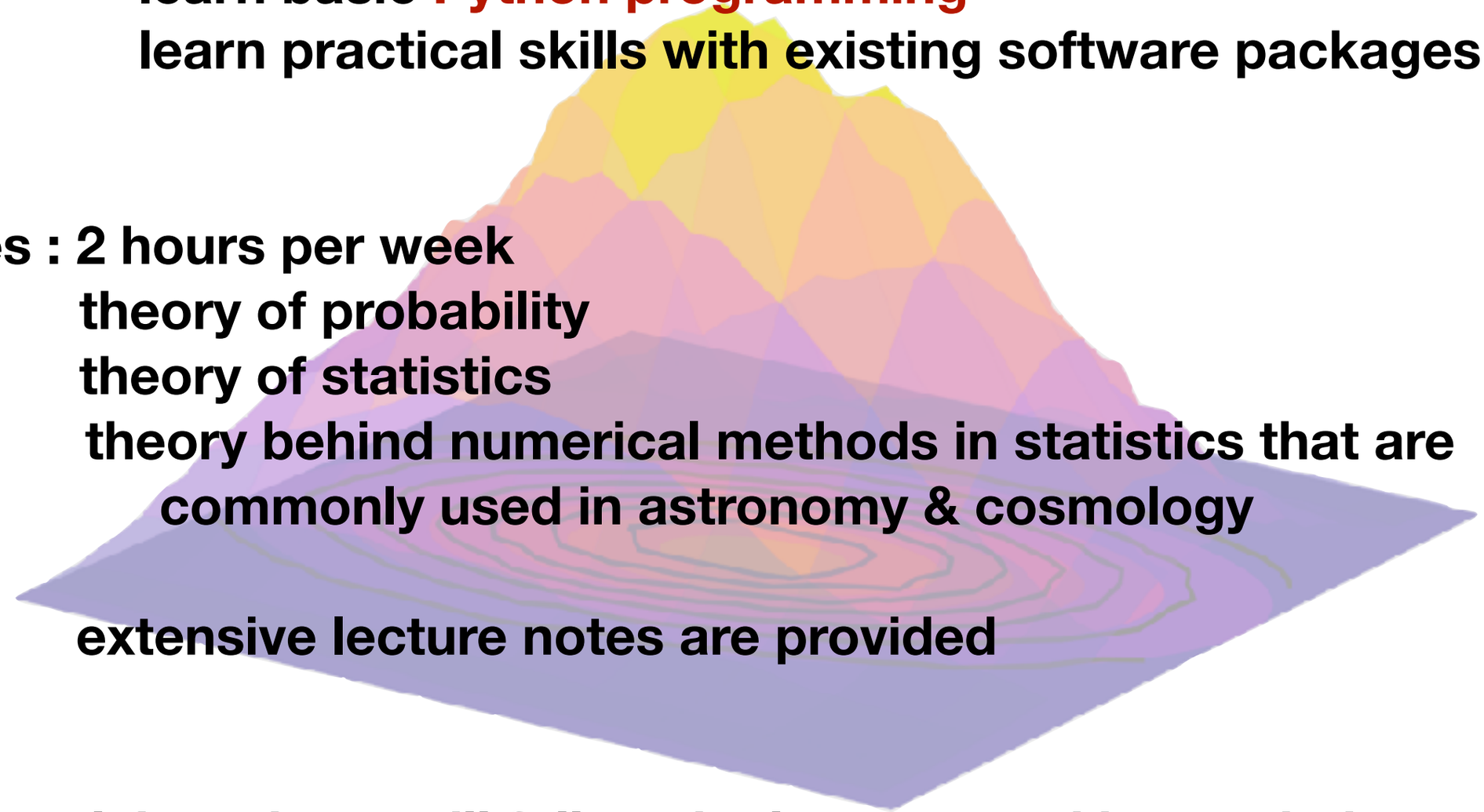
- ***prediction***
- ***categorisation, identification***
- ***data mining***



Practical Statistics for Physics and Astrophysics

Prof: Ben Metcalf

laboratory : 4 hours per week
learn basic Python programming
learn practical skills with existing software packages



lectures : 2 hours per week
theory of probability
theory of statistics
theory behind numerical methods in statistics that are commonly used in astronomy & cosmology

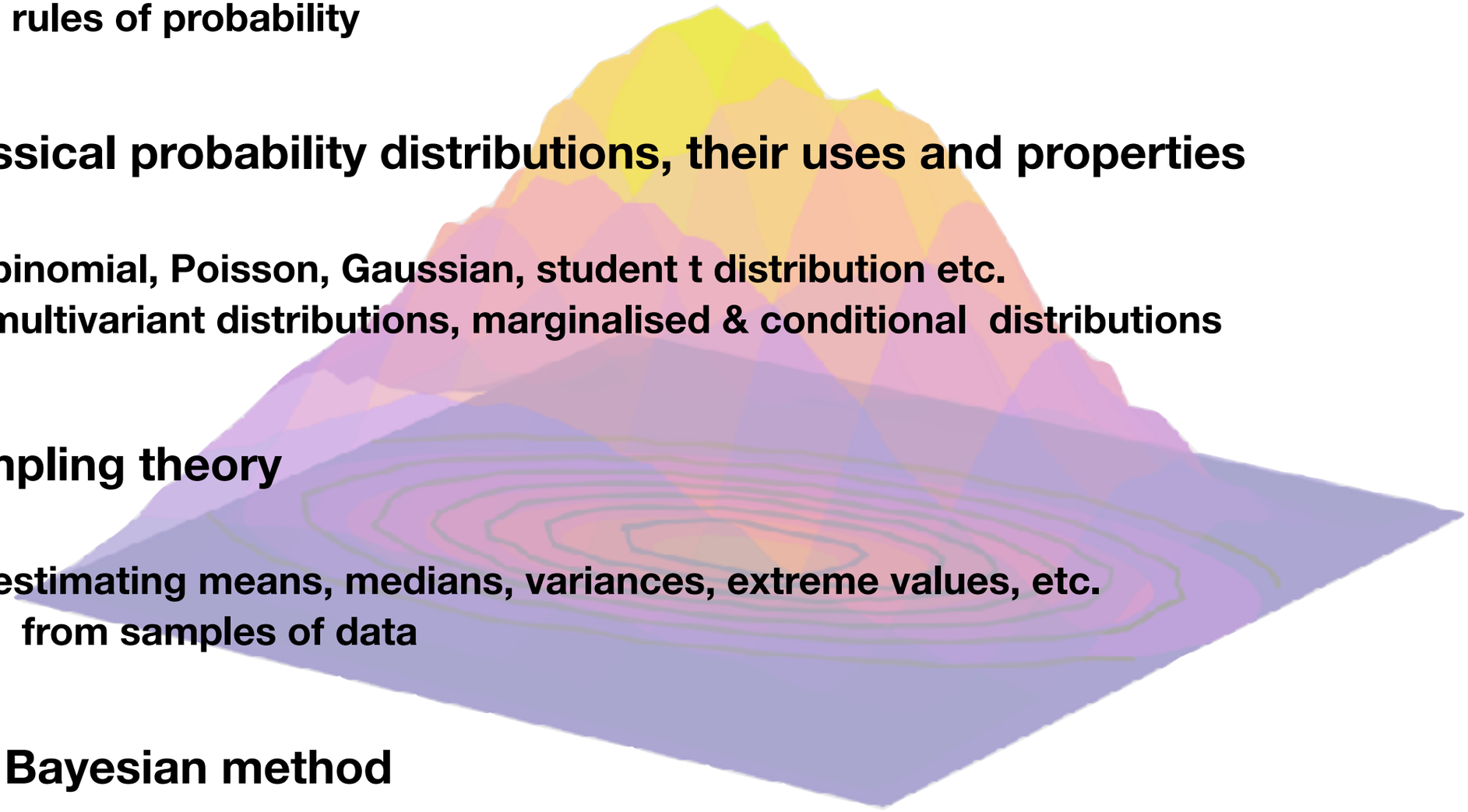
extensive lecture notes are provided

Grades : lab projects will follow the lectures and be graded
oral exam

Practical Statistics for Physics and Astrophysics

What is probability?

theory of probability
Bayesian and frequentist interpretation of probability
rules of probability



Classical probability distributions, their uses and properties

binomial, Poisson, Gaussian, student t distribution etc.
multivariant distributions, marginalised & conditional distributions

Sampling theory

estimating means, medians, variances, extreme values, etc.
from samples of data

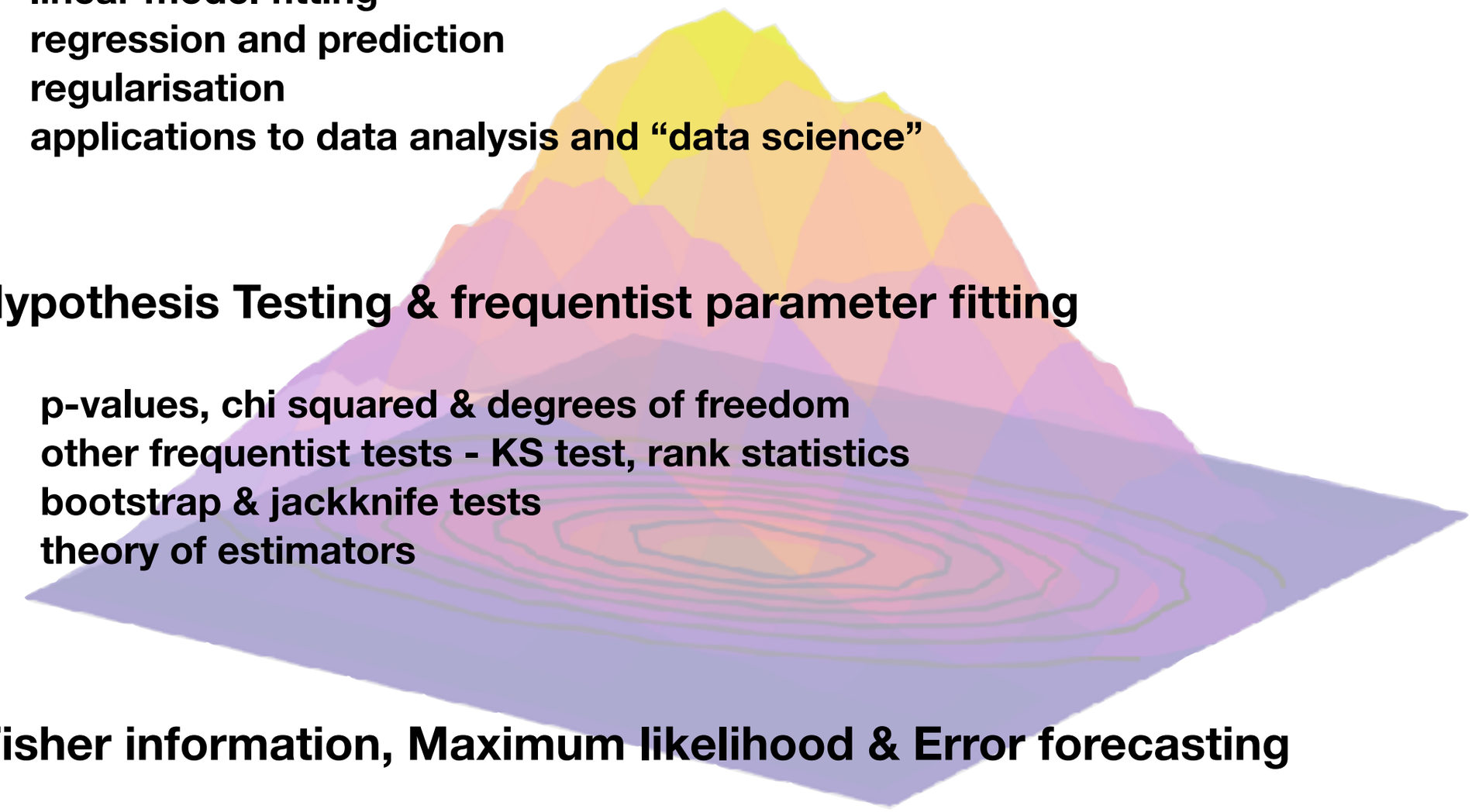
The Bayesian method

Bayesian statistical inference (parameter estimation)
Bayesian model selection
Bayesian model checking

Practical Statistics for Physics and Astrophysics

Linear models, curve fitting, least squares, and regression

fitting curves to data with errors
linear model fitting
regression and prediction
regularisation
applications to data analysis and “data science”



Hypothesis Testing & frequentist parameter fitting

p-values, chi squared & degrees of freedom
other frequentist tests - KS test, rank statistics
bootstrap & jackknife tests
theory of estimators

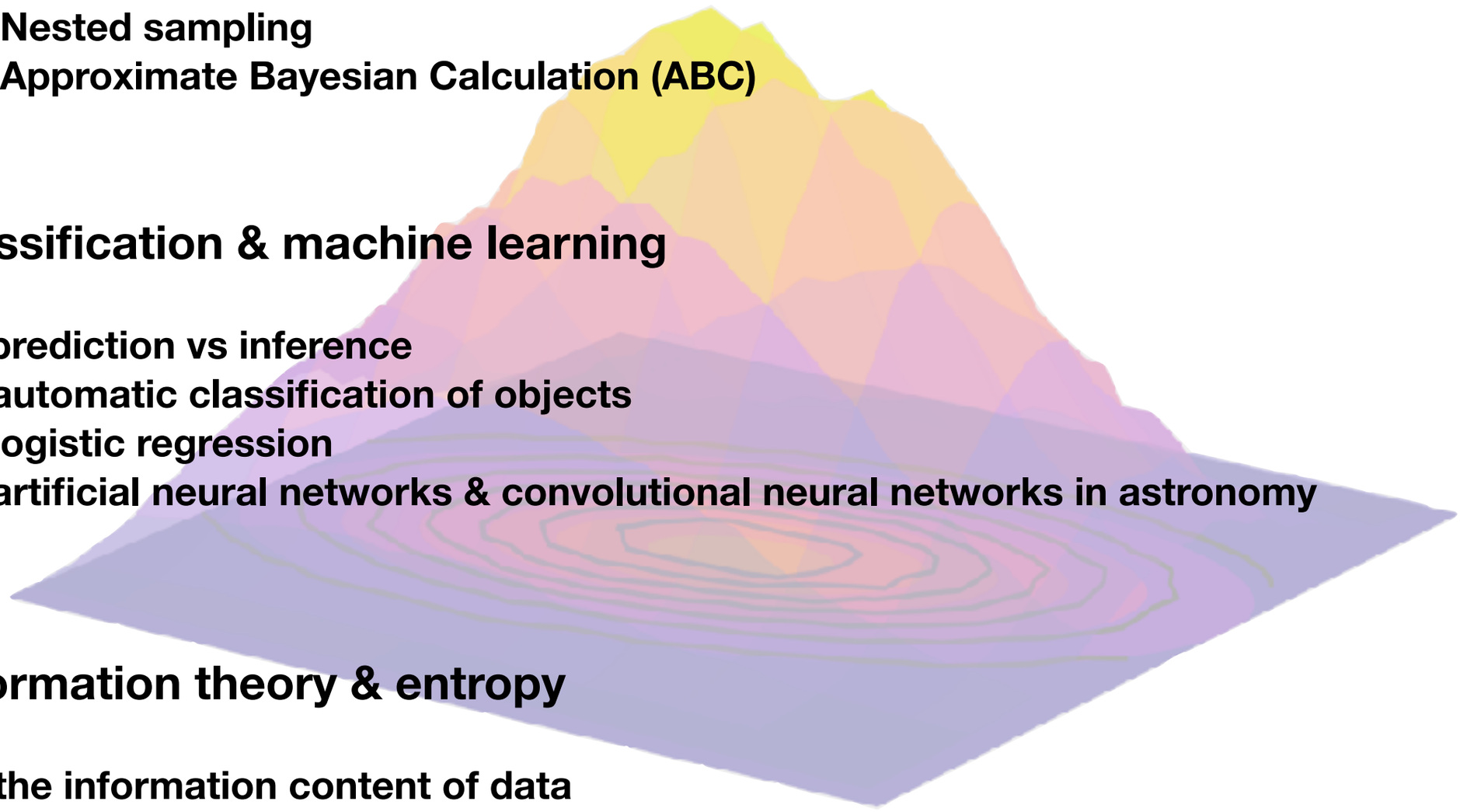
Fisher information, Maximum likelihood & Error forecasting

the Fisher information matrix
forecasting the errors of a proposed experiment / observation
the maximum likelihood estimator

Practical Statistics for Physics and Astrophysics

Numerical methods for the Bayesian inference problem

Forward modelling
Markov Chain Monte Carlo (MCMC)
Nested sampling
Approximate Bayesian Calculation (ABC)



Classification & machine learning

prediction vs inference
automatic classification of objects
logistic regression
artificial neural networks & convolutional neural networks in astronomy

Information theory & entropy

the information content of data
the connection between information theory, statistical mechanics & thermodynamics