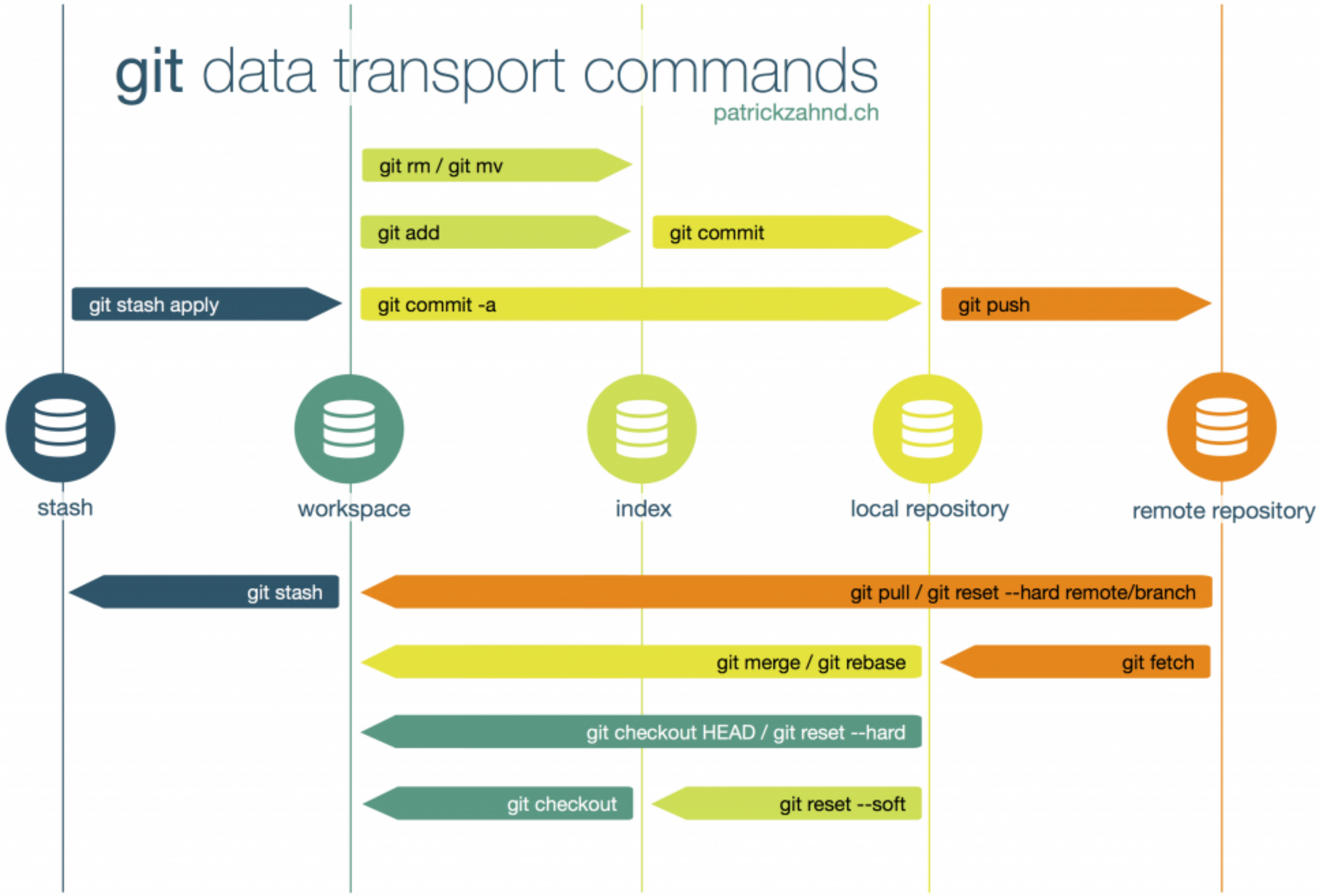
GAUTHAM NARAYAN

FUNDAMENTALS OF DATA SCIENCE
WEEK 1

RECAP

- You installed the `conda` package manager, python and a bunch of packages
- You forked the repo from github to your machine
- You added a file to your git workspace
- You created some SSH keys and added the public key to github
- You pushed your change to your fork of the repo
- You opened a pull request and I merged your change

14/20 of you made it to the last step!
What questions do you have?



More help with git/os commands/python in the help/ directory

SYNCING YOUR FORK

- https://help.github.com/en/github/collaborating-with-issues-and-pull-requests/syncing-a-fork
- git remote -v
  - Check if my repo is already there! If not:
  - git remote add upstream https://github.com/gnarayan/ast596_2020_Spring.git
- git fetch upsteam
- get checkout master
- git merge upstream/master
- git push origin master

We're not focusing on git/python/os usage – pick it up as we go.

What the class is **NOT**

A Statistics Class

A Math Methods Class

A Computer Science Class

A Programming Class

1.0

DATA

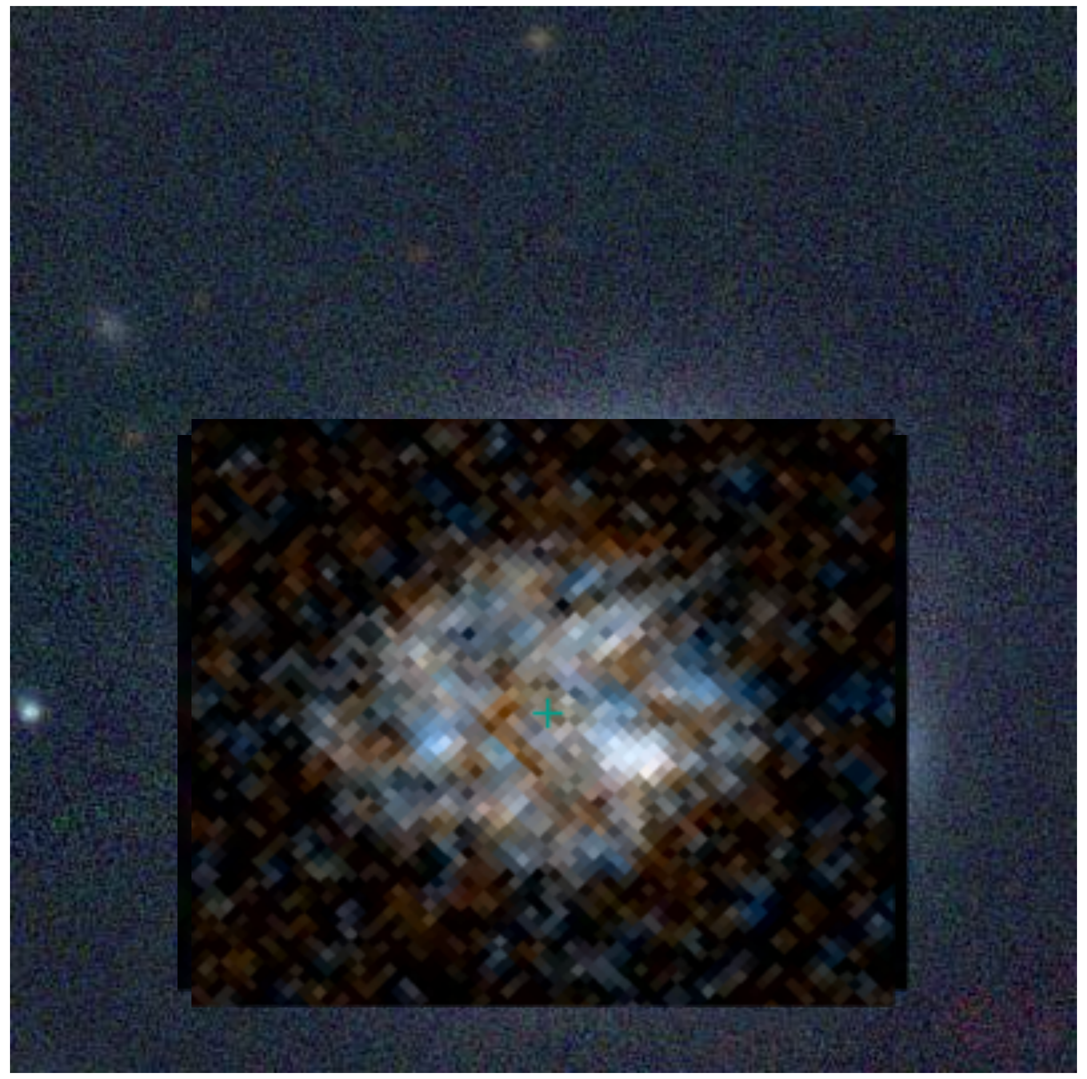
WHAT ASTRONOMERS CAN MEASURE

- Angular separations
- Time differences
- Energy differences

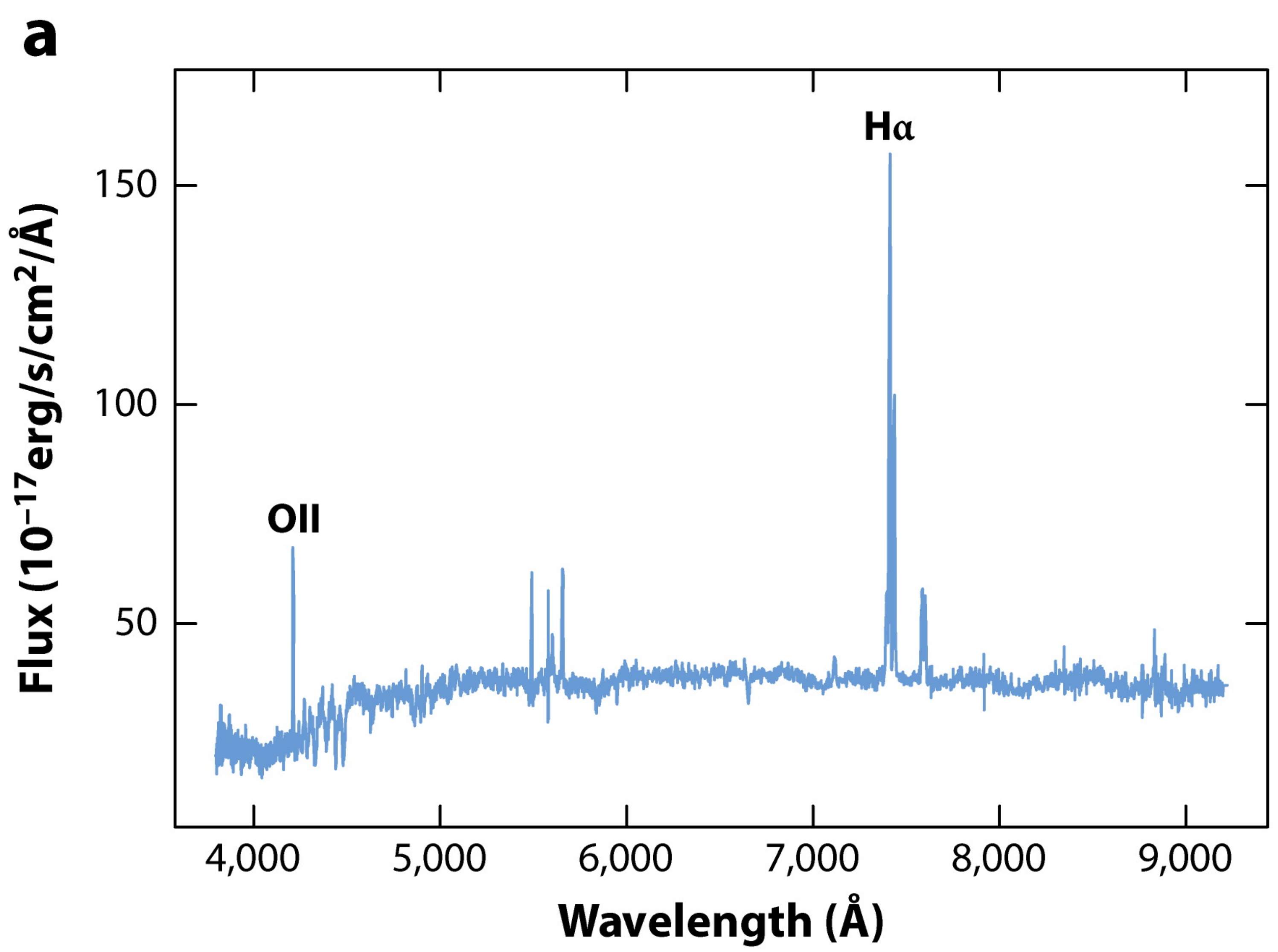
That's it.

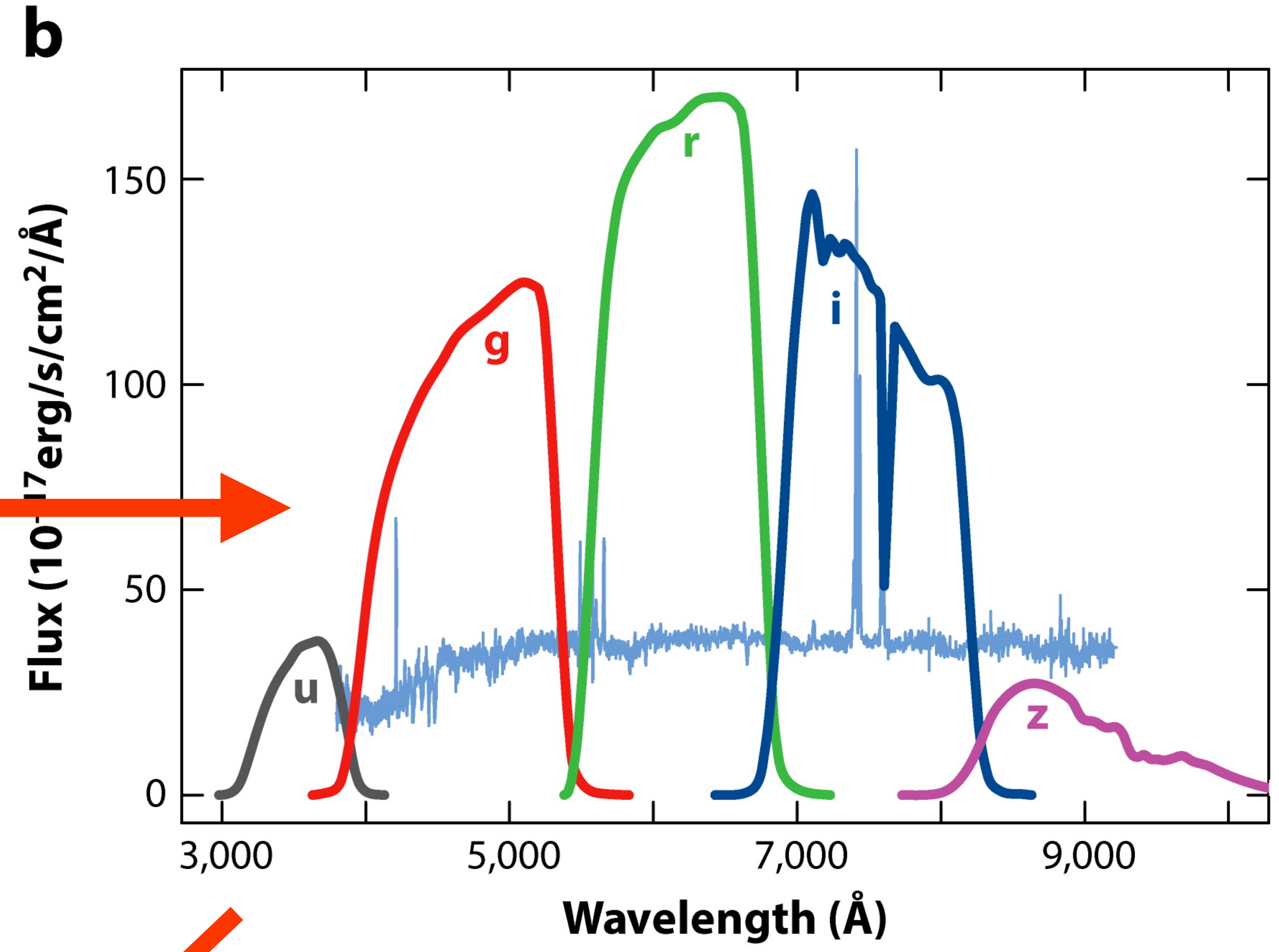
WHAT ASTRONOMERS CAN MEASURE

- **Astrometry** (angular position on the sky) - arcseconds
  - Related definition: 1 parsec (pc) = distance at which a distance of 1 AU (i.e. Earth-Sun) subtends and angle of 1 arcsecond i.e. 1 pc = 1 AU/tan(1") ~ 31 trillion kilometers or 3.26 light years (ly)
- **Photometry** (how bright something is)
  - Flux = photons (or energy in ergs)/sec/cm^2
  - (Apparent) Magnitude = -2.5 log10(Flux) + const
  - (Absolute) Magnitude = -2.5log10(Luminosity) + const = magnitude you'd measure if you could move the source to 10 pc
- **Light curves** = photometry vs time
  - Evolution in source brightness either because of intrinsic (supernovae, AGN) or extrinsic (asteroids, eclipsing binaries)
- **Spectroscopy** = Energy vs wavelength/frequency
- **Images/maps** = Energy vs position on the sky (clustering, spatial correlation functions)
- **Proper Motion** = Astrometry vs time (e.g. stars, satellite galaxies, asteroids…)



Optical/IR/UV





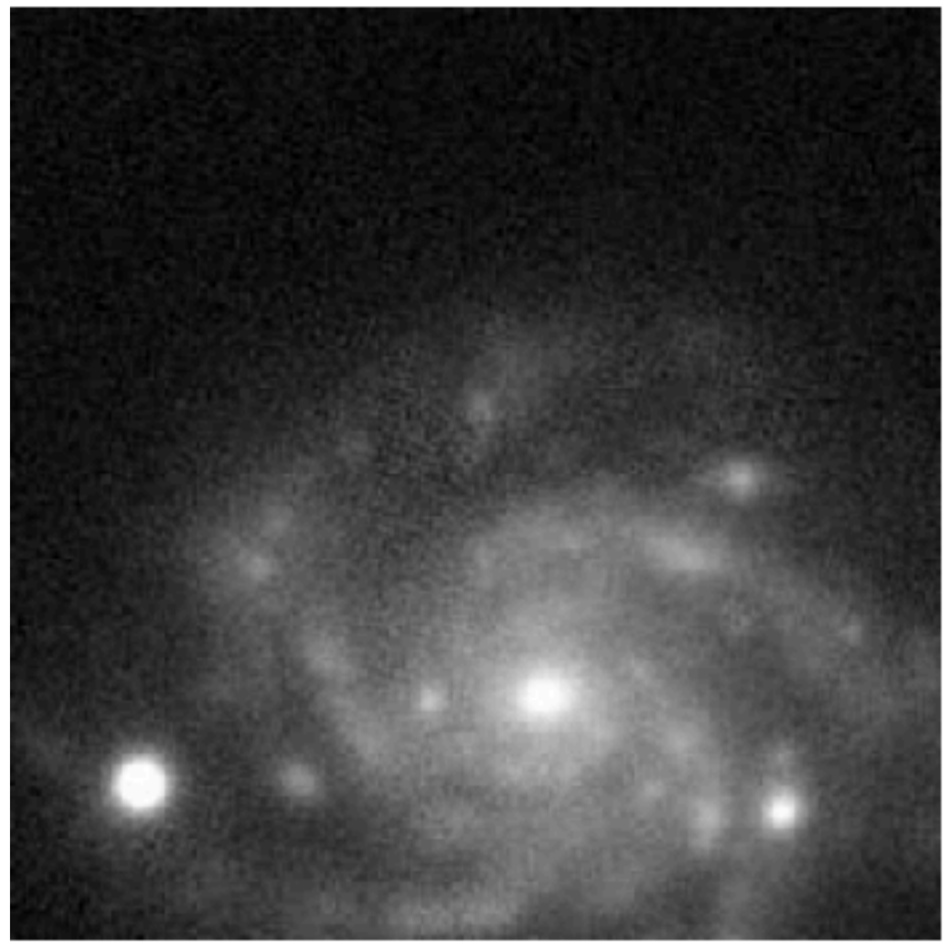
Schafer CM. 2015.
Annu. Rev. Stat. Appl. 2:141–62

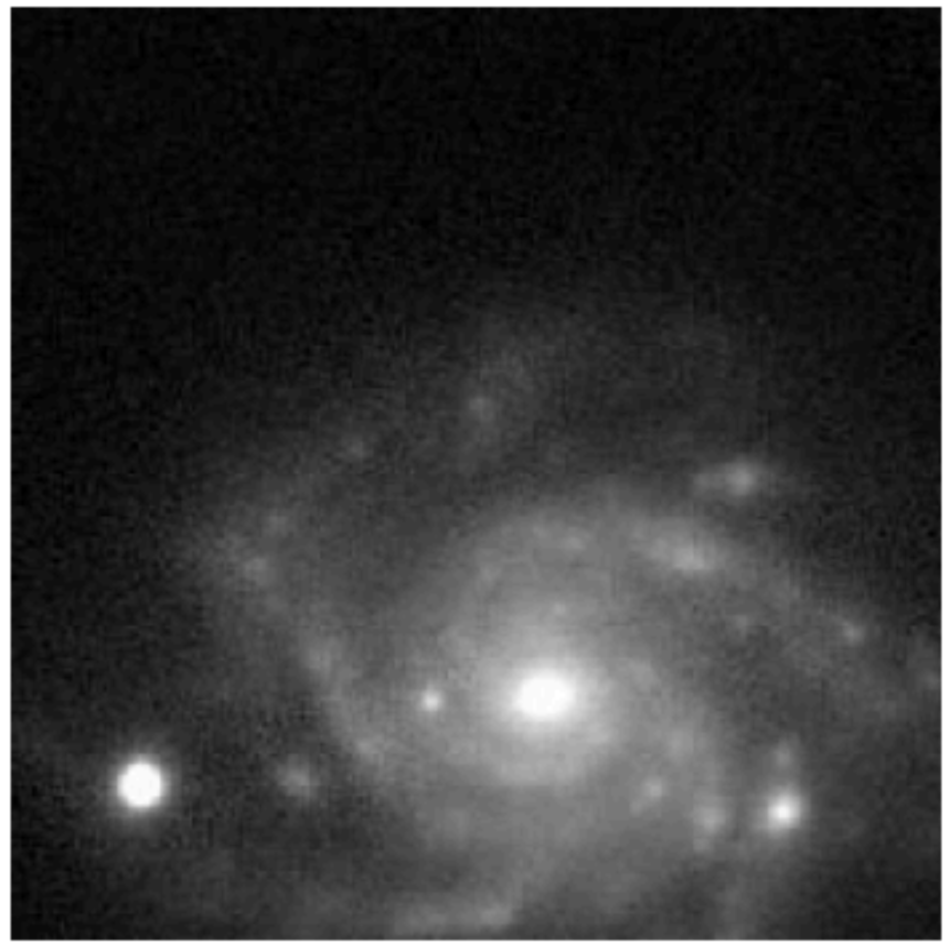
Galaxy Photometry

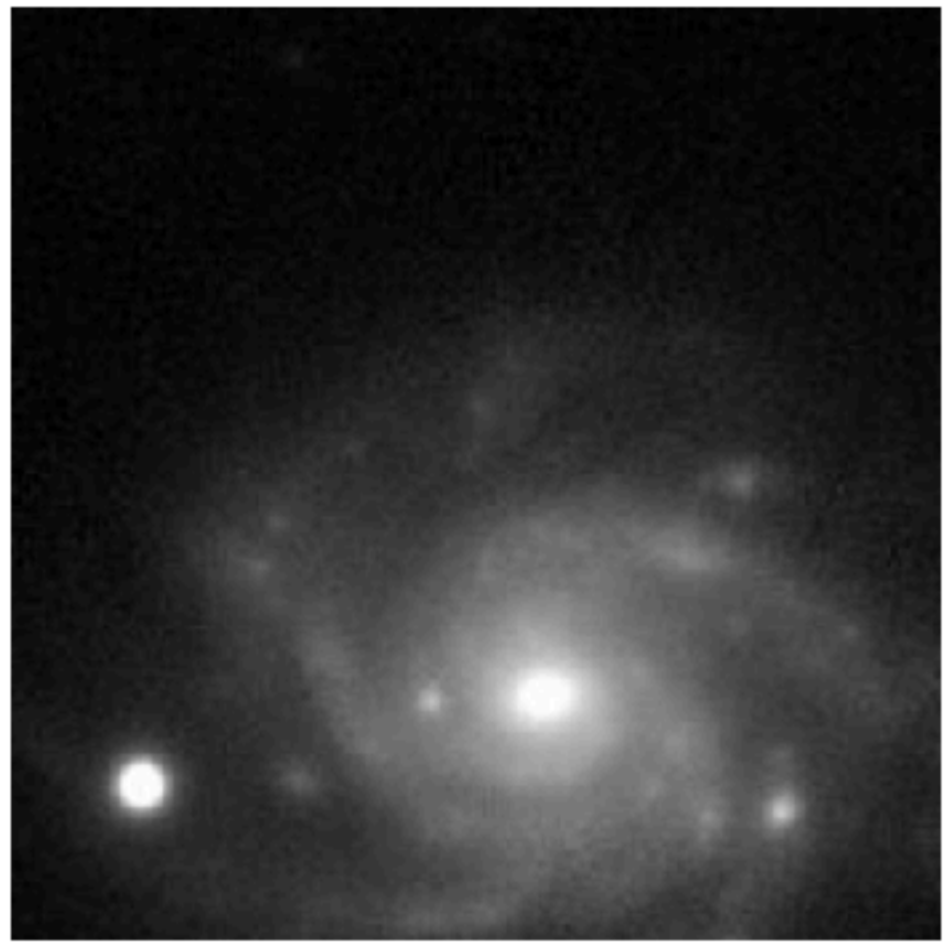
(Brightness: Flux / Magnitude)

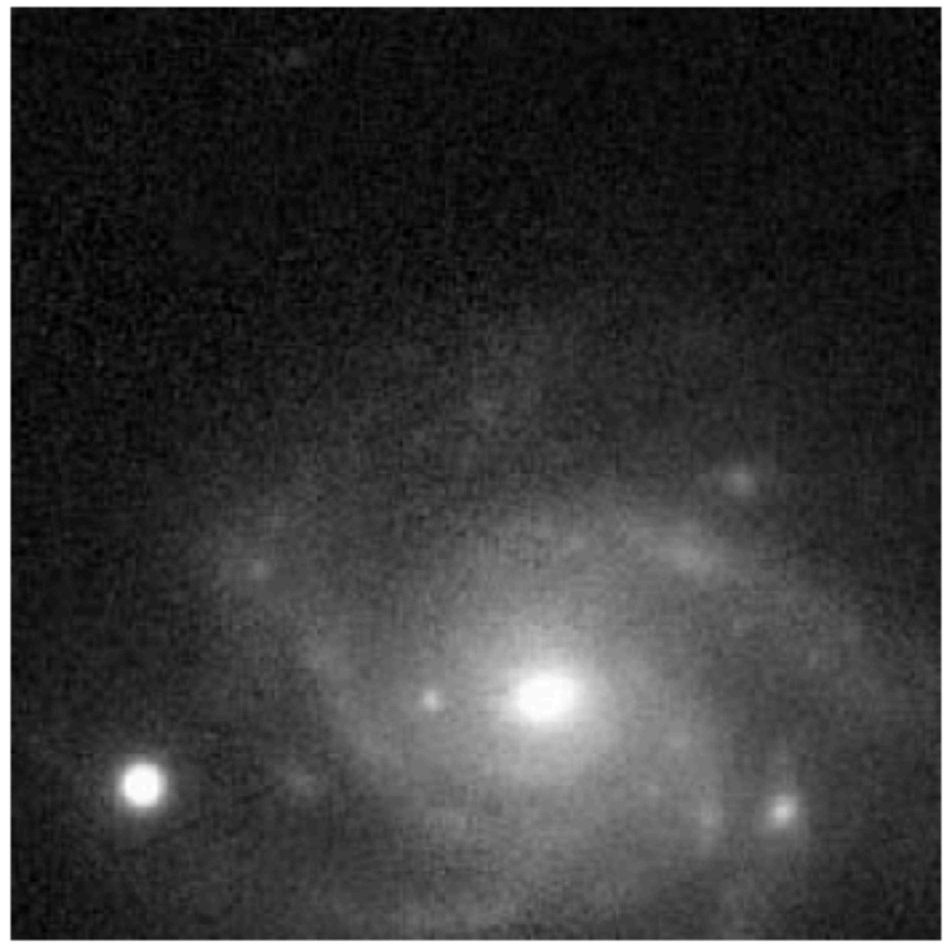
Galaxy Spectrum

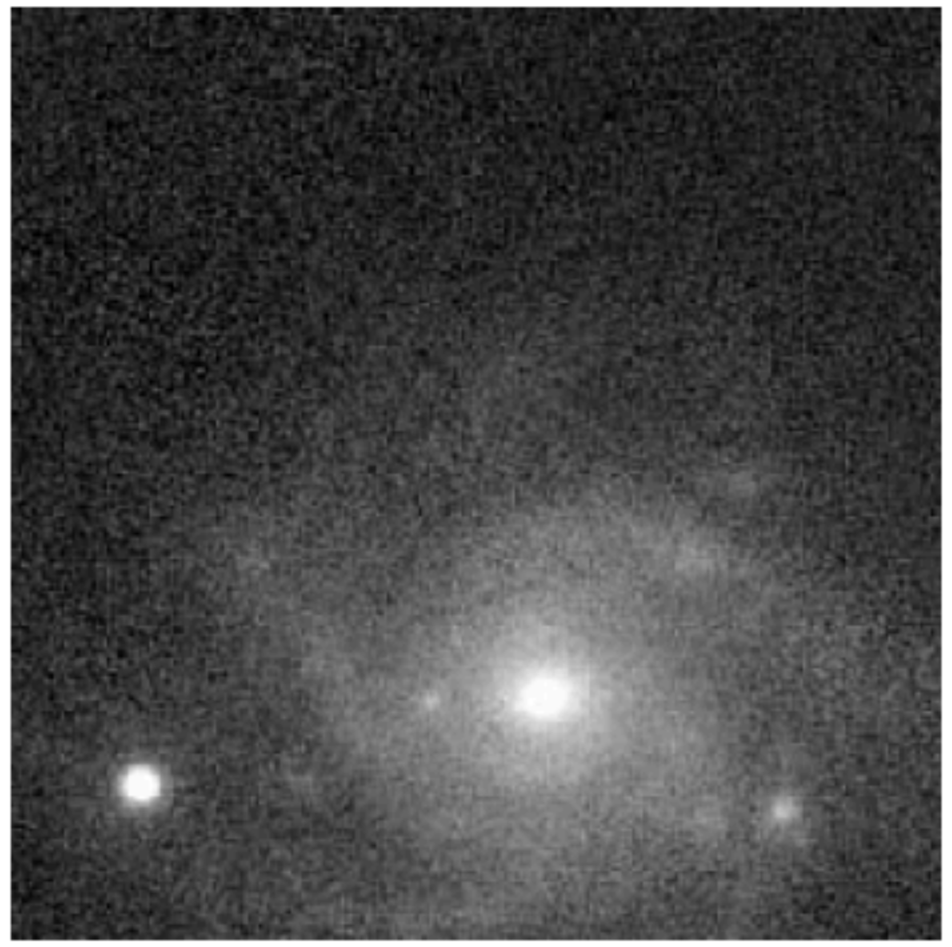
stack 920.038 g	stack 920.038 r	stack 920.038 i	stack 920.038 z	stack 920.038 y
Display FITS FITS-cutout	Display FITS FITS-cutout	Display FITS FITS-cutout	Display FITS FITS-cutout	Display FITS FITS-cutout



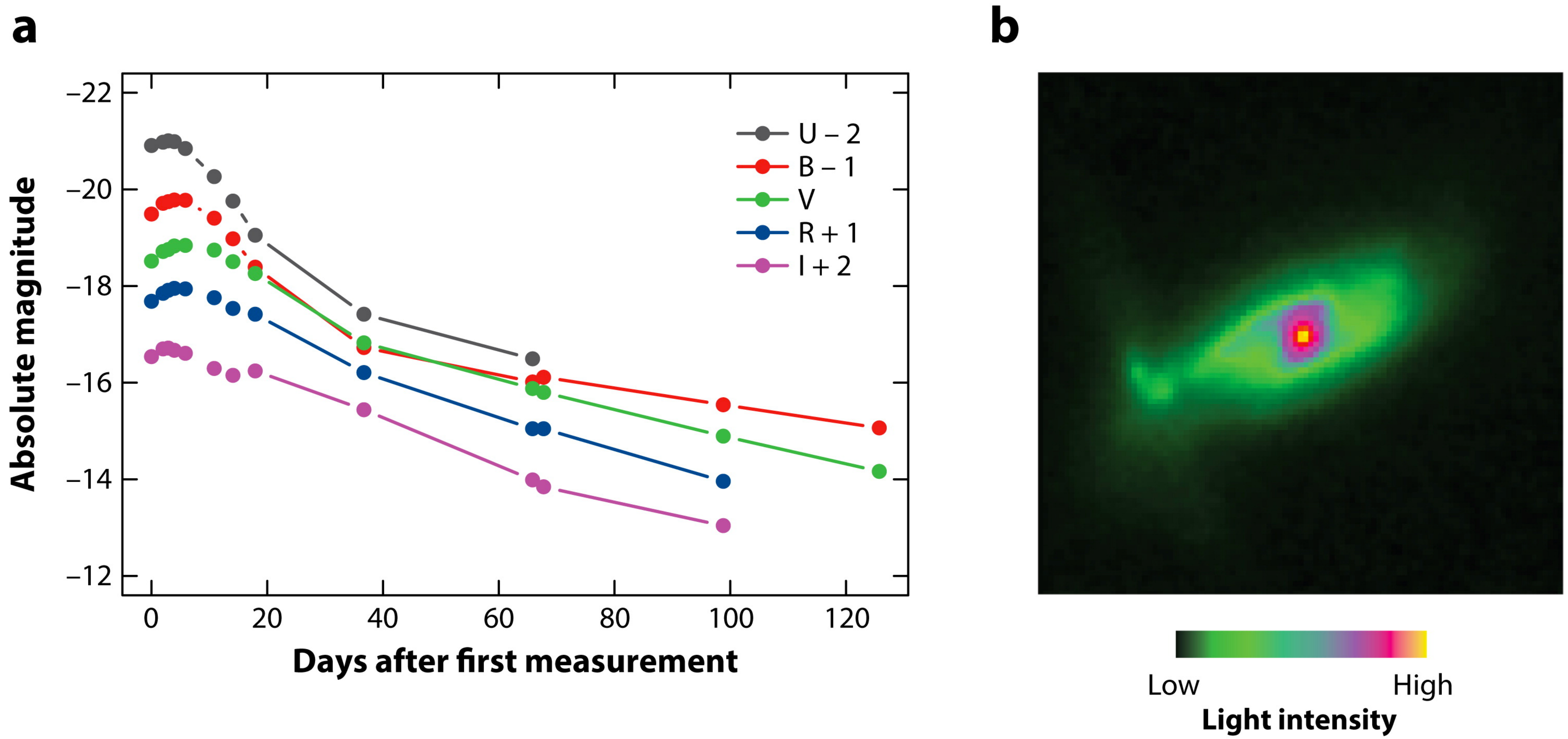








Temporal & Spatial Variation



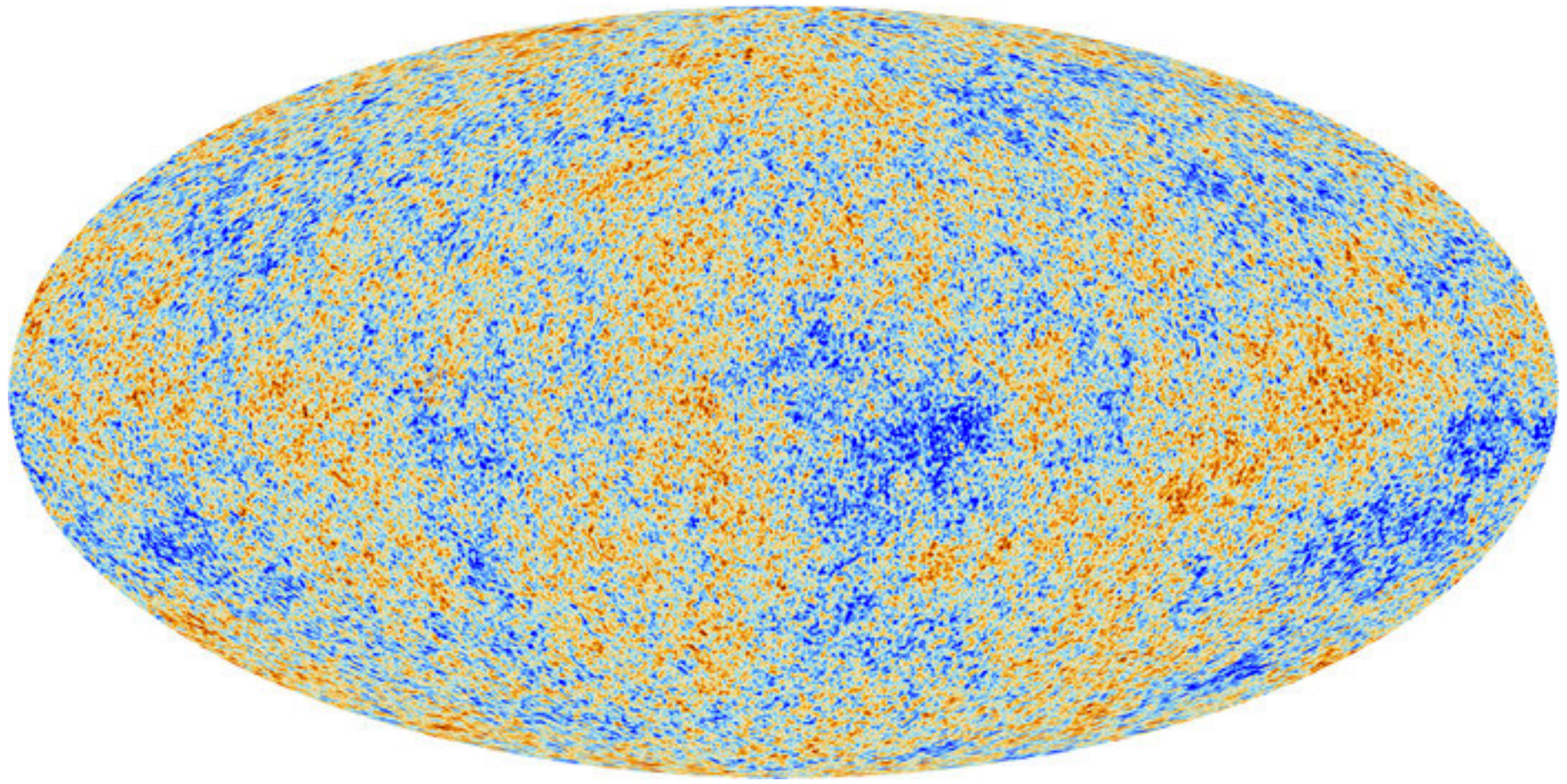
Schafer CM. 2015.
Annu. Rev. Stat. Appl. 2:141–62

Time Series (Light Curve)
Supernova

Galaxy Image
(Intensity Map)

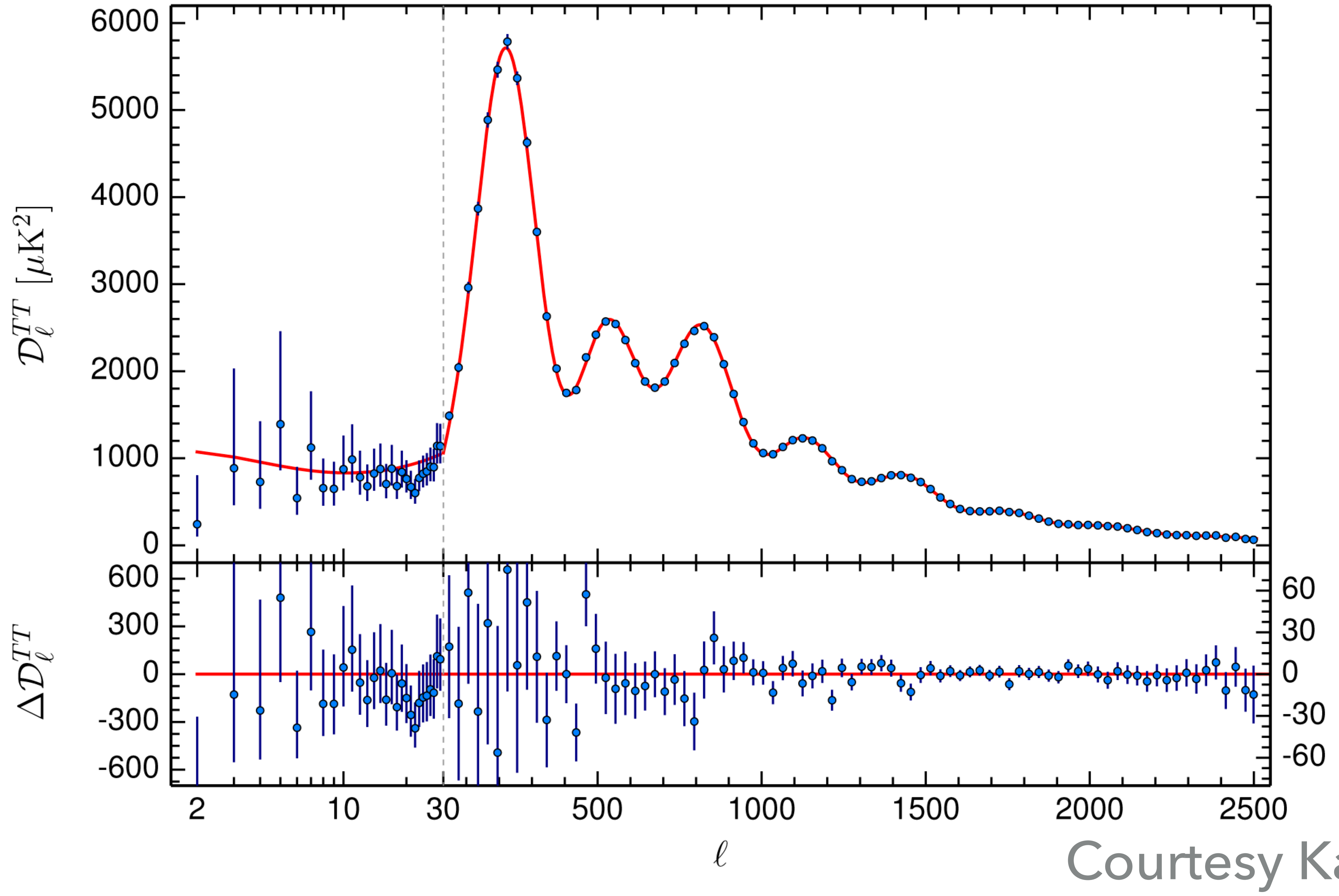
Courtesy Kaisey Mandel

Spatial Variation



Cosmic Microwave Background (Planck) ~ Gaussian Random Field (mean = 2.7 K, std dev ~ 10^{-5})

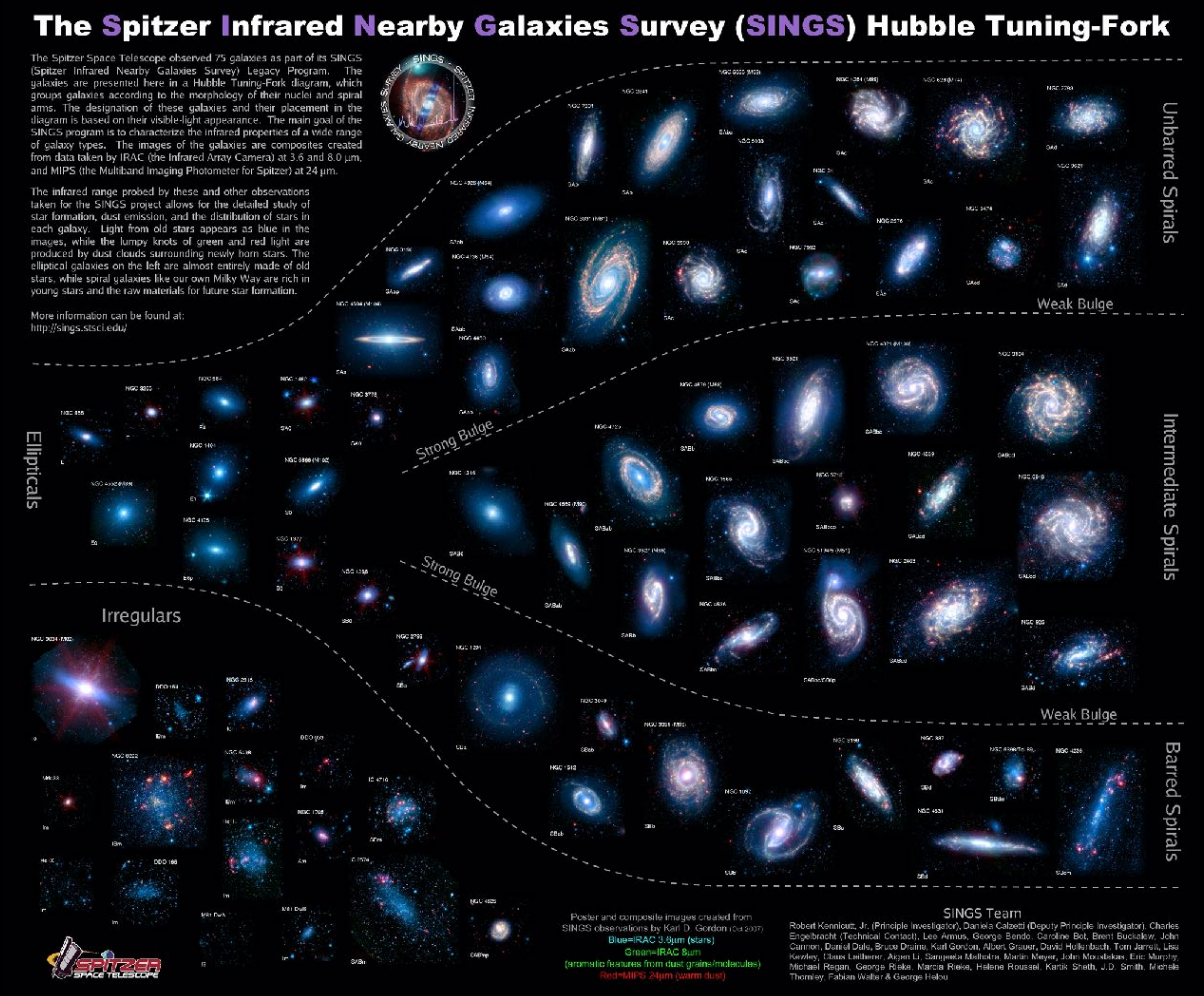
Power Spectrum (~Fourier Transform of Correlation Function) sensitive to cosmological parameters



Courtesy Kaisey Mandel

IN CLASS EXERCISE

- Dealing with spectra (1-D data):
  - Programmatic: Use pandas to load the ascii spectrum and plot it
  - Canned: Use specviz to look at the spectra and zoom in on the region around 6000-7000 A
- Dealing with images (2-D data):
  - Programmatic: Use astropy.io.fits to load a .fits image
    Display it with matplotlib
  - Canned: Use ds9 to look at the images and adjust the scale


The Spitzer Infrared Nearby Galaxies Survey (SINGS) Hubble Tuning-Fork
The Spitzer Space Telescope observed 75 galaxies as part of its SINGS (Spitzer Infrared Nearby Galaxies Survey) Legacy Program. The galaxies are presented here in a Hubble Tuning-Fork diagram, which groups galaxies according to the morphology of their nuclei and spiral arms. The designation of these galaxies and their placement in the diagram is based on their visible-light appearance. The main goal of the SINGS program is to characterize the infrared properties of a wide range of galaxy types. The images of the galaxies are composites created from data taken by IRAC (the Infrared Array Camera) at 3.6 and 8.0 µm, and MIPS (the Multiband Imaging Photometer for Spitzer) at 24 µm.
The infrared range probed by these and other observations taken for the SINGS project allows for the detailed study of star formation, dust emission, and the distribution of stars in each galaxy. Light from old stars appears as blue in the images, while the lumpy knots of green and red light are produced by dust clouds surrounding newly born stars. The elliptical galaxies on the left are almost entirely made of old stars, while spiral galaxies like our own Milky Way are rich in young stars and the raw materials for future star formation.
More information can be found at:
http://sings.stsci.edu/
Unbarred Spirals
Weak Bulge
Strong Bulge
Ellipticals
Intermediate Spirals
Strong Bulge
Irregulars
Weak Bulge
Barred Spirals
Poster and composite images created from SINGS observations by Karl D. Gordon (Oct 2007)
Blue=IRAC 3.6µm (stars)
Green=IRAC 8µm
(aromatic features from dust grains/molecules)
Red=MIPS 24µm (warm dust)
SINGS Team
Robert Kennicutt, Jr. (Principle Investigator), Daniela Calzetti (Deputy Principle Investigator), Charles Engelbracht (Technical Contact), Lee Armus, George Bendo, Caroline Bot, Brent Buckalew, John Cannon, Daniel Dale, Bruce Draine, Karl Gordon, Albert Grauer, David Hollenbach, Tom Jarrett, Lisa Kewley, Claus Leitherer, Aigen Li, Sangeeta Malhotra, Martin Meyer, John Moustakas, Eric Murphy, Michael Regan, George Rieke, Marcia Rieke, Helene Roussel, Kartik Sheth, J.D. Smith, Michele Thornley, Fabian Walter & George Helou
SPITZER SPACE TELESCOPE

Statistical inference is a logical framework with which to test our beliefs of a noisy world against data.

We formalize our beliefs in a probabilistic model.

1.1

AXIOMS OF PROBABILITY

x is a scalar quantity, measured N times

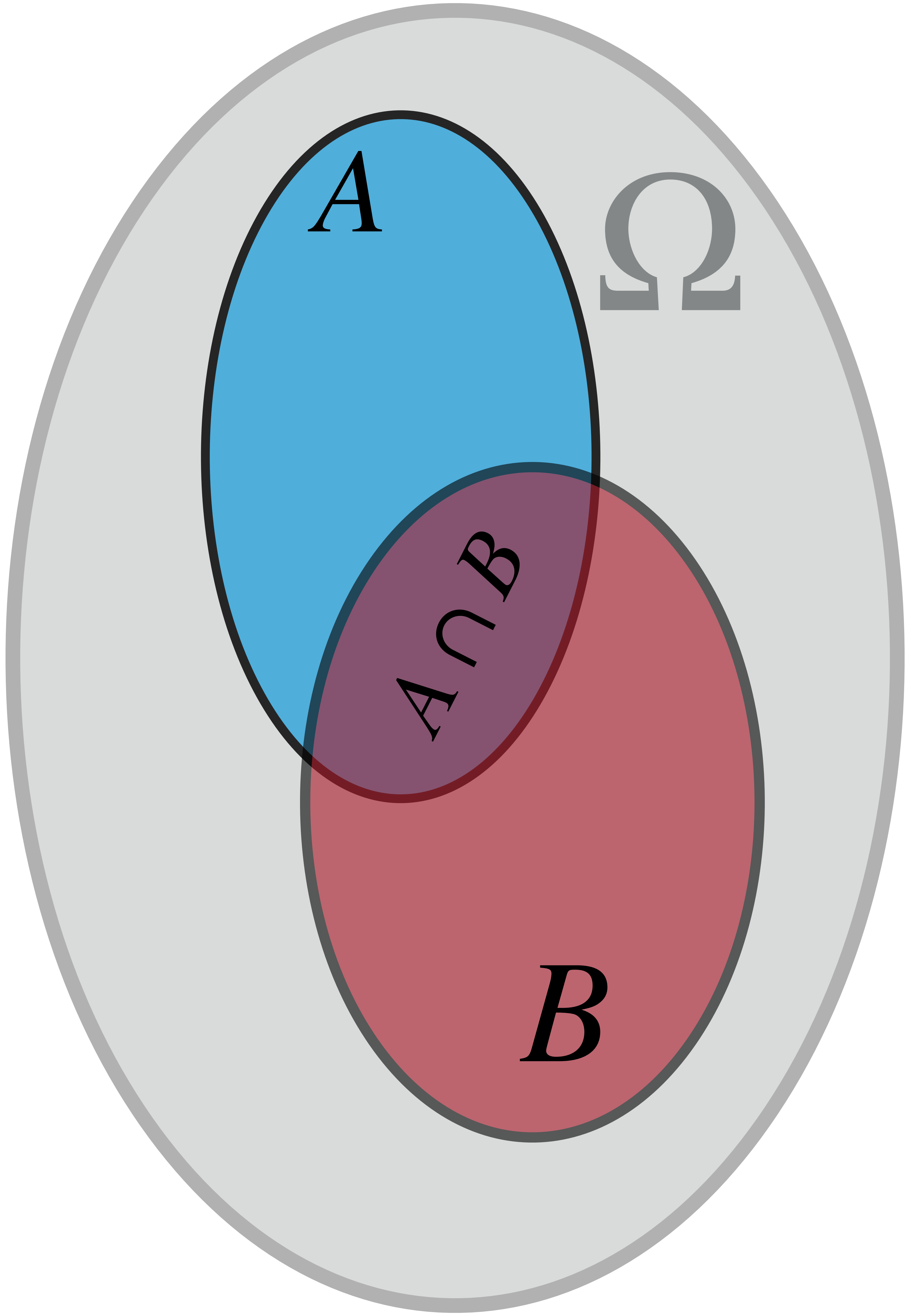
x_i is a single measurement with $i = 1, ..., N$

x_i refers to the set of all N measurements

Let Ω be a collection of possible elementary events, and A and B events such that $A, B \in \Omega$

- $P(A) \geq 0$ for all A;
- $P(\Omega) = 1$;
- $$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

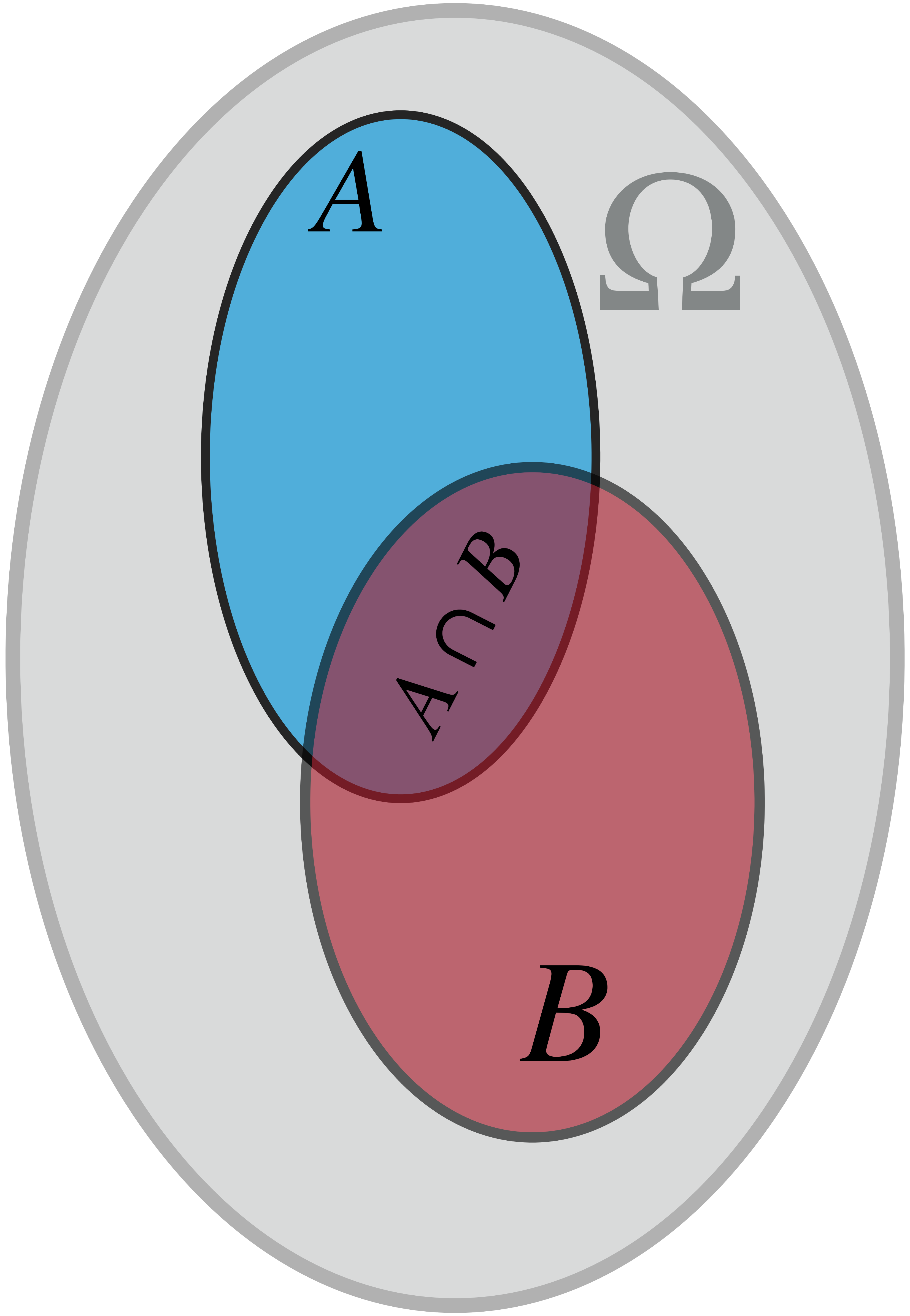
for all countable disjoint sets $A_1, A_2 .. \in \Omega$



Let Ω be a collection of possible elementary events, and A and B events such that $A, B \in \Omega$

- $P(A) \geq 0$ for all A;
- $P(\Omega) = 1$;
- $$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

  for all countable disjoint sets $A_1, A_2 .. \in \Omega$
- Some consequences: $P(A) + P(A^C) = 1$



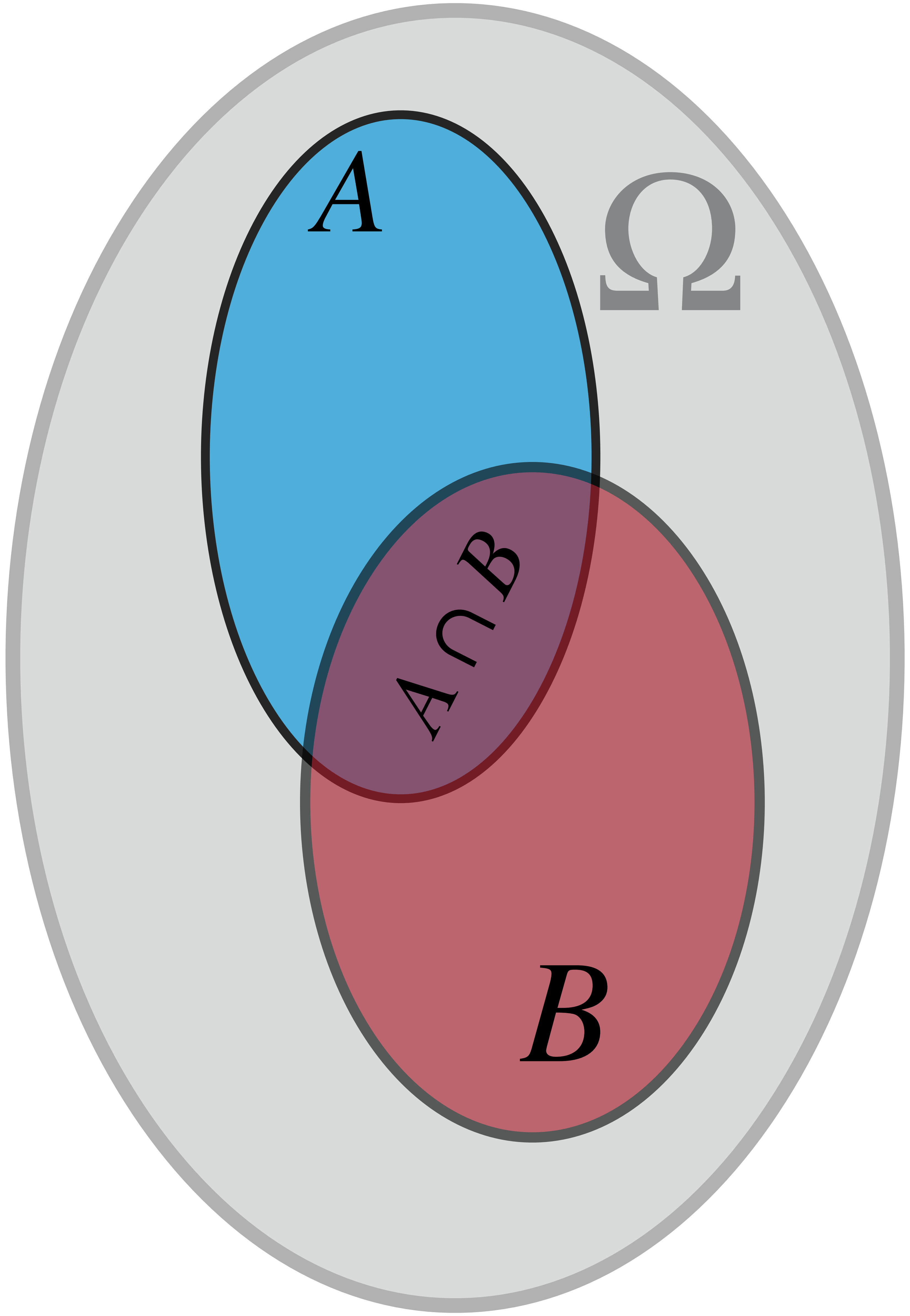
Largely because a 100 years is not enough time for people to agree on very much, there's a few different notations used in the readings.

$$P(A \cap B) = P(A, B) = P(A \text{ and } B)$$

Let Ω be a collection of possible elementary events, and A and B events such that $A, B \in \Omega$

- $P(A) \geq 0$ for all A;
- $P(\Omega) = 1$;
- $$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

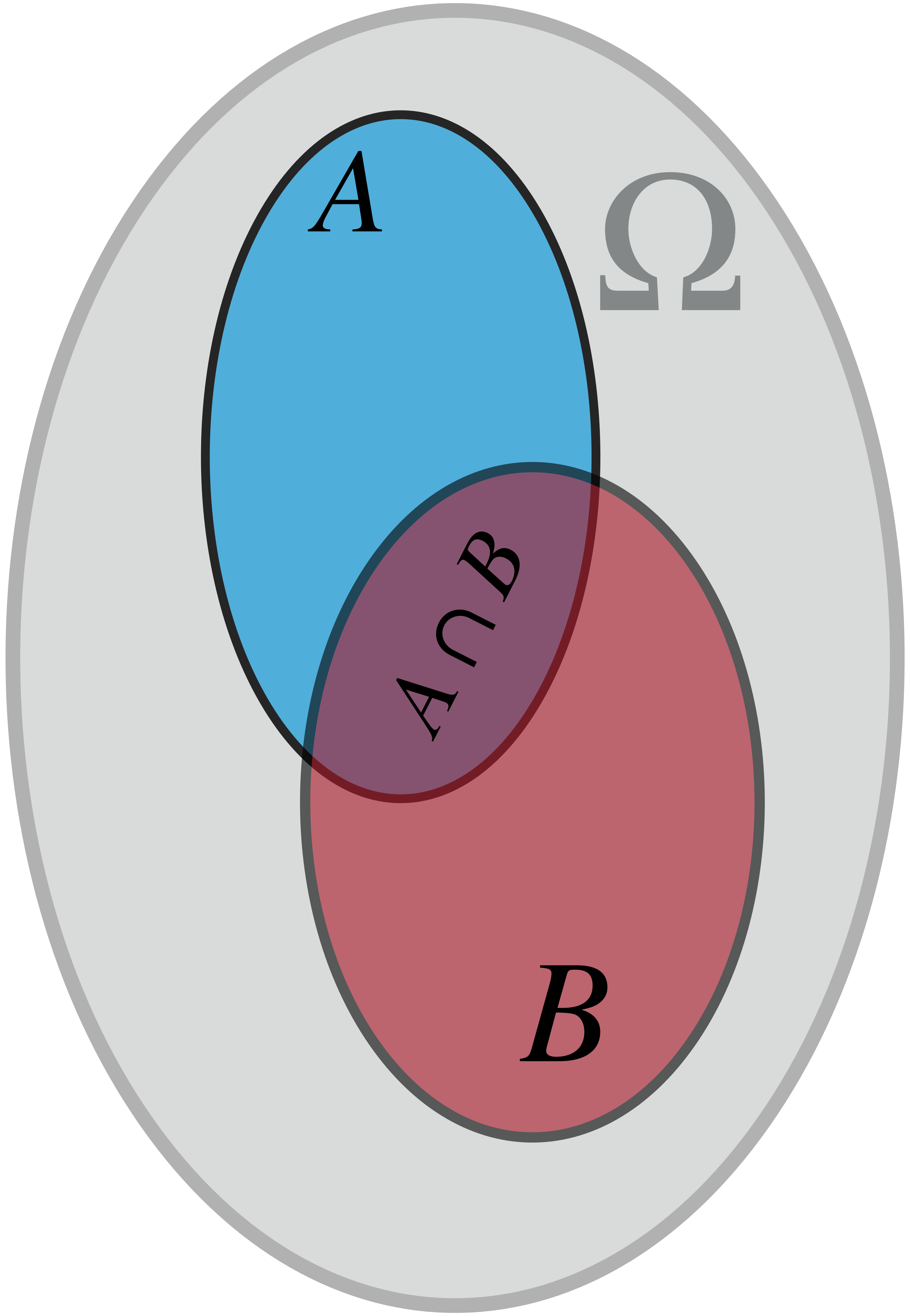
  for all countable disjoint sets $A_1, A_2 .. \in \Omega$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



Let Ω be a collection of possible elementary events, and A and B events such that $A, B \in \Omega$

- $P(A) \geq 0$ for all A;
- $P(\Omega) = 1$;
- $$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

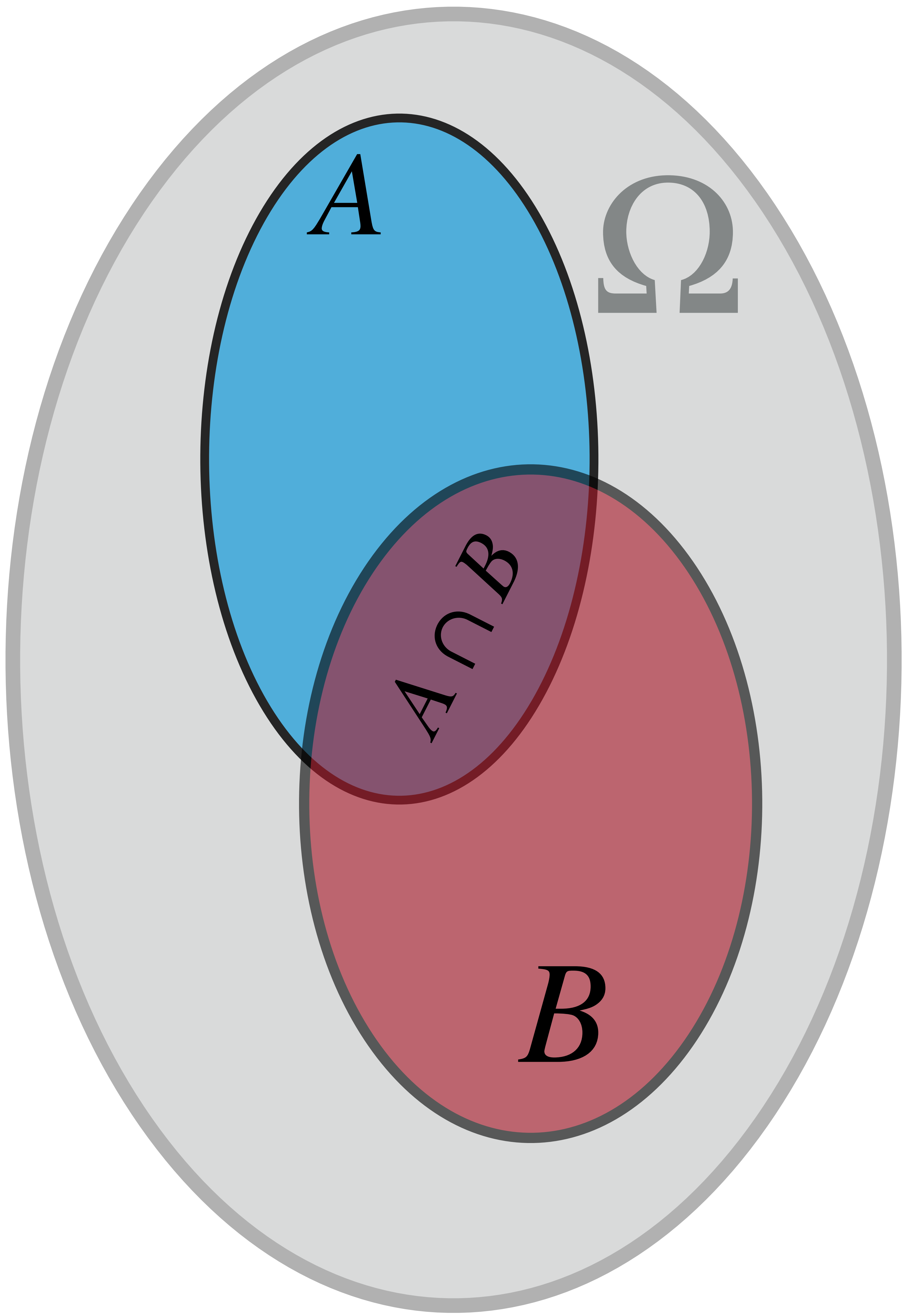
  for all countable disjoint sets $A_1, A_2 .. \in \Omega$
- $$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$ Conditional Probability



Let Ω be a collection of possible elementary events, and A and B events such that $A, B \in \Omega$

- $P(A) \geq 0$ for all A;
- $P(\Omega) = 1$;
- $$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

  for all countable disjoint sets $A_1, A_2 .. \in \Omega$
- $P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A) = P(A \cap B)$

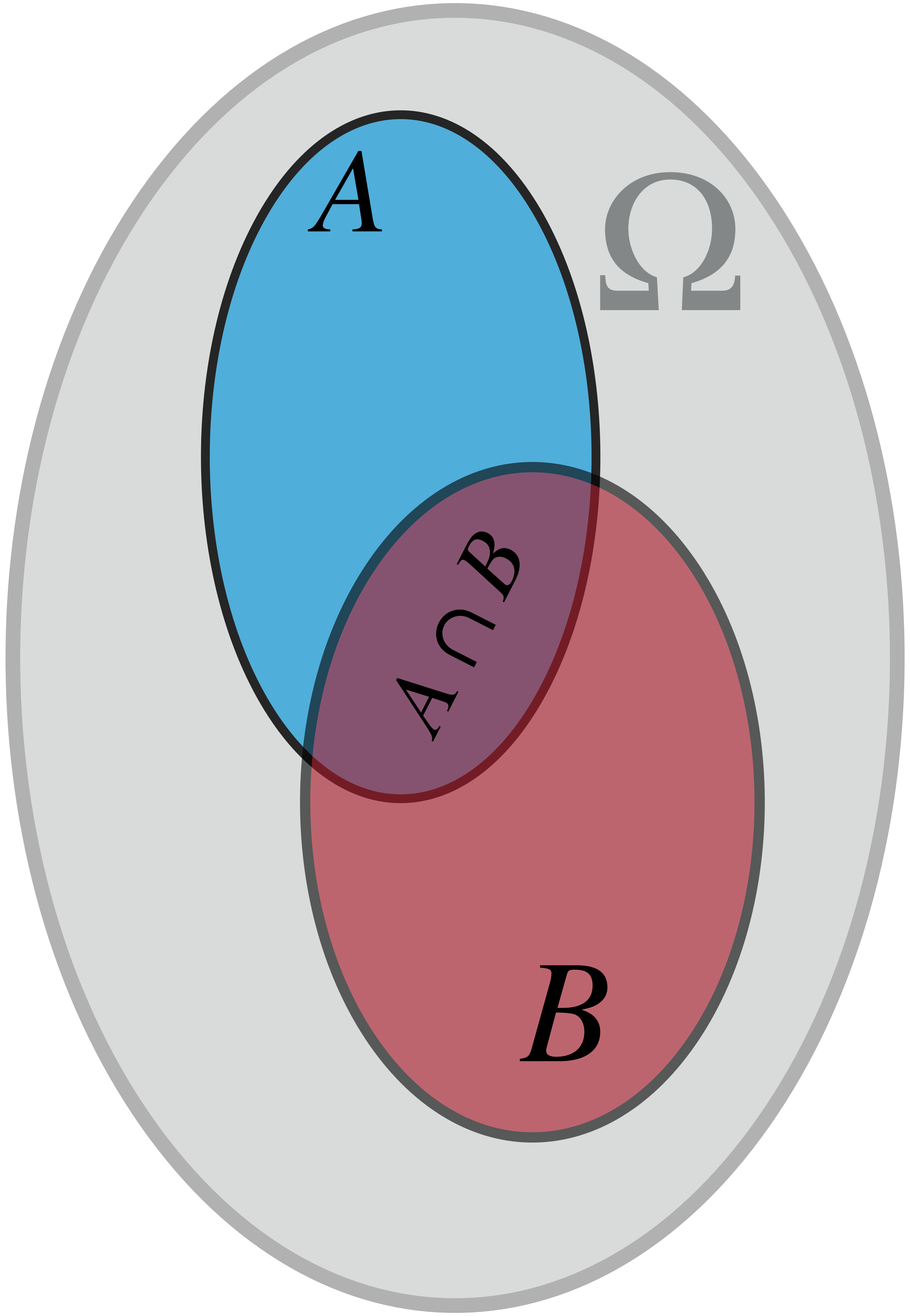


Let Ω be a collection of possible elementary events, and A and B events such that $A, B \in \Omega$

- $P(A) \geq 0$ for all A;
- $P(\Omega) = 1$;
- $$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

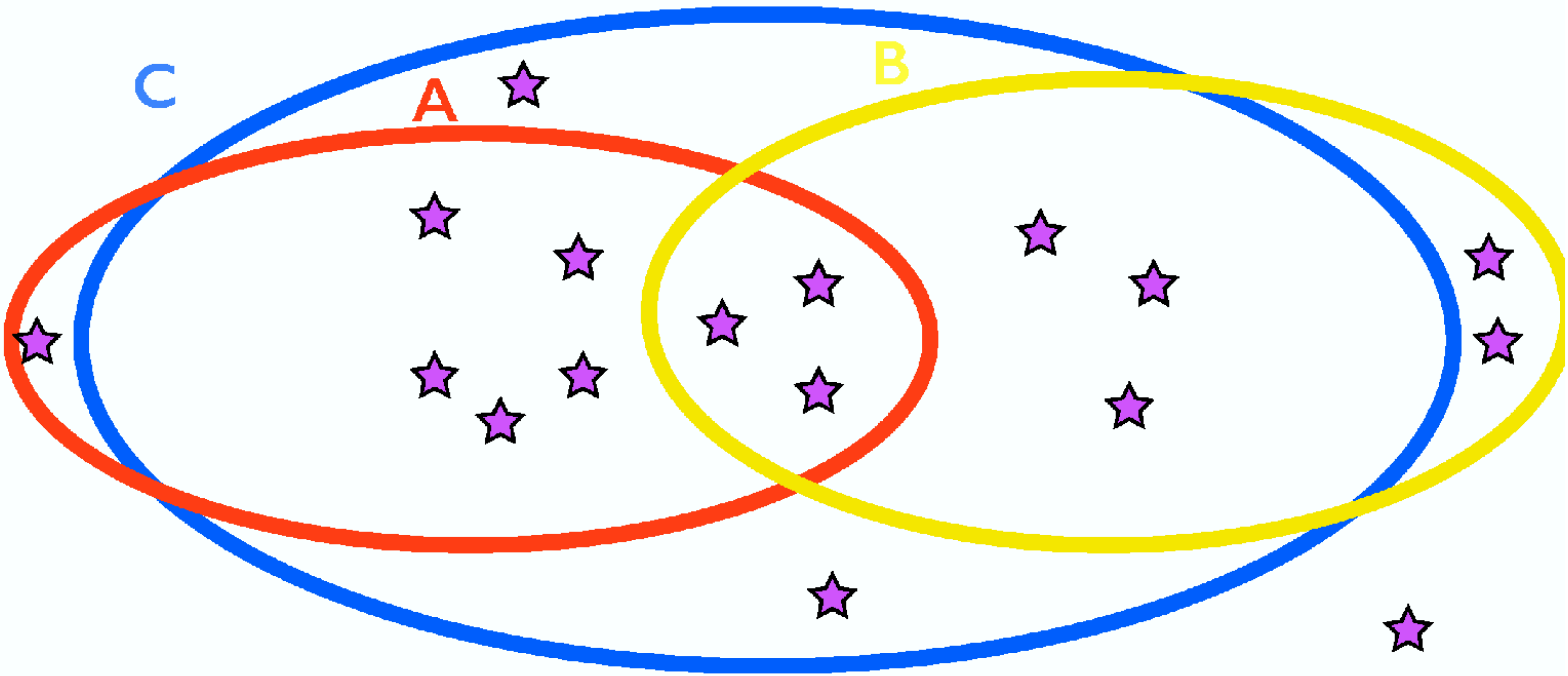
  for all countable disjoint sets $A_1, A_2 .. \in \Omega$
- $$P(A) = \sum_{i=i}^{\infty} P(A \mid B_i) \cdot P(B_i)$$

The law of total probability



Rules of probability

- $P(A \cup B|C) = P(A|C)+P(B|C)-P(A \cap B|C)$
- $11/13 = 8/13 + 6/13 - 3/13$



Courtesy Jo Bovy

A real world example:

Schawinski et al.

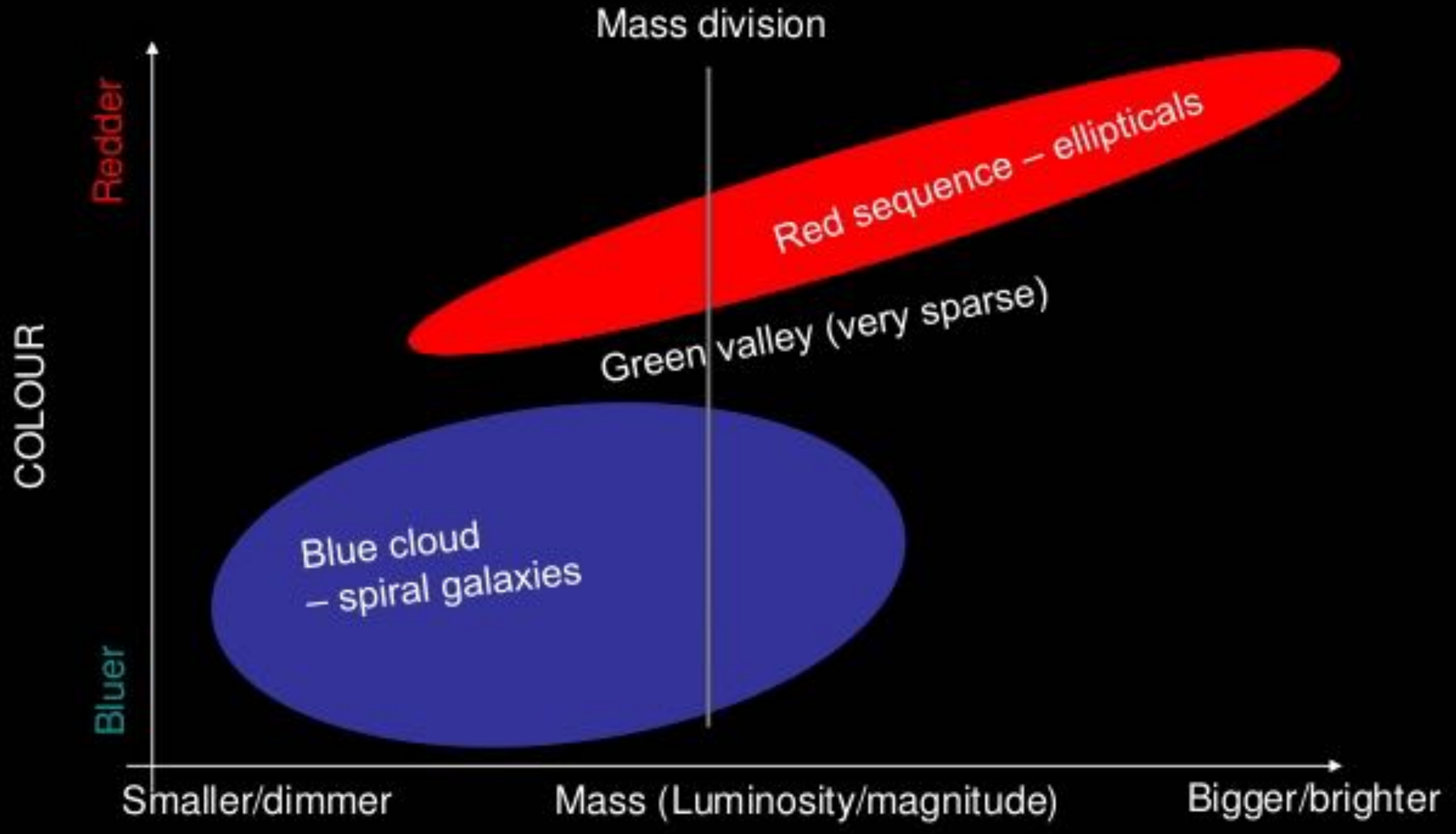
“The Green Valley is a Red Herring: GalaxyZoon reveals two evolutionary pathways towards quenching of star formation in early- and late-type galaxies”

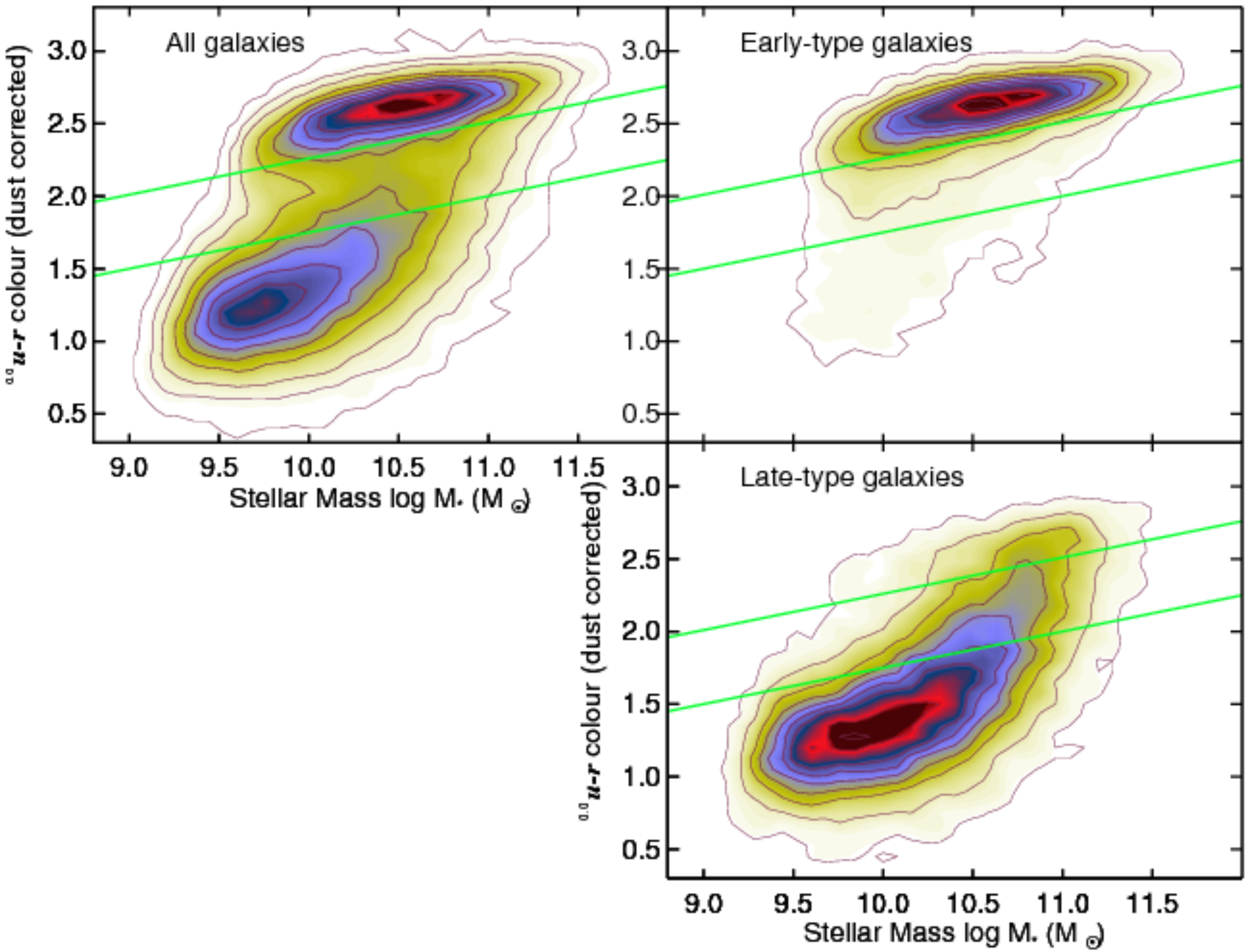
MNRAS 440, no. 1, 889-907 (2014)

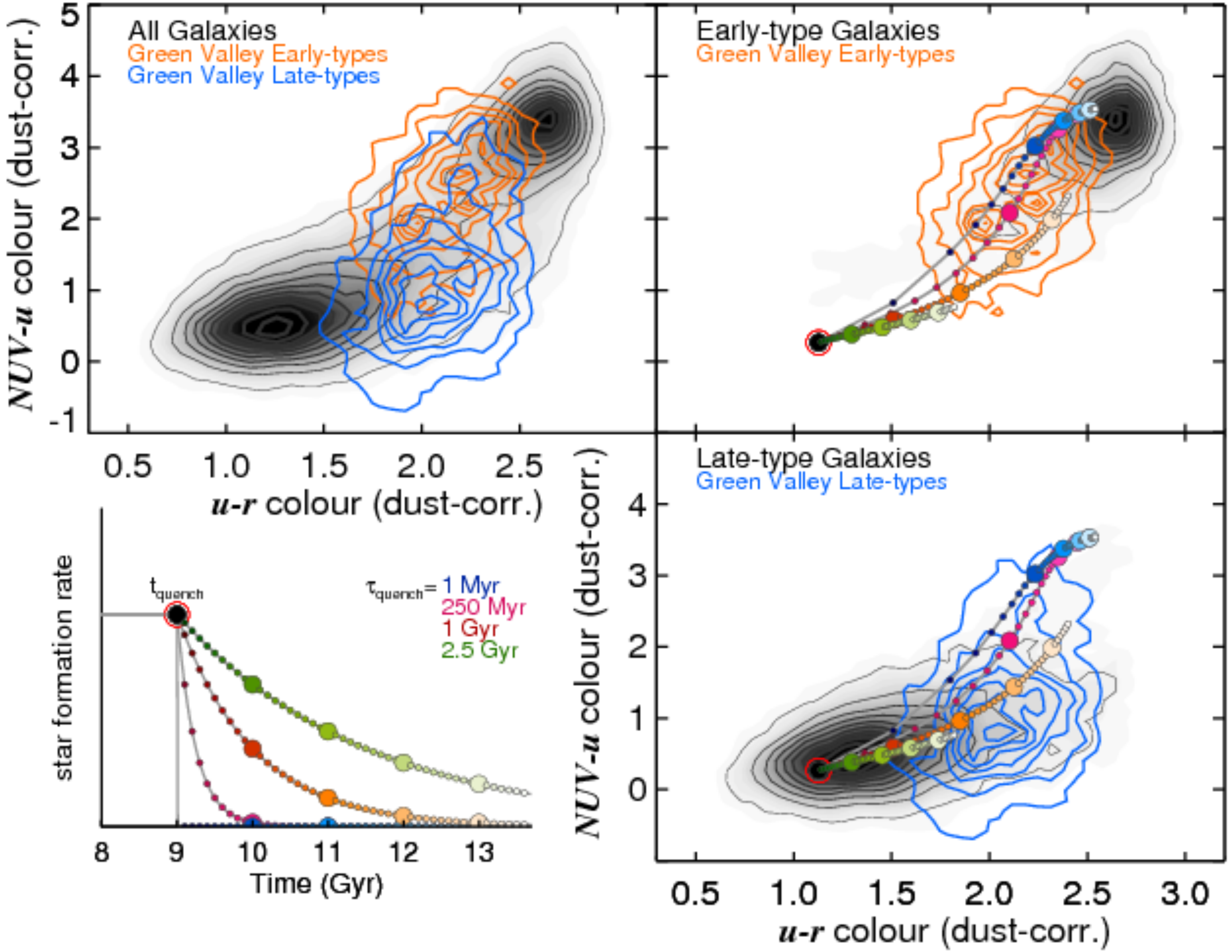
arXiv 1402.4814


early, blue
early, green
early, red
indeterminate, blue
indeterminate, green
indeterminate, red
late, blue
late, green
late, red

Colour Magnitude Diagram




All galaxies
Early-type galaxies
Late-type galaxies
$^{0.0}u$-r colour (dust corrected)
Stellar Mass log M* (M ☉)
0.5
1.0
1.5
2.0
2.5
3.0
9.0
9.5
10.0
10.5
11.0
11.5



All Galaxies
Green Valley Early-types
Green Valley Late-types
Early-type Galaxies
Green Valley Early-types
Late-type Galaxies
Green Valley Late-types
NUV-u colour (dust-corr.)
u-r colour (dust-corr.)
star formation rate
t_{quench}
τ_{quench}= 1 Myr
250 Myr
1 Gyr
2.5 Gyr
Time (Gyr)

SOME KEY IDEAS

- All of the data we collect include some degree of randomness
- Any conclusions we draw must therefore incorporate some notion of uncertainty
- There is a a correct answer - the Universe as we know it exists after all.
    - Theory gives us a useful model for it. The challenging is evaluating how likely that model is given the data
- Data are constants.
    - Even if they were randomly generated by the Universe, the data that we have already collected are fixed numbers.
- We describe things we don't know with perfect precision as "random"

1.2

RANDOM VARIABLES

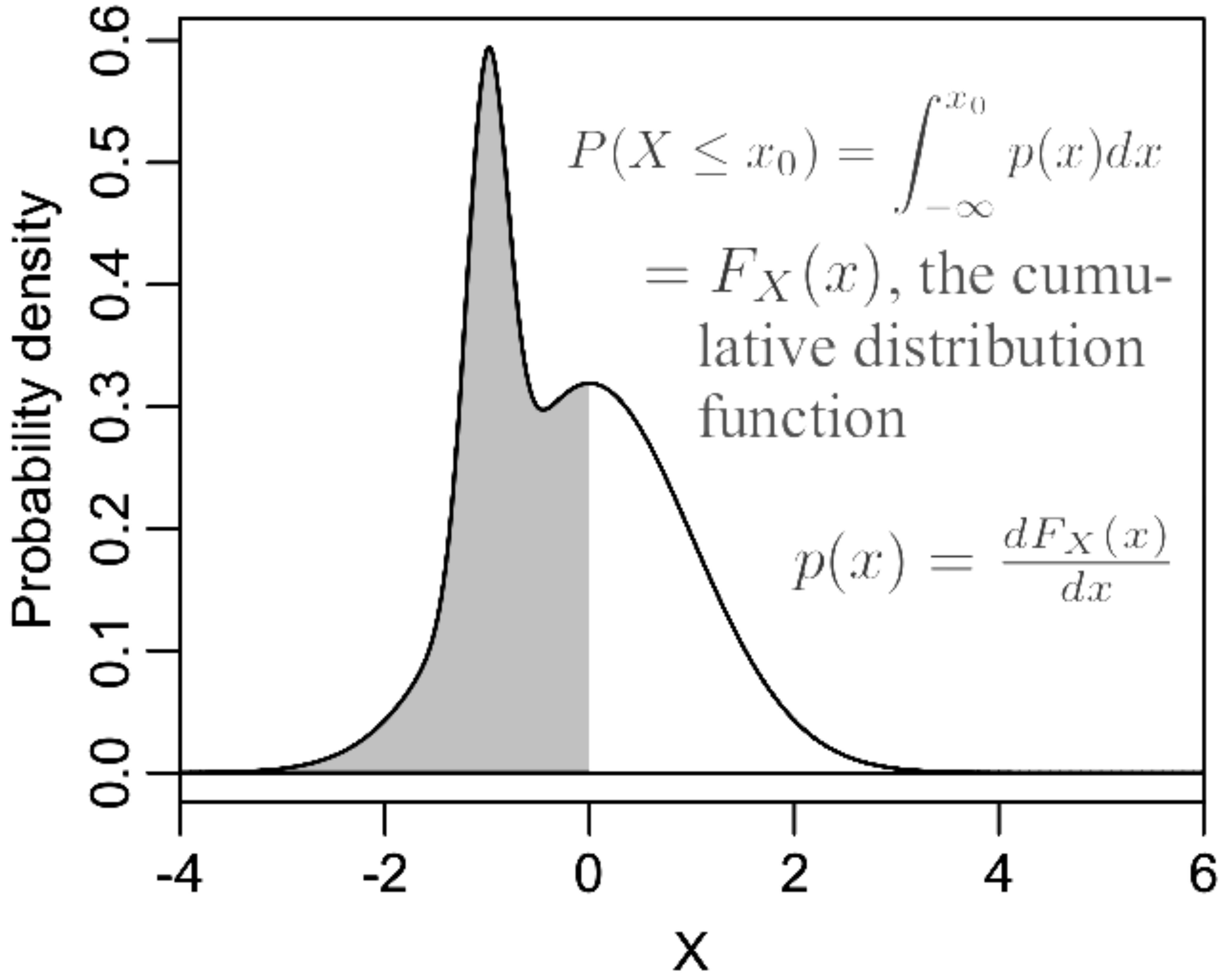
Random variable:

Modified from Maria Suveges, Laurent Eyer

the outcome of an “experiment”, with a probability for each outcome

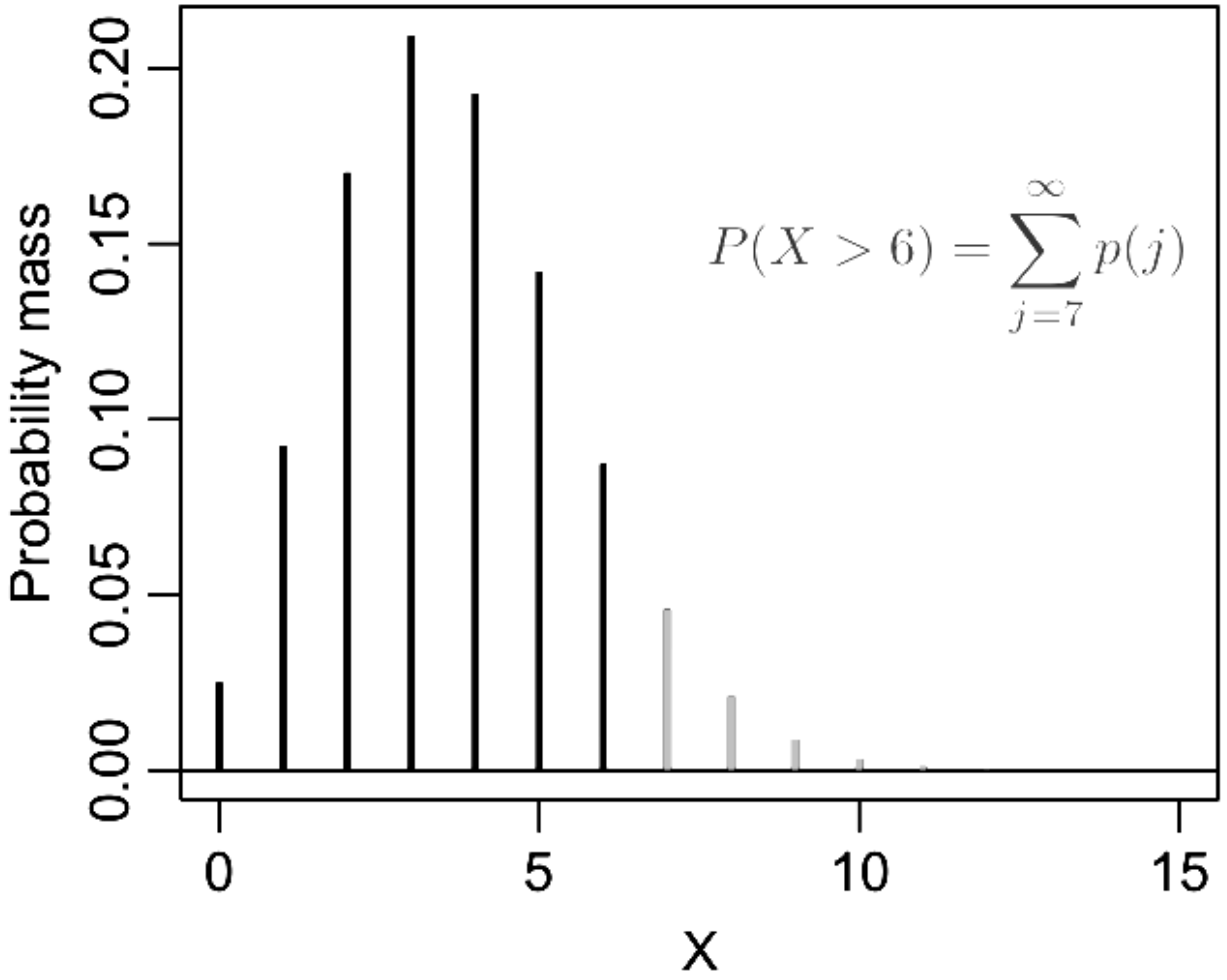
Continuous

$$P(X \in A) = \int_A p(x)dx$$



Discrete

$$P(X \in A) = \sum_{x_j \in A} p(x_j)$$



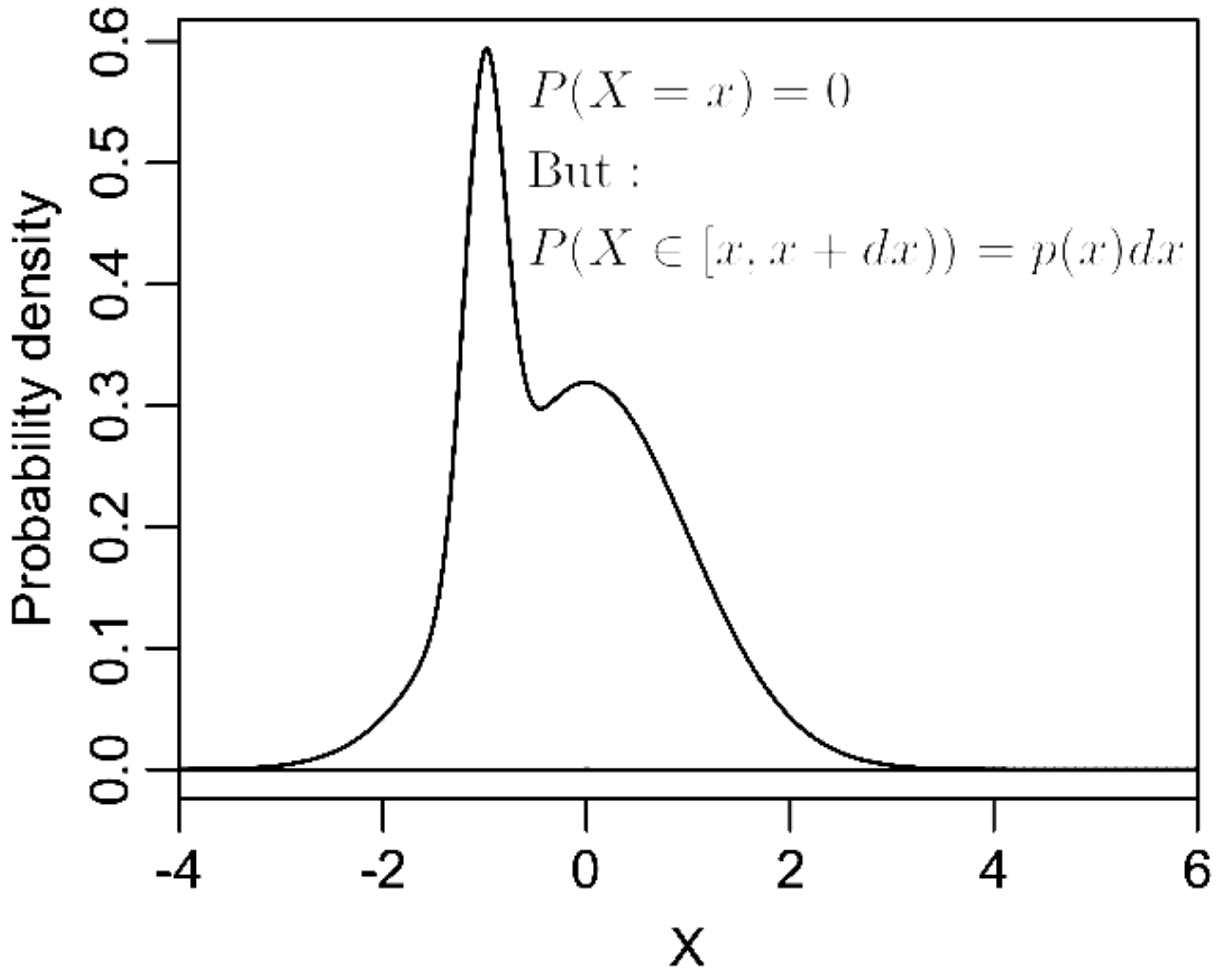
Modified from Maria Suveges, Laurent Eyer

Random variable:

the outcome of an “experiment”, with a probability for each outcome

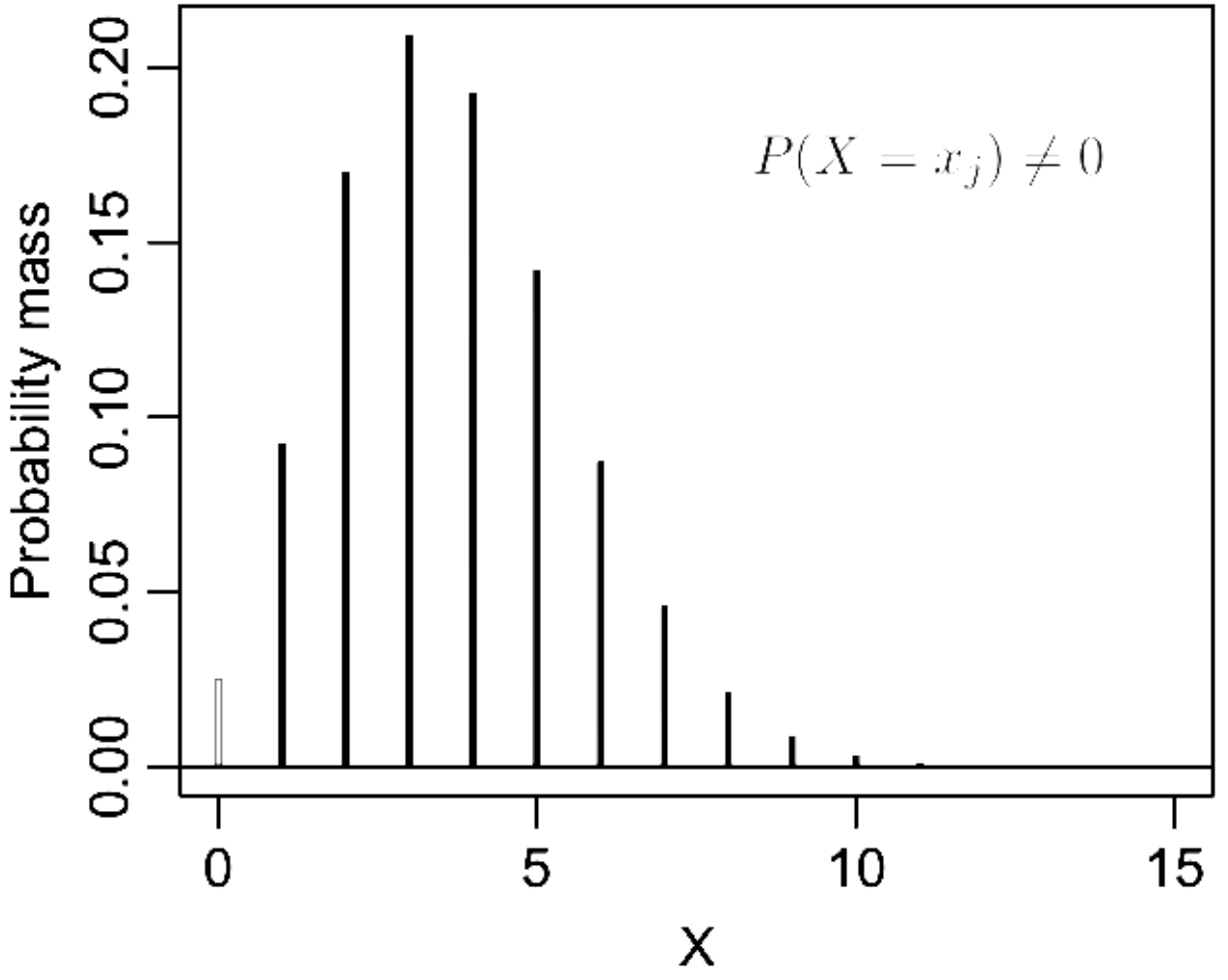
Continuous

$$P(X \in A) = \int_A p(x)dx$$



Discrete

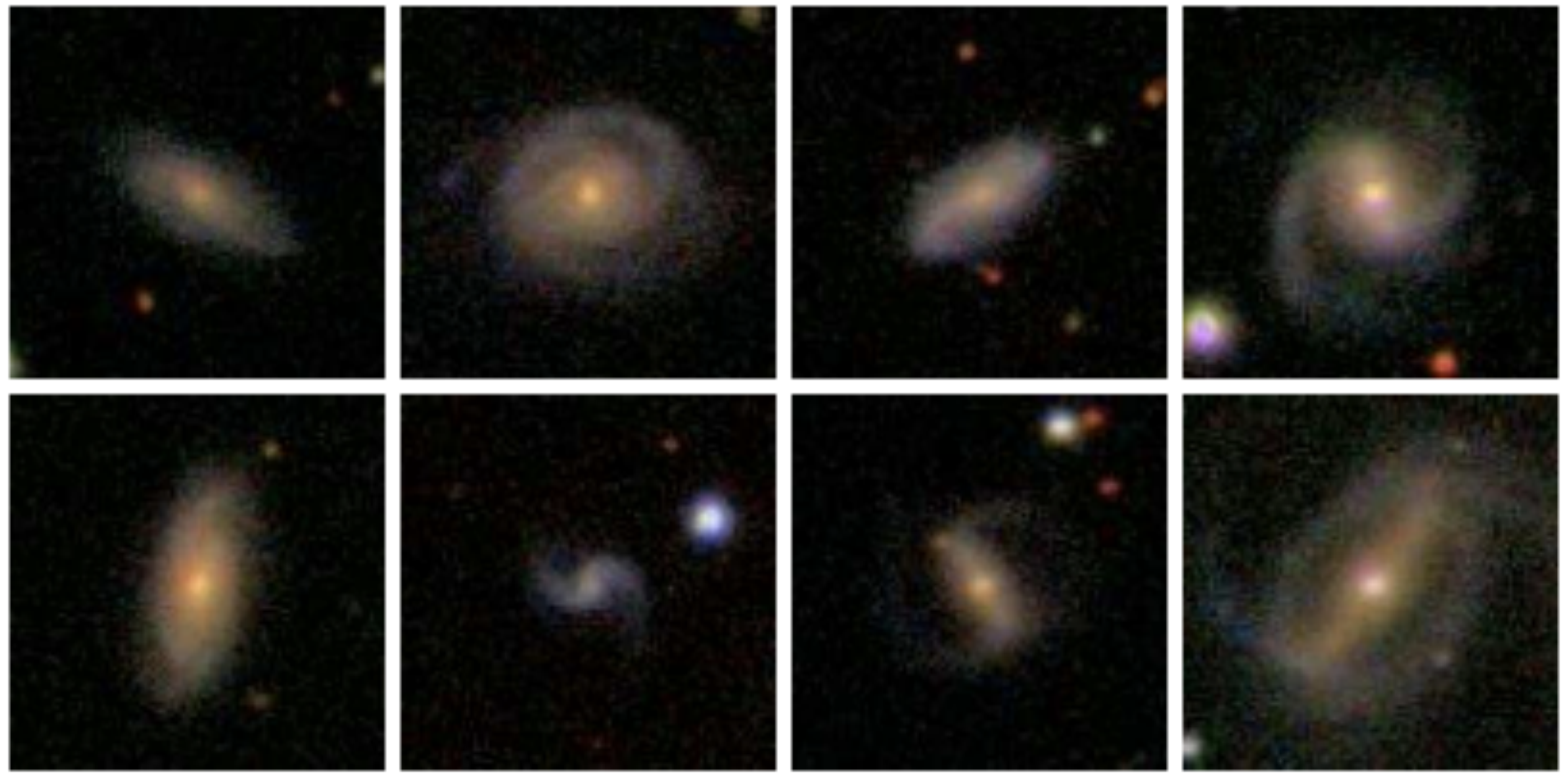
$$P(X \in A) = \sum_{x_j \in A} p(x_j)$$



Random Variables

Modified from Maria Suveges, Laurent Eyer

- Discrete:
  - Spectral type (G2V, KIII)
  - Galaxy type, galaxy zoo



Class	Button	Description
1		Elliptical galaxy
2		Clockwise/Z-wise spiral galaxy
3		Anti-clockwise/S-wise spiral galaxy
4		Spiral galaxy other (e.g. edge on, unsure)
5		Star or Don't Know (e.g. artefact)
6		Merger

- Continuous:
  - magnitude, flux, colour, radial velocity, parallax/distance, temperature, elemental abundances, magnetic field, age, etc...

We are generally trying to estimate $p(x)$, the *true* distribution from which x is drawn.

$p(x)$ is the "Probability Density Function" of x.

$p(x) \cdot dx$ is the probability of a value lying between x and $x + dx$.

While $p(x)$ is the true or population pdf, we don't observe it.

We measure the *empirical* pdf, $f(x)$.

With infinite data $f(x) \to p(x)$.

Not really.

So when we say:

Statistical inference is a logical framework with which to test our beliefs of a noisy world against data.

We formalize our beliefs in a probabilistic model.

What that means we're doing:

Estimate $f(x)$ from (possibly multi-D) data.

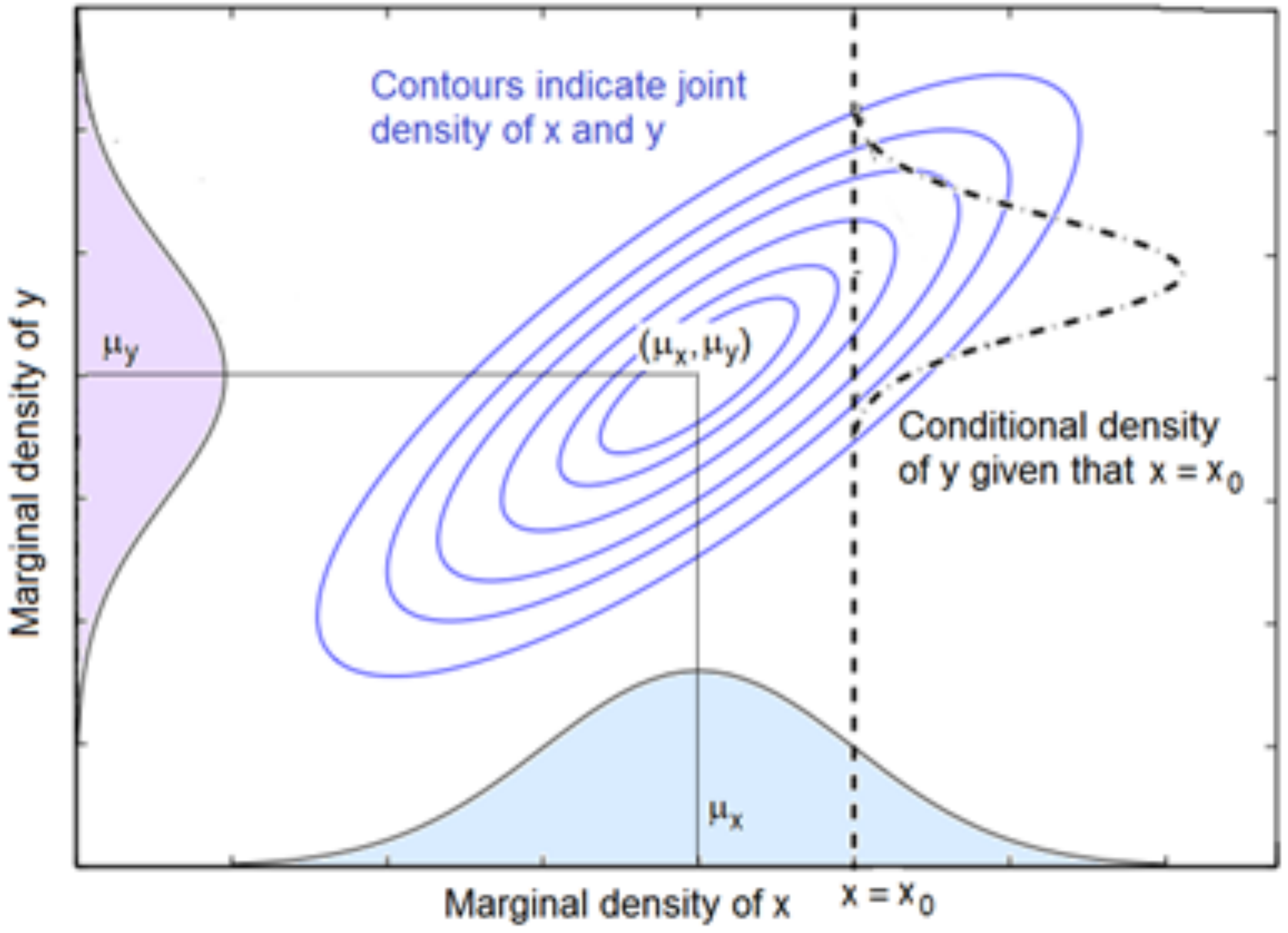
Describe $f(x)$ and its uncertainty.

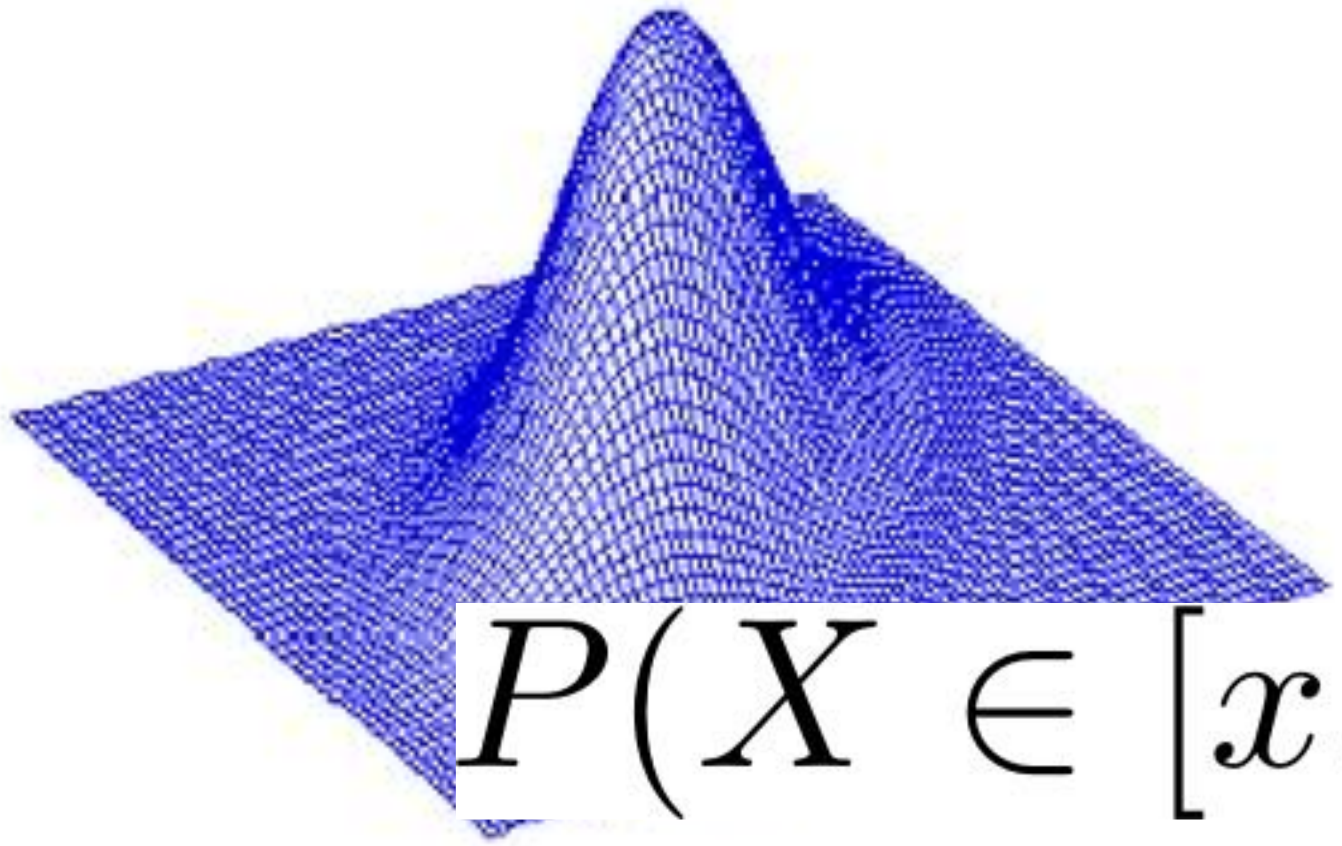
Compare it to models of $p(x)$

and use the knowledge gained
to interpret new measurements

$$P(X \in A) = \int_A p(x_i, x_2, ..., x_N) dx_1 dx_2 ... dx_N$$

THE MULTIVARIATE CASE





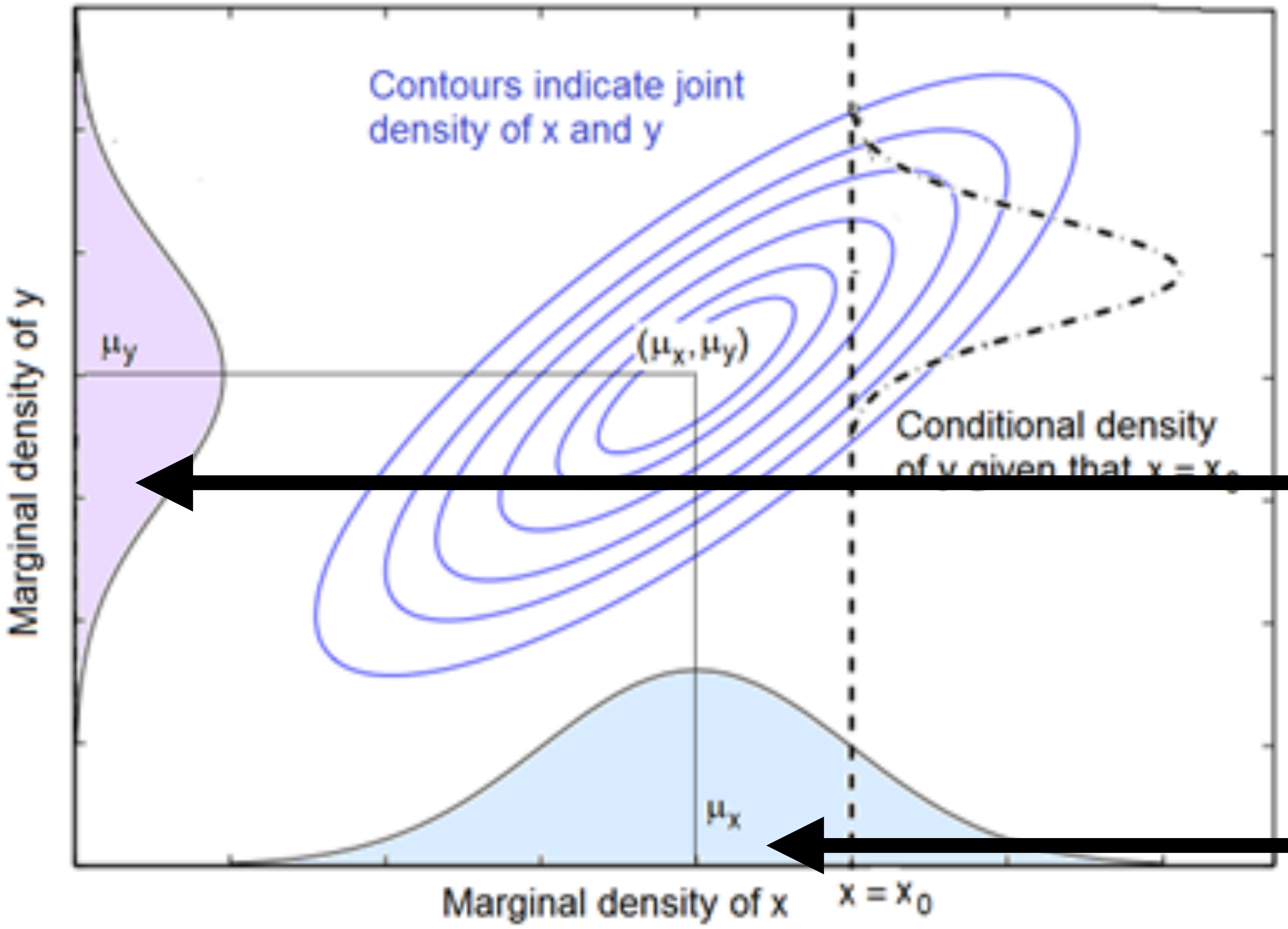
Given this in 1D,

the probability of *X*

in a 2D box of area *dx.dy* is:

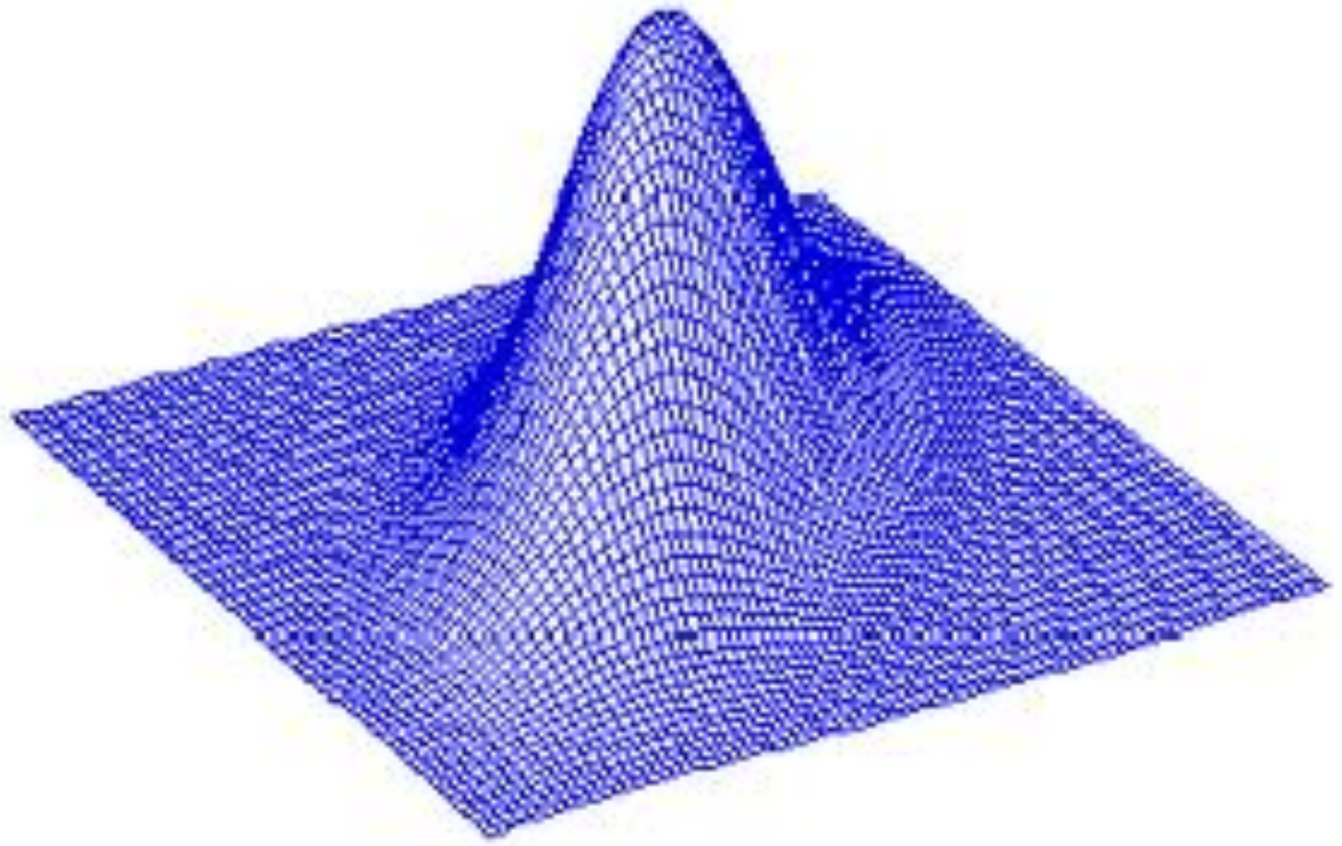
$$P(X \in [x, x+dx] \times [y, y+dy]) = p(x,y) dx dy$$

THE MULTIVARIATE CASE



$$p(y) = \int_X p(x, y)dx$$

$$p(x) = \int_Y p(x, y)dy$$



Conditional Probability

$$p(x, y) = p(x|y)p(y)$$

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

$$p(x) = \int_y p(x, y)dy$$

$$p(x) = \int_y p(x|y)p(y)dy$$

The law of total probability

Bayes Rule:

$$p(y|x) = \frac{p(x|y)p(y)}{\int_Y p(x|y)p(y)dy}$$

$$p(y|x) = \frac{p(x|y)p(y)}{\int_Y p(x|y)p(y)dy}$$

Posterior

How probable is the hypothesis given the data we observed

Likelihood

How probable is the data given the hypothesis is true

Prior

How probable was the hypothesis before we observed anything

$$p(\text{Hypothesis}|\text{Data}) = \frac{p(\text{Data}|\text{Hypothesis})p(\text{Hypothesis})}{p(\text{Data})}$$

Evidence

How probable is the data over all possible hypotheses

IN CLASS EXERCISE

- Download this file (too big for git!): https://bit.ly/38PDnGy
  - Use h5py to look at this data - h5py.File() to open, and then use the keys() method to find what elements are store
  - You want "**chain**" and then "**position**"
- Use numpy to get the stored data as an array
- Use matplotlib to visualize this point cloud (CAREFUL)
- Use pandas to convert the first two columns of the numpy array to a dataframe
- Use seaborn's jointplot to visualize this dataframe (try hex, or a kde with every 100th sample)

1.3

MOMENTS AND DISTRIBUTIONS

Expected Value

$$E(x) = \langle x \rangle = \int_X x \cdot p(x) dx$$

$$E(f(x)) = \langle f(x) \rangle = \int_X f(x) \cdot p(x) dx$$

Variance

$$\mathrm{Var}(x) = E([x - \langle x \rangle]^2)$$

nth moment (non-central)

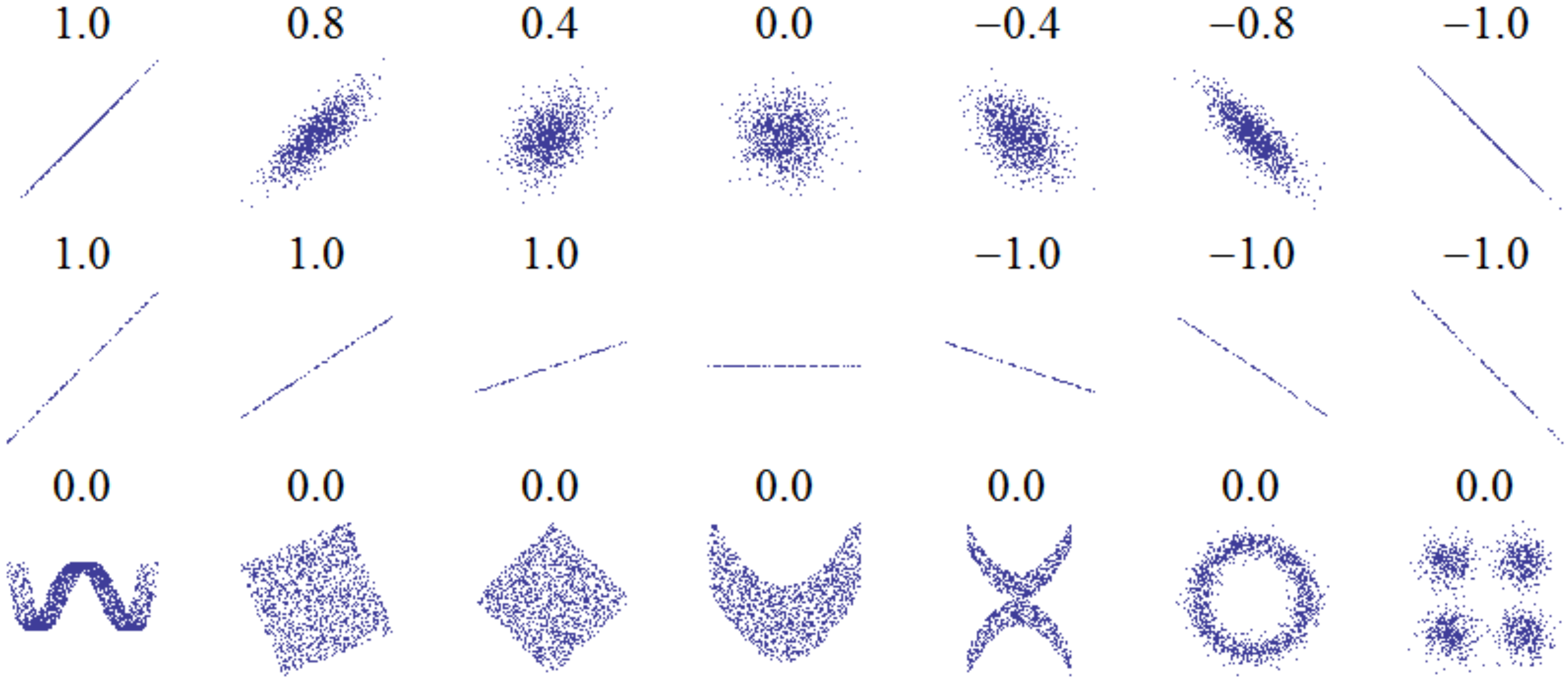
$$\mu_n(x) = E(x^n)$$

nth moment (central)

$$\widetilde{\mu}_n(x) = E([x - \langle x \rangle]^n)$$

$$\mathrm{Cov}(x,y) = E([x - \langle x \rangle])E([y - \langle y \rangle])$$

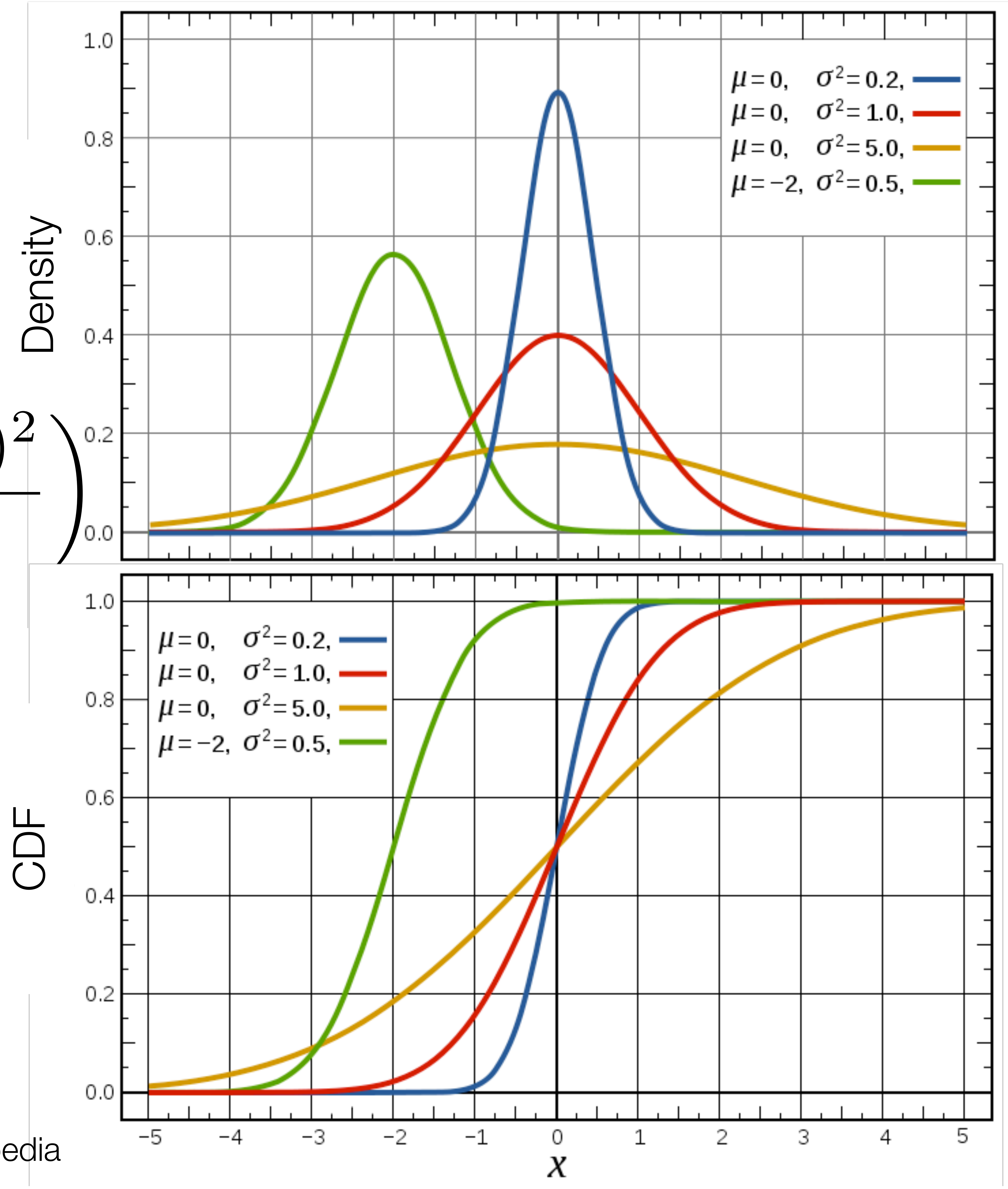
$$\mathrm{Cor}(x,y) = \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}(x)\mathrm{Var}(y)}}$$



Example: Gaussian / Normal distribution

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



from wikipedia

Poisson distribution

Discrete probability distribution (no density)

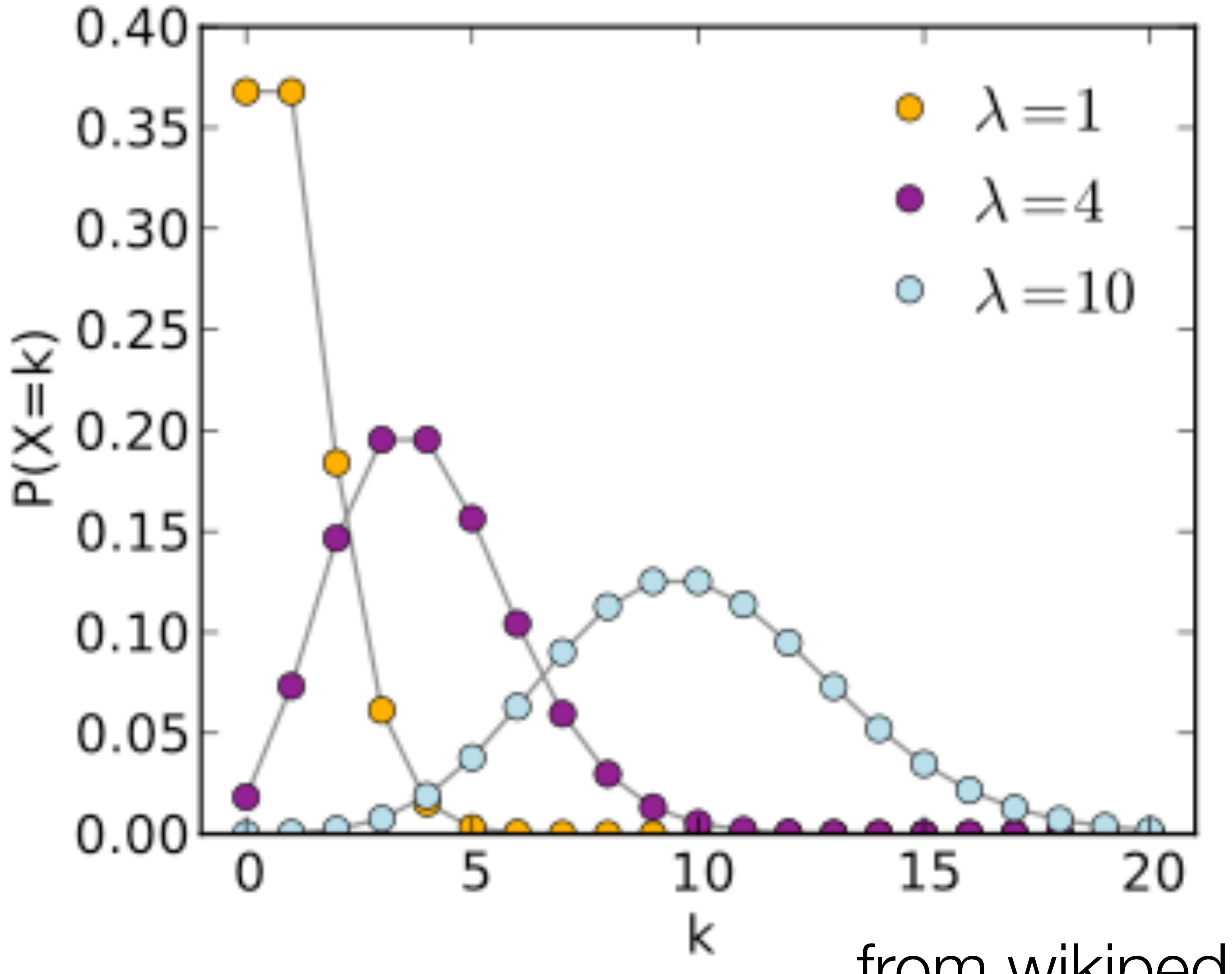
$$X \sim \text{Poisson}(\lambda)$$

Number of photons on a detector
Number of people in a shop

$$\Pr(X = k) = \exp(-\lambda)\frac{\lambda^k}{k!}$$

For large λ

$$\mathcal{N}(\lambda, \lambda)$$



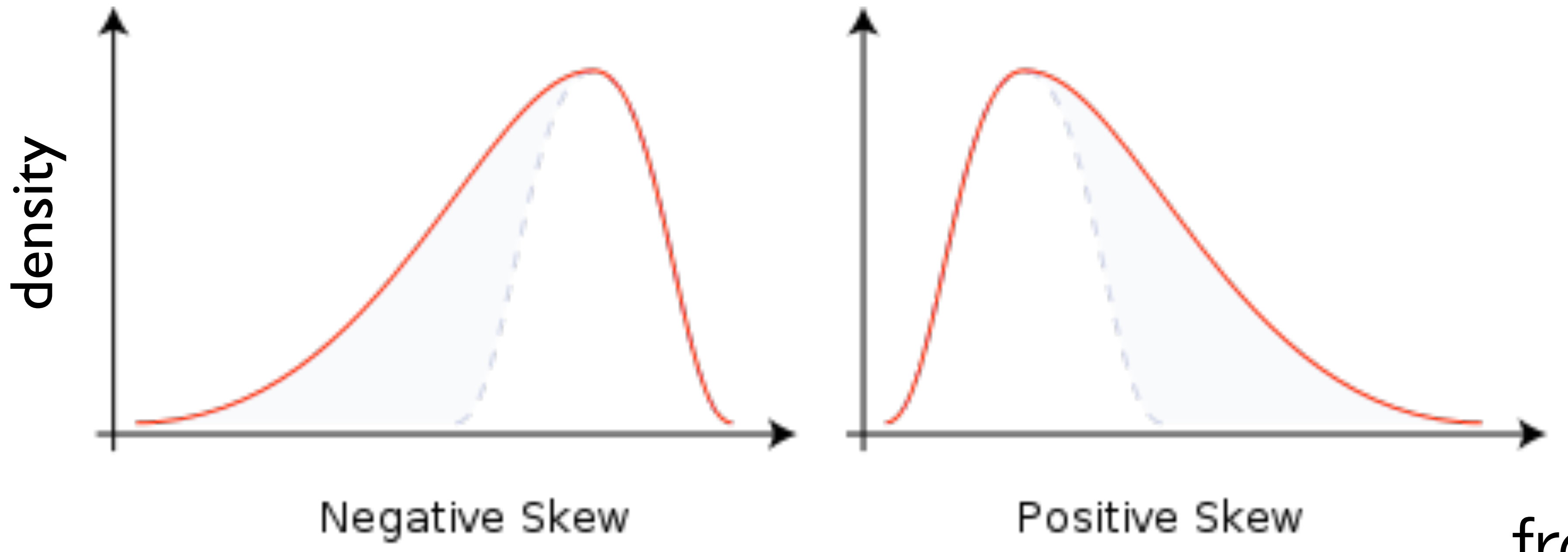
from wikipedia

3d and 4th moments of a distribution

- Skewness, asymmetry

$$\mu_3/\sigma^3 = \int_{-\infty}^{\infty} (x-\mu)^3 f(x)dx/\sigma^3$$

Normal: 0
Poisson: 1/√λ



from wikipedia

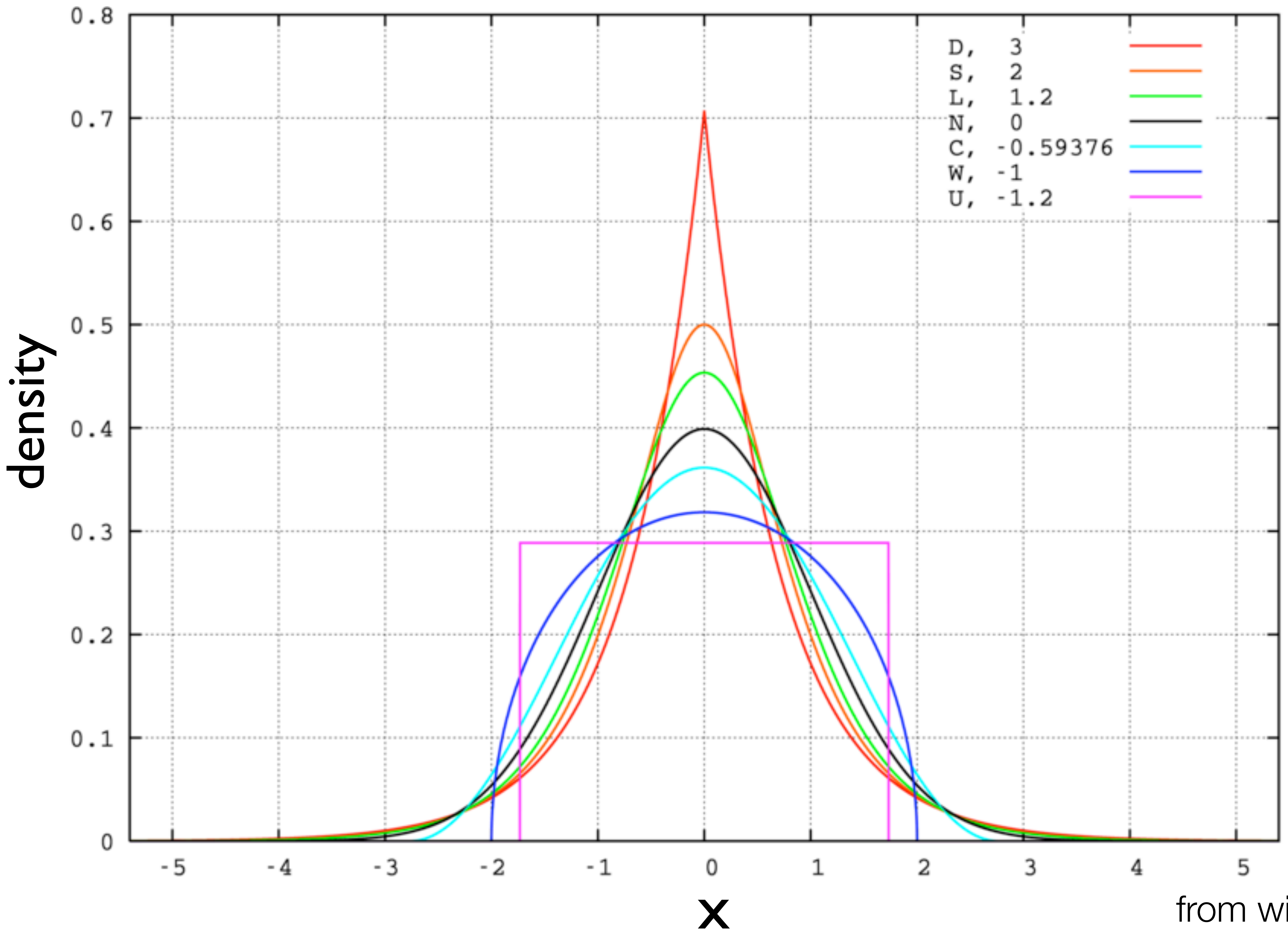
- Kurtosis

$$\mu_4/\sigma^4 = \int_{-\infty}^{\infty} (x-\mu)^4 f(x)dx/\sigma^4$$

$$\mu_4/\sigma^4 - 3$$

Normal: 0
Poisson: 1/λ

Example of different values of kurtosis: "boxiness" -- tail heaviness



from wikipedia

Quantiles

- x_p : p-quantiles of f(x)

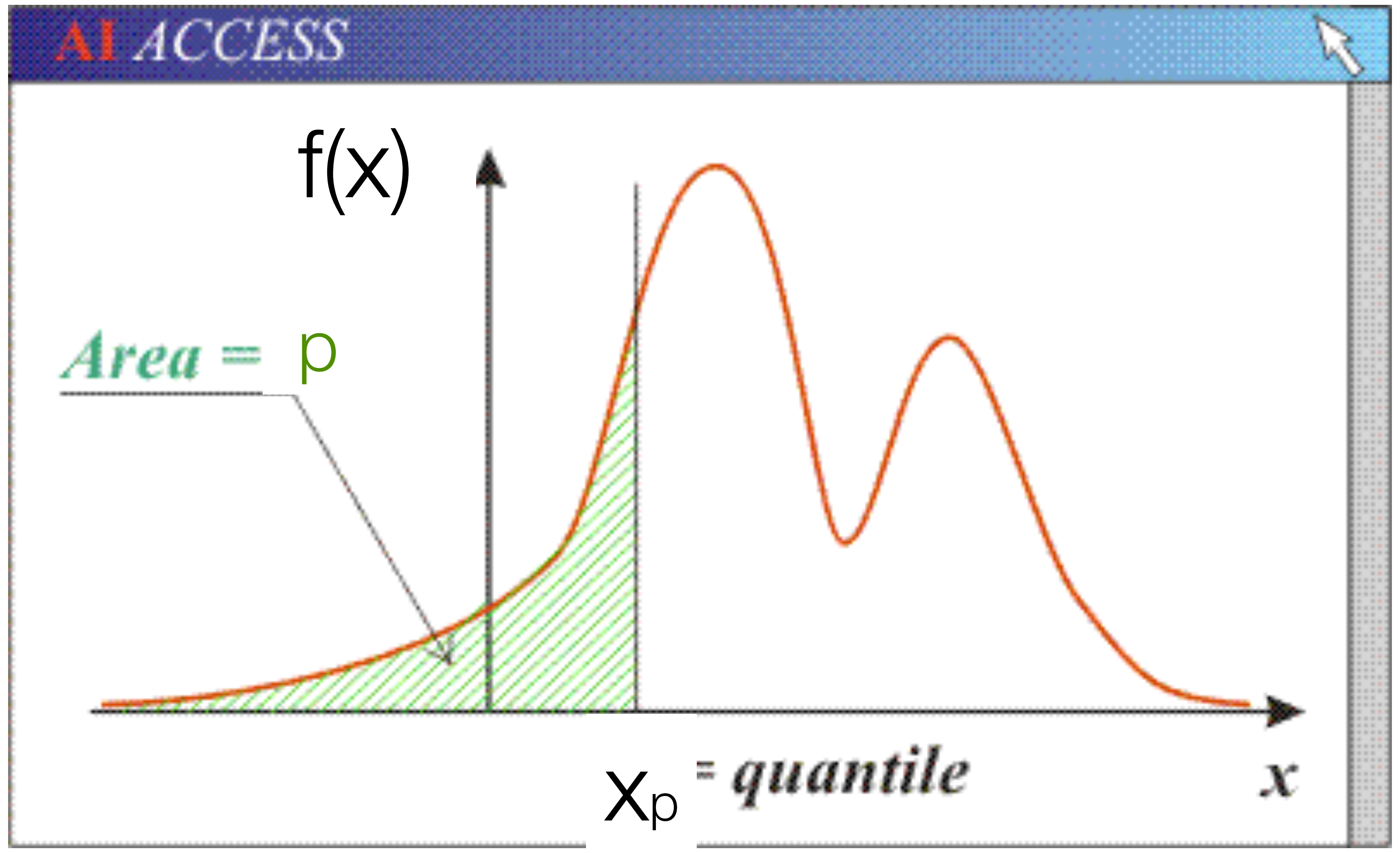
$$p = \int_{-\infty}^{x_p} f(x)dx$$

- Measure of location: Median

$$1/2 = \int_{-\infty}^{x_{1/2}} f(x)dx$$

- Measure of dispersion: Inter-quantile range

$$\mathrm{IQR} = x_{3/4} - x_{1/4}$$



from www.aiacces.net

Data, samples

- Usually we have observations, e.g. additive process

$$y_i = f(t_i) + \epsilon_i \qquad\qquad i = 1, \ldots, n$$

Deterministic random variable

- We want a characterisation of the deterministic and random parts

- Suppose something about the random variable, often normality: $\mathcal{N}(0, \sigma^2)$

- Assumption of models

- Estimate the parameters of a distribution, moments

- Exercise 1: Sample mean: $$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad E(\bar{X}) = \mu$$

- Exercise 2: Sample variance (bias):

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad E(\hat{\sigma^2}) = \frac{n}{n-1}\sigma^2$$

$$\xrightarrow{\text{redefine}} \quad \hat{\sigma^2} = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- Sample quantiles are estimators of quantiles

Central limit theorem

The distribution of the mean of a sufficiently large number of random variables can be approximated by a Gaussian distribution!

$$X_i,\ i = 1, \ldots, n \ \ \text{iid} \ \ \text{with} \ \ \mathrm{E(X_i)} = \mu \ \ \mathrm{Var(X_i)} = \sigma^2$$

iid= Independent identically distributed

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \text{follows approximately} \quad \mathcal{N}(0, 1)$$

$$\Downarrow$$

One reason why
the Gaussian distribution is so important

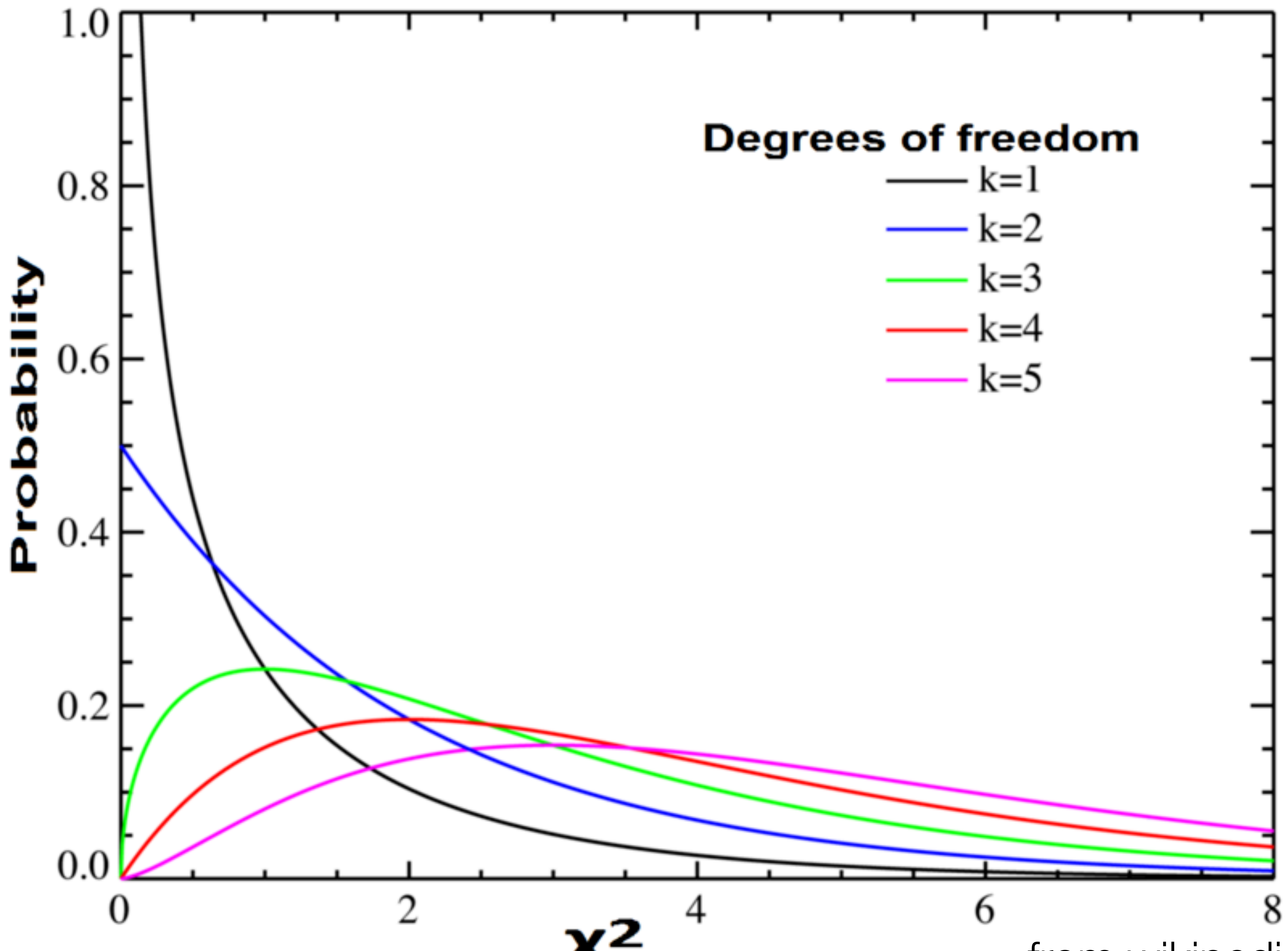
Distribution derived from Normal distribution
1) Chi square distribution

If $X_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ $\Rightarrow$ $$\sum_{i=1}^{k} X_i^2 \sim \chi_k^2$$

iid= Independent identically distributed

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} \exp\left(-x/2\right)$$

mean: k
variance: 2 k
skewness: √(8/k)
kurtosis: 12/k



from wikipedia

$$X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu,\sigma)$$

$$\sum_{i=1}^{k} (X_i - \bar{X})^2/\sigma^2 \sim \chi_{k-1}^2$$

When k is large χ_k^2 approximates a $\mathcal{N}(k, 2k)$

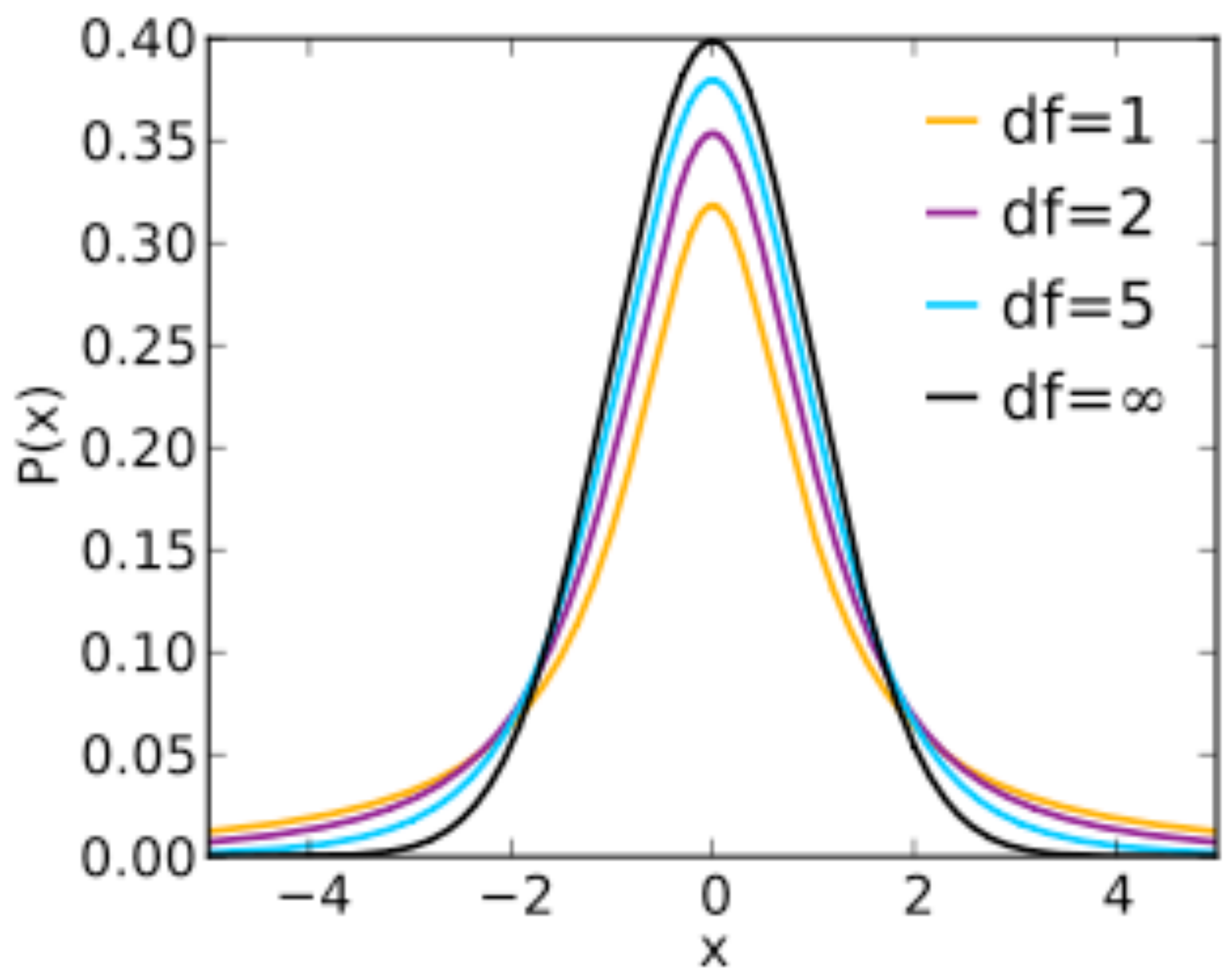
Distribution derived from Normal distribution
2) Student distribution

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

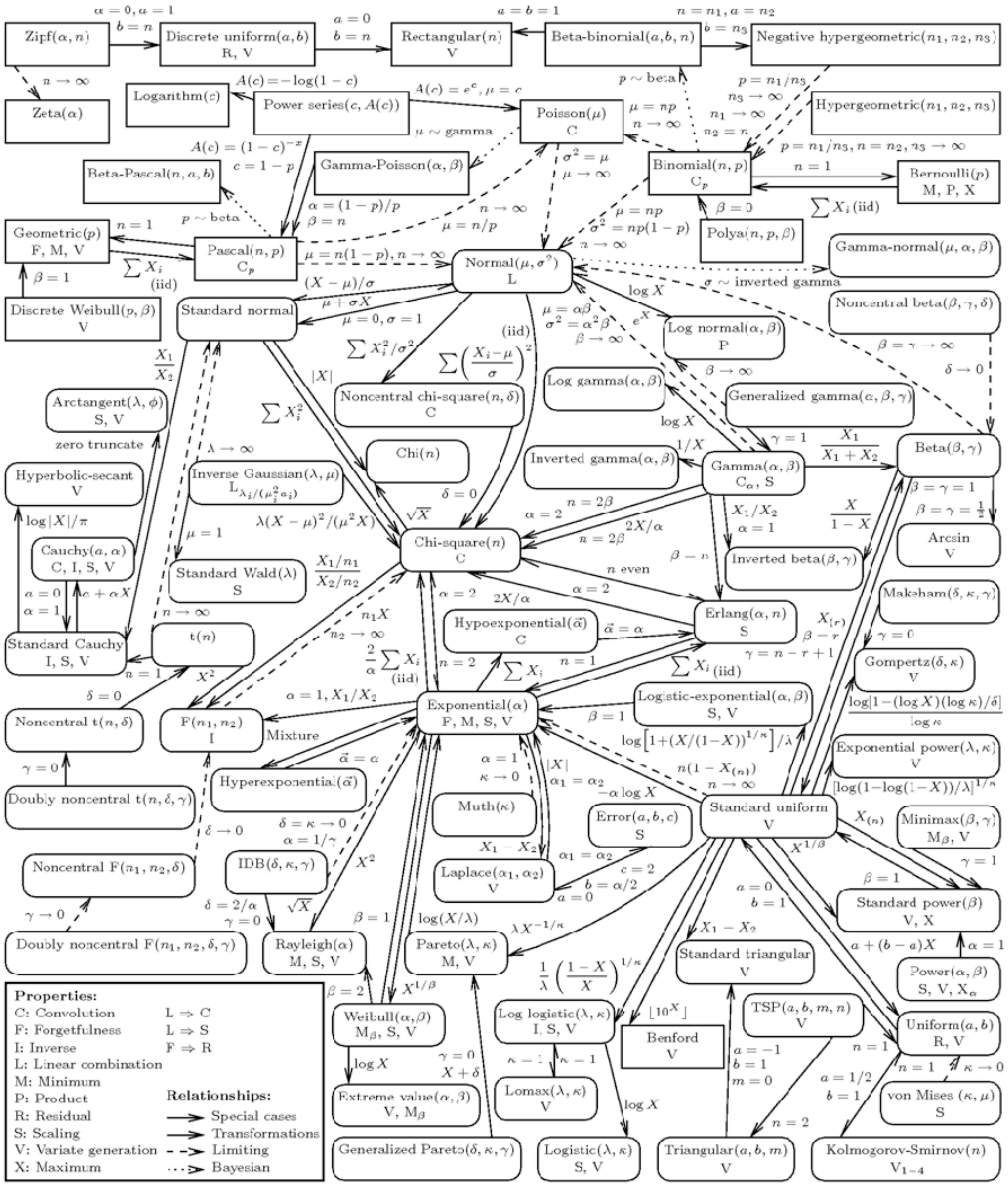
$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(n/2)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}$$

Note $t_\infty = \mathcal{N}(0,1)$



mean: 0 n>1
NaN n=0,1
variance: n/(n-2) n>2
∞ 1<n≤2
otherwise NaN
skewness: 0 n>3
kurtosis: 6/(n-4) n>4

Lawrence M Leemis & Jacquelyn T McQueston (2008)
"Univariate Distribution Relationships",
The American Statistician, 62:1, 45-53,
DOI: [10.1198/000313008X270448](https://doi.org/10.1198/000313008X270448)

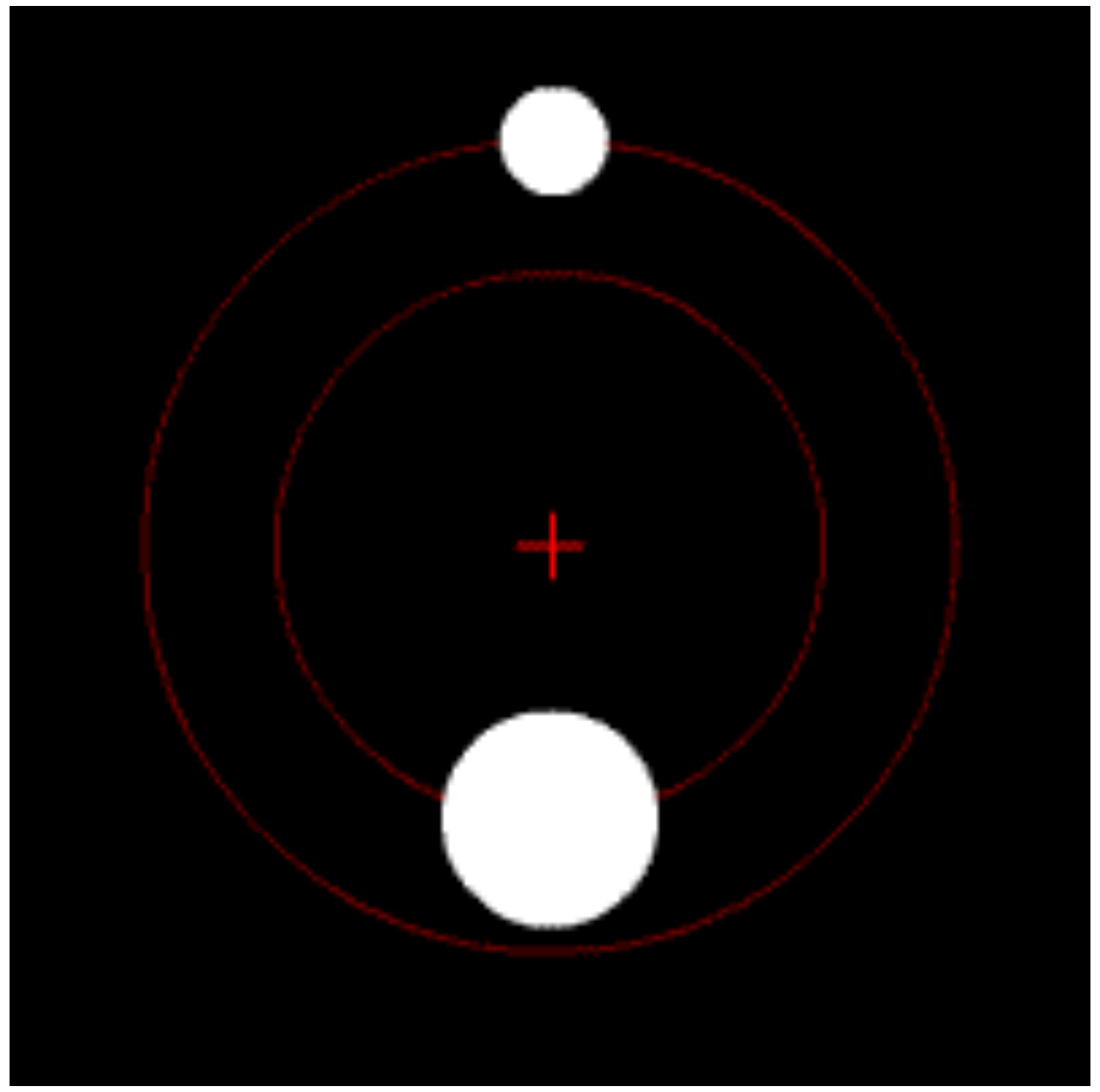


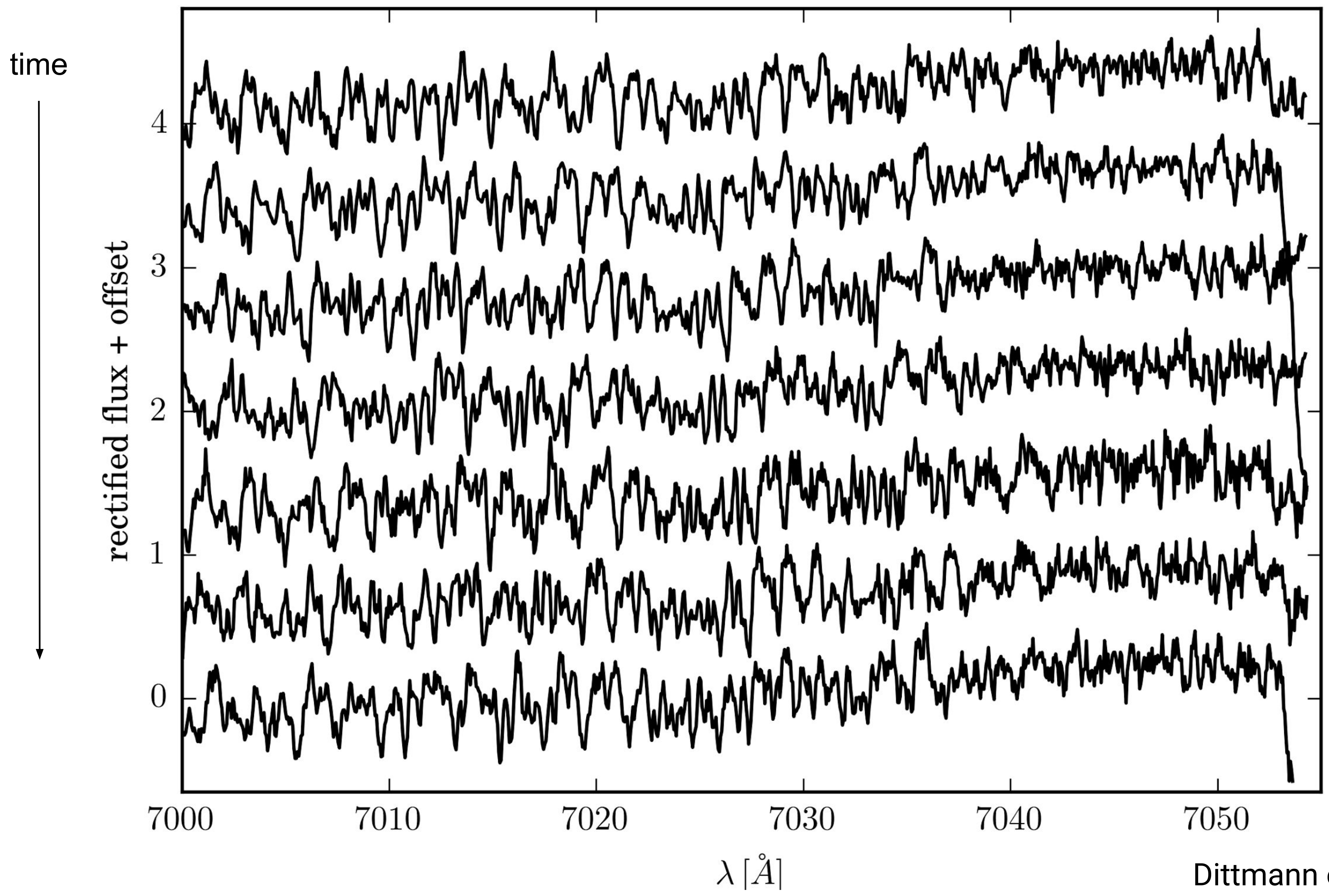
1.4

PUTTING THINGS TOGETHER

Astrostatistics Case Studies:
Disentangling Time Series Spectra with Gaussian Processes: Applications to Radial Velocity Analysis
(Czekala et al. 2017, ApJ, 840, 49. arXiv:1702.05652)

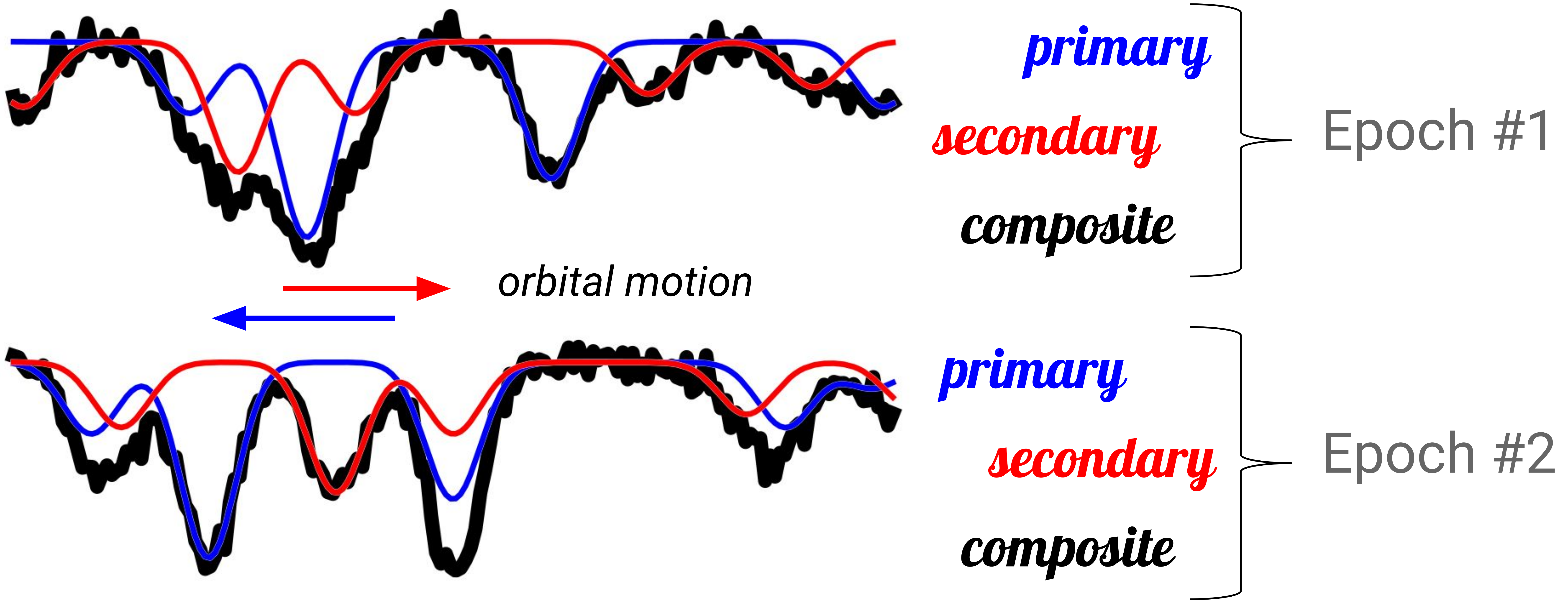
Raw Observations of the LP661-13 M4 Binary





Dittmann et al. 17
Czekala et al. 17a

Spectroscopic Binary Stars

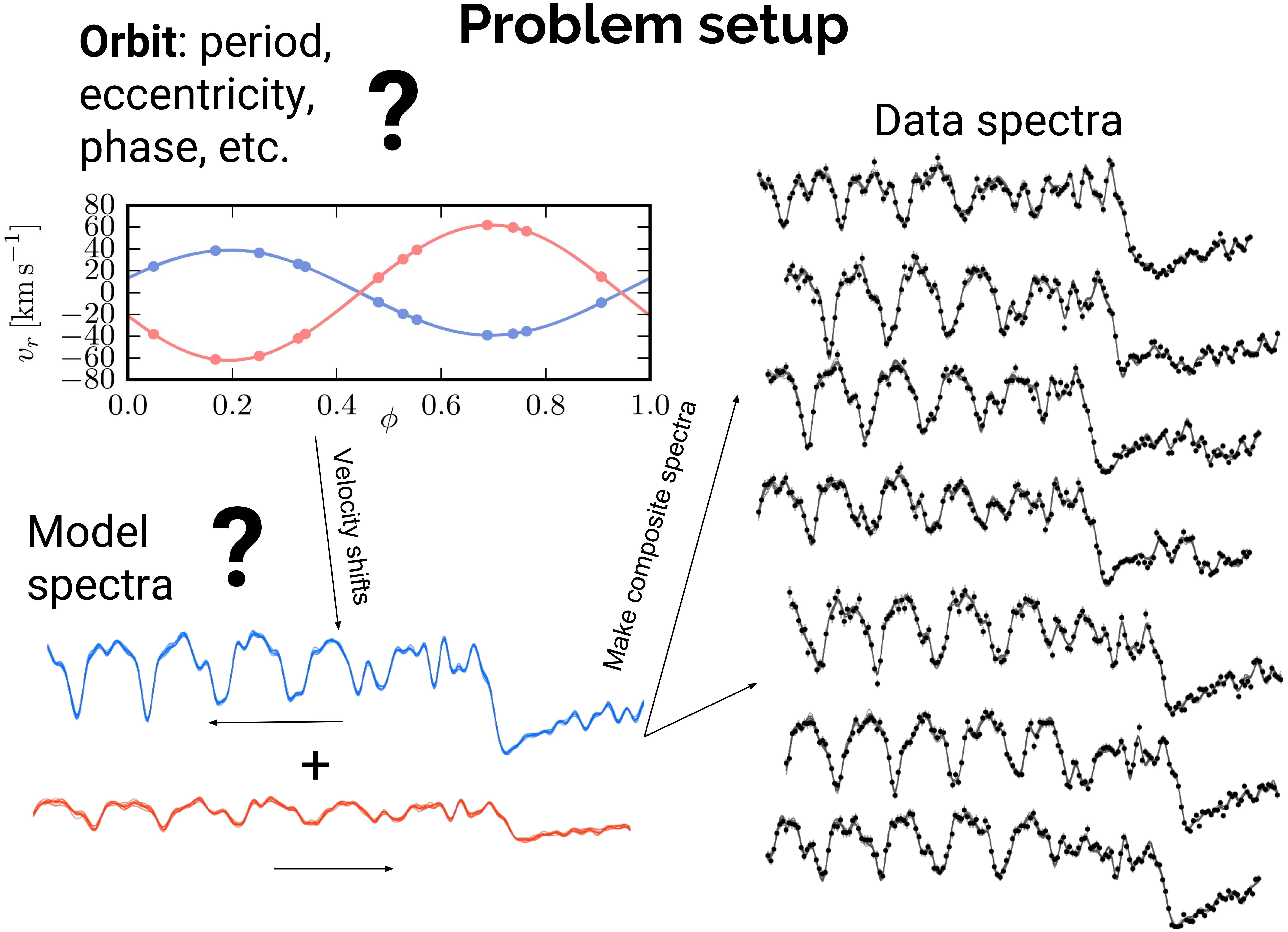


(Epoch = Time)

We only observe the “noisy” sum of two (latent) spectra.
Latent (underlying) spectra are unknown functions
Observed spectrum = Measured Data

Forward Model = Generates Data

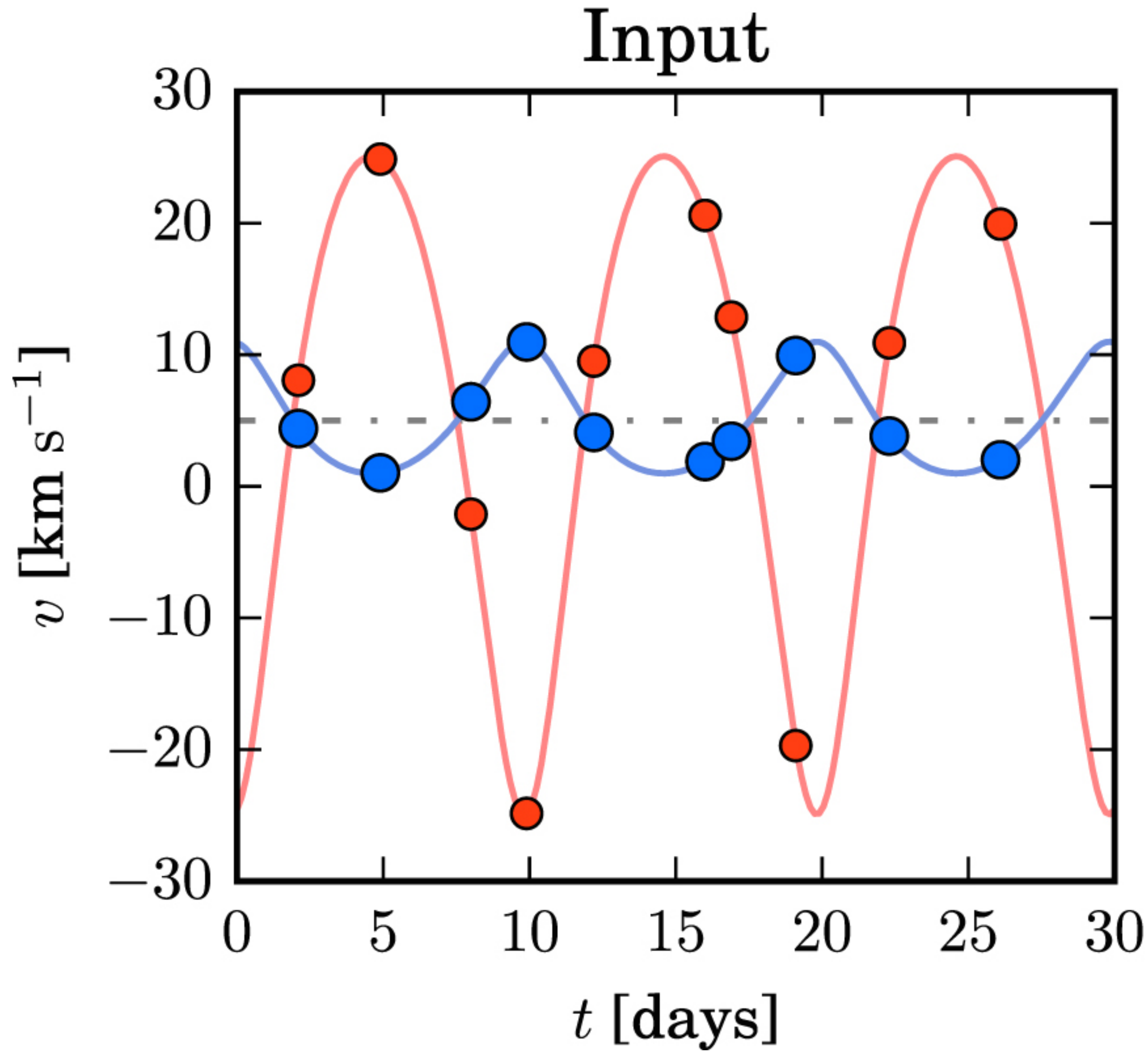
Problem setup



https://www.youtube.com/watch?v=kHjN42ft6aU

Goal: Go Backwards and Infer the Component Spectra & Orbital Parameters from noisy, observed (composite) spectra time series

Orbital Parametric Model



Seven Parameters:

- Mass Ratio
- Velocity Amplitude
- eccentricity
- Arg of Periastron
- Epoch of Periastron
- Orbital Period
- Systemic Velocity

Nonparametric Bayes

Gaussian processes

We will model the latent stellar spectrum f_λ as a Gaussian process

$$f_\lambda \sim \mathrm{GP}(\mu(\lambda),\ k(\lambda, \lambda'))$$

A function is said to have a Gaussian process if for any collection of inputs the random vector ***f*** has a multivariate Gaussian distribution with mean ***mu*** and covariance matrix given by *k* evaluated over ***lambda***

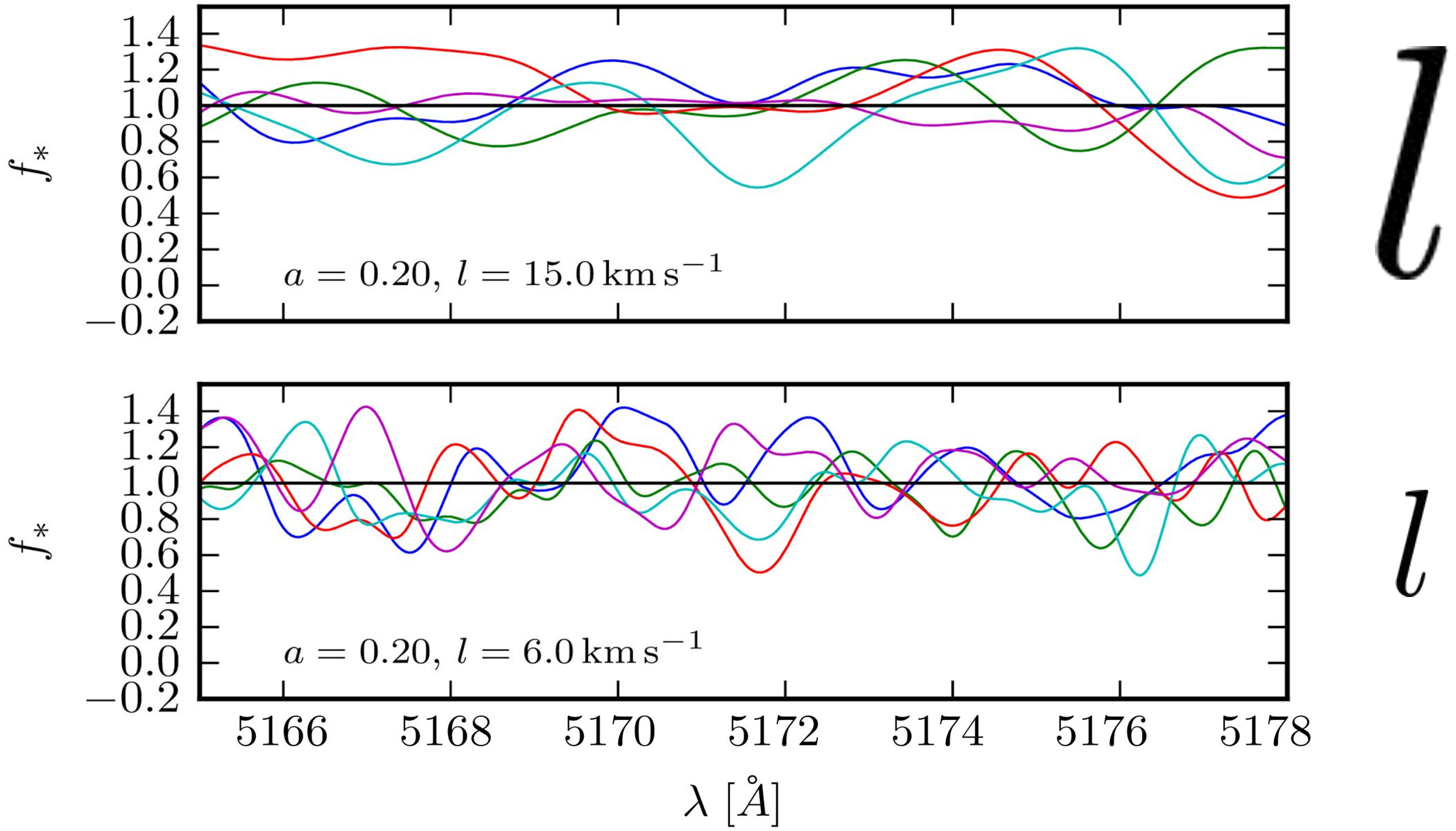
For a covariance kernel, we will use the commonly used squared exponential kernel, which relates pixels in the spectrum based upon their distance in log-wavelength (∝ velocity)

$$k_{ij}(r_{ij} | a, \boxed{l}) = a^2 \exp\left(-\frac{r_{ij}^2}{2l^2}\right)$$

Gaussian Process = a prior on functions (latent spectra)

Gaussian Process model for a single, stationary star

(Zoomed) draws from the prior



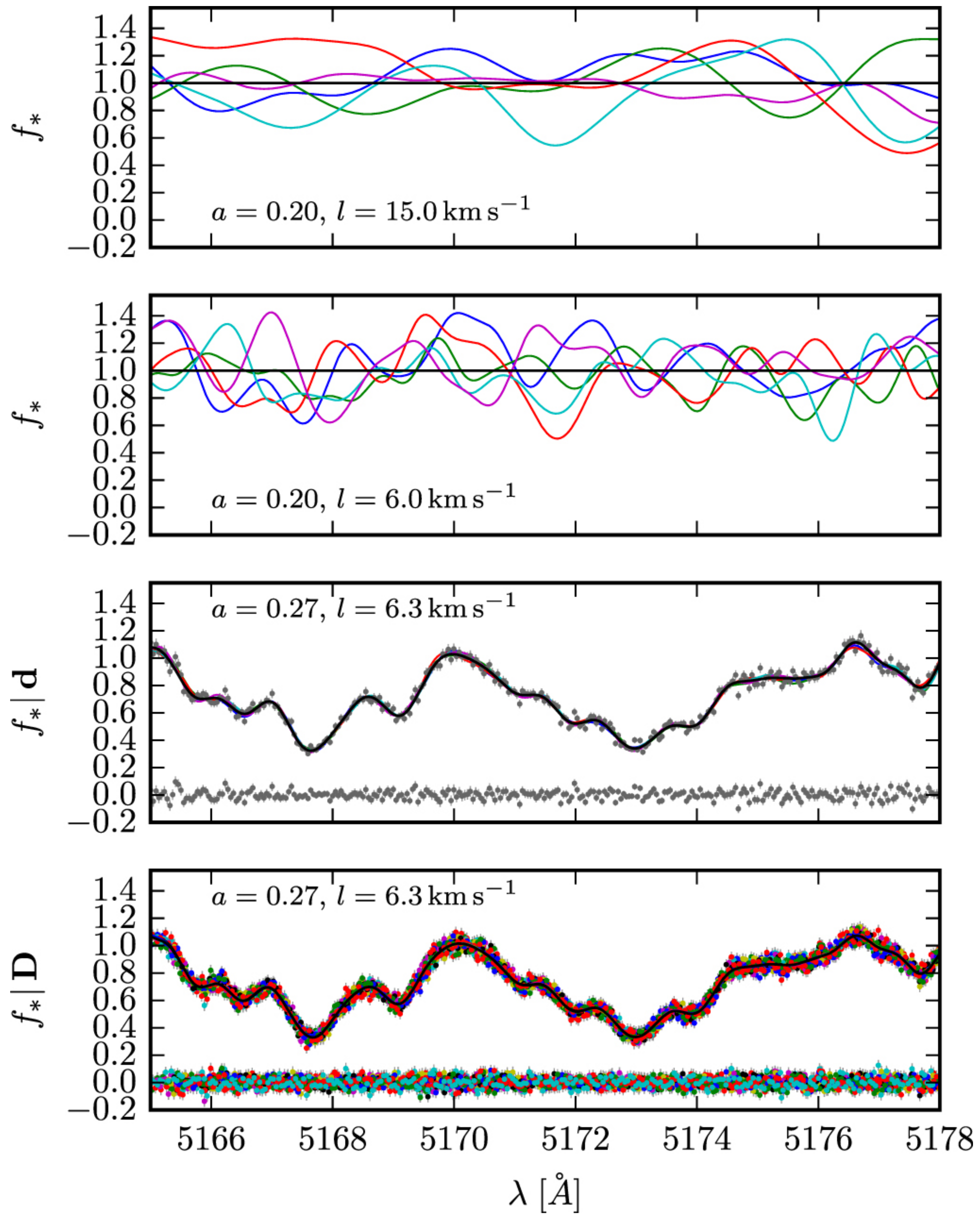
Inference = Which function is most consistent with the data?

Gaussian Process: Priors & Posteriors

GP prior
(long length scale)

(short length scale)

GP Posterior
(conditioned on
data spectrum d)
Inference of latent
spectrum



Known Unknowns

$$\Theta$$

7-dim Orbital Parameters = Period, Phase, eccentricity, Velocity Amplitude

$$f(\lambda), g(\lambda)$$

(∞-dim) Latent Functions = the unobserved component spectra of the primary (f) and secondary (g) stars

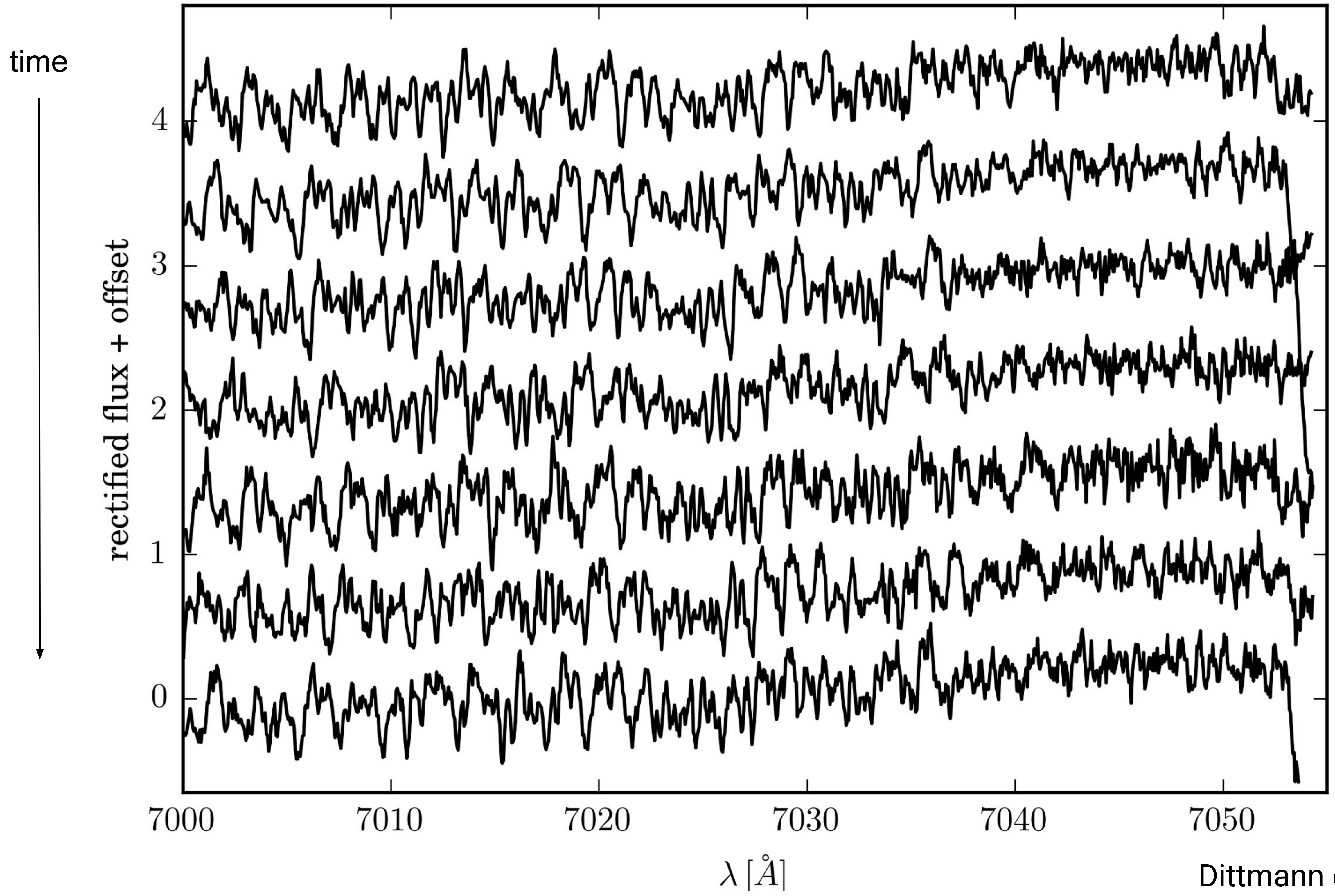
$$\boldsymbol{\alpha} = (a_f, l_f, a_g, l_g)$$

4-dim GP hyperparameters = controlling the amplitude and smoothness of Gaussian Process prior on latent spectra

Knowns (Data)

Raw Observations of the LP661-13 M4 Binary

$$\boldsymbol{D} =$$



Dittmann et al. 17
Czekala et al. 17a

Bayesian Inference

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In this case:

$$P(\boldsymbol{\Theta}, \boldsymbol{f}, \boldsymbol{g}, \boldsymbol{\alpha}|\boldsymbol{D}) \propto$$
$$P(\boldsymbol{D}|\boldsymbol{\Theta}, \boldsymbol{f}, \boldsymbol{g}, \boldsymbol{\alpha}) \times P(\boldsymbol{\Theta}, \boldsymbol{f}, \boldsymbol{g}, \boldsymbol{\alpha})$$

a probability density on (4+7+∞)-dim parameter space

Bayesian Computation

1. Run Markov Chain Monte Carlo (MCMC)
   (e.g. *emcee* affine-invariant ensemble sampler)
   on the 4+7 small dimensional marginal posterior

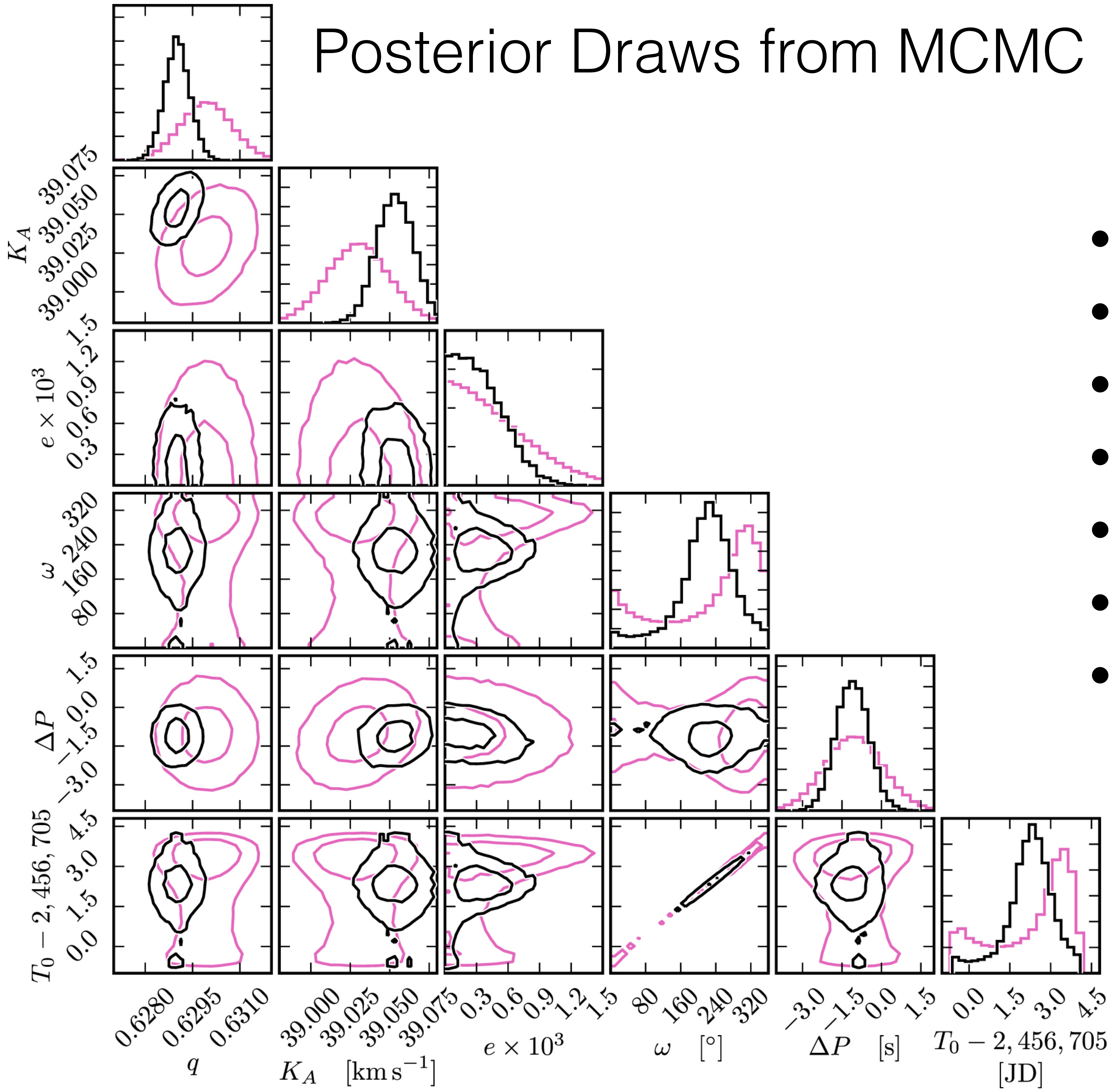
$$P(\boldsymbol{\Theta}, \boldsymbol{\alpha} | \boldsymbol{D}) = \int d\boldsymbol{f} \int d\boldsymbol{g}\, P(\boldsymbol{\Theta}, \boldsymbol{f}, \boldsymbol{g}, \boldsymbol{\alpha} | \boldsymbol{D})$$

MCMC generates samples: $\boldsymbol{\Theta}_i, \boldsymbol{\alpha}_i \sim P(\boldsymbol{\Theta}, \boldsymbol{\alpha} | \boldsymbol{D})$

2. Draw high-dim (**f**, **g**) spectra from the posterior predictive distribution

$$\boldsymbol{f}_i, \boldsymbol{g}_i \sim P(\boldsymbol{f}, \boldsymbol{g} | \boldsymbol{\Theta}_i, \boldsymbol{\alpha}_i, \boldsymbol{D})$$

Application to the Mid-M-Dwarf Binary LP661-13



Posterior Draws from MCMC

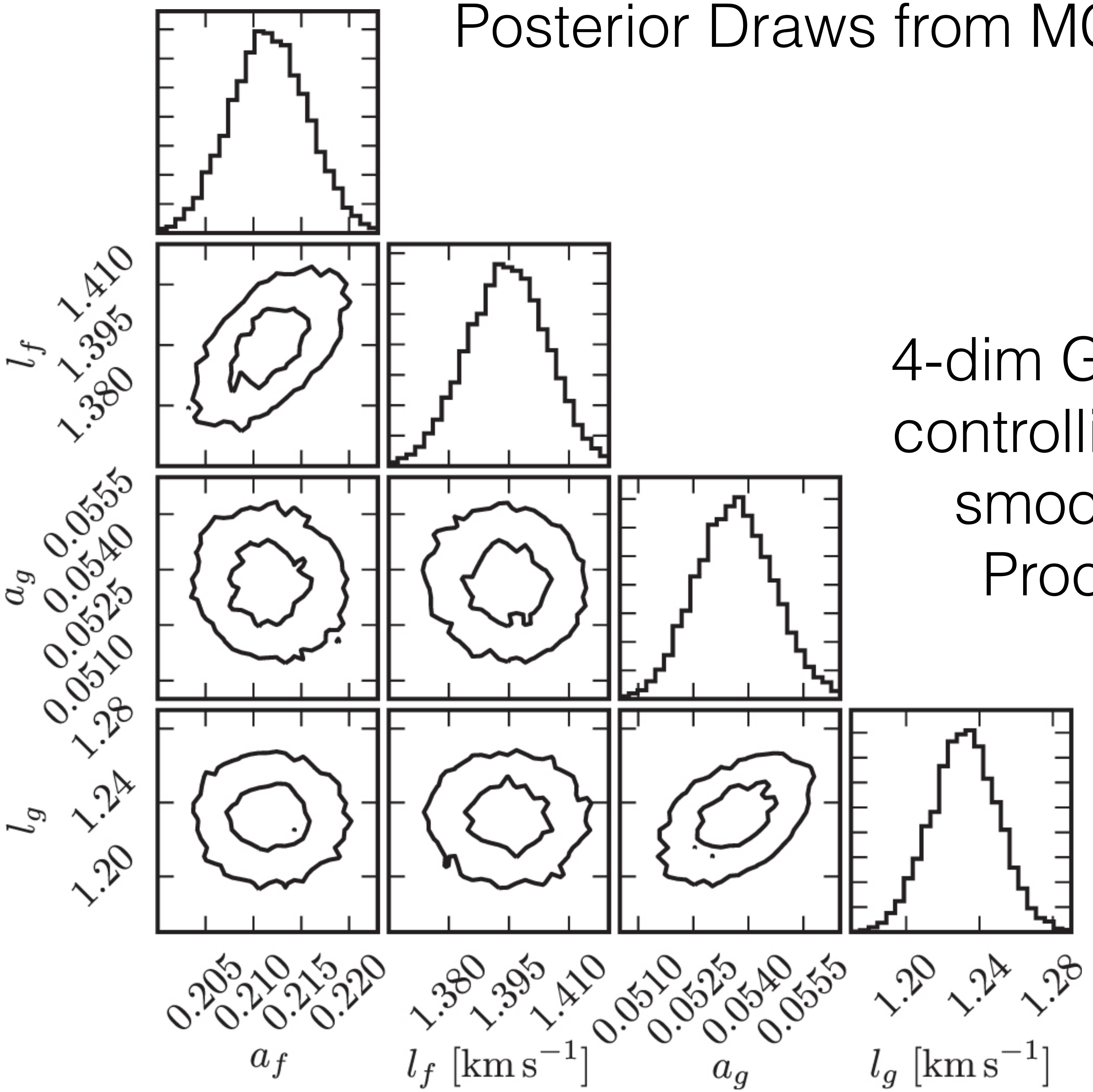
Seven Orbital Parameters:

- Mass Ratio
- Velocity Amplitude
- eccentricity
- Arg of Periastron
- Epoch of Periastron
- Orbital Period
- Systemic Velocity

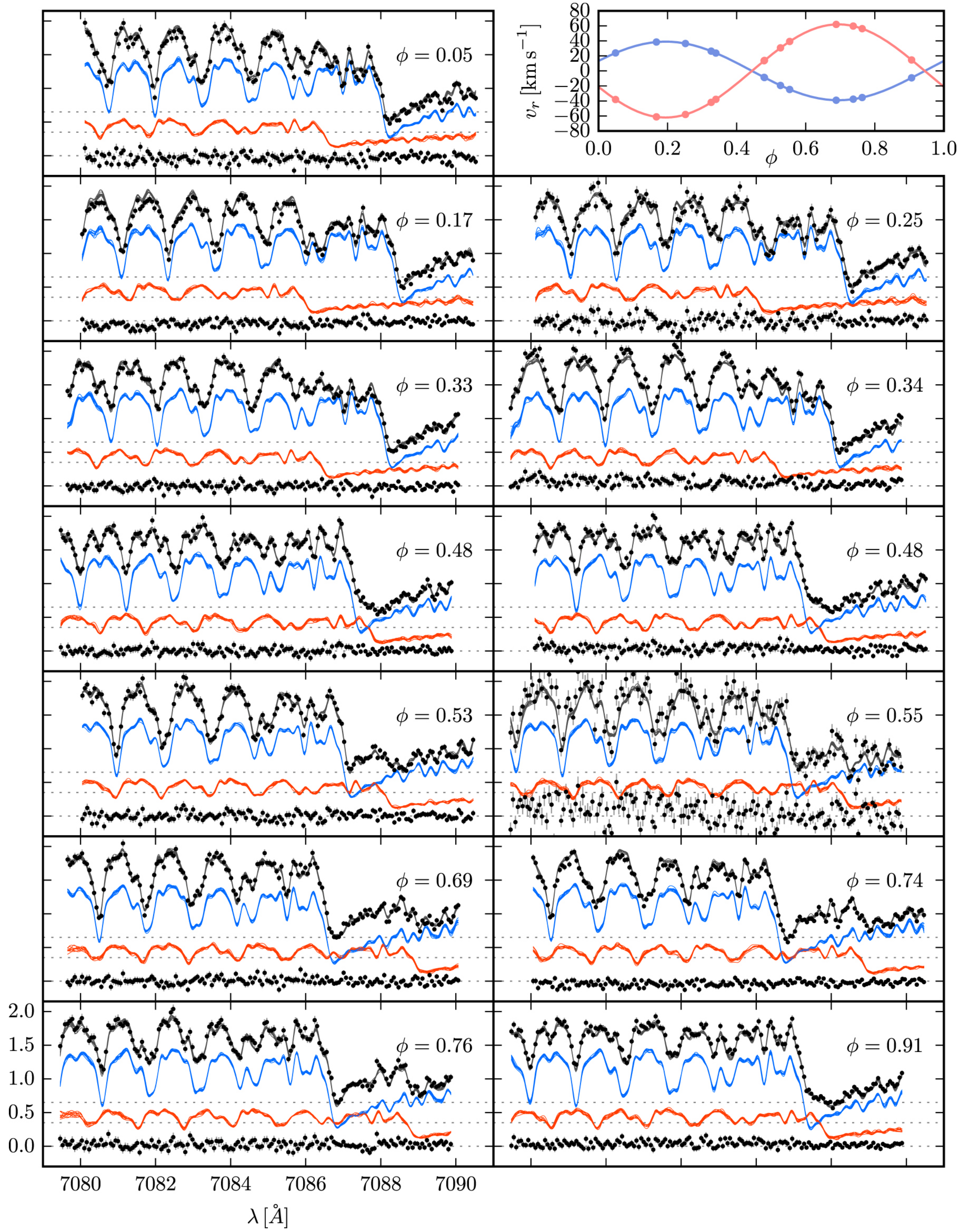
(Purple: Conventional Analysis)

Application to the Mid-M-Dwarf Binary LP661-13

Posterior Draws from MCMC $\boldsymbol{\alpha} = (a_f, l_f, a_g, l_g)$



4-dim GP hyperparameters = controlling the amplitude and smoothness of Gaussian Process prior on latent spectra



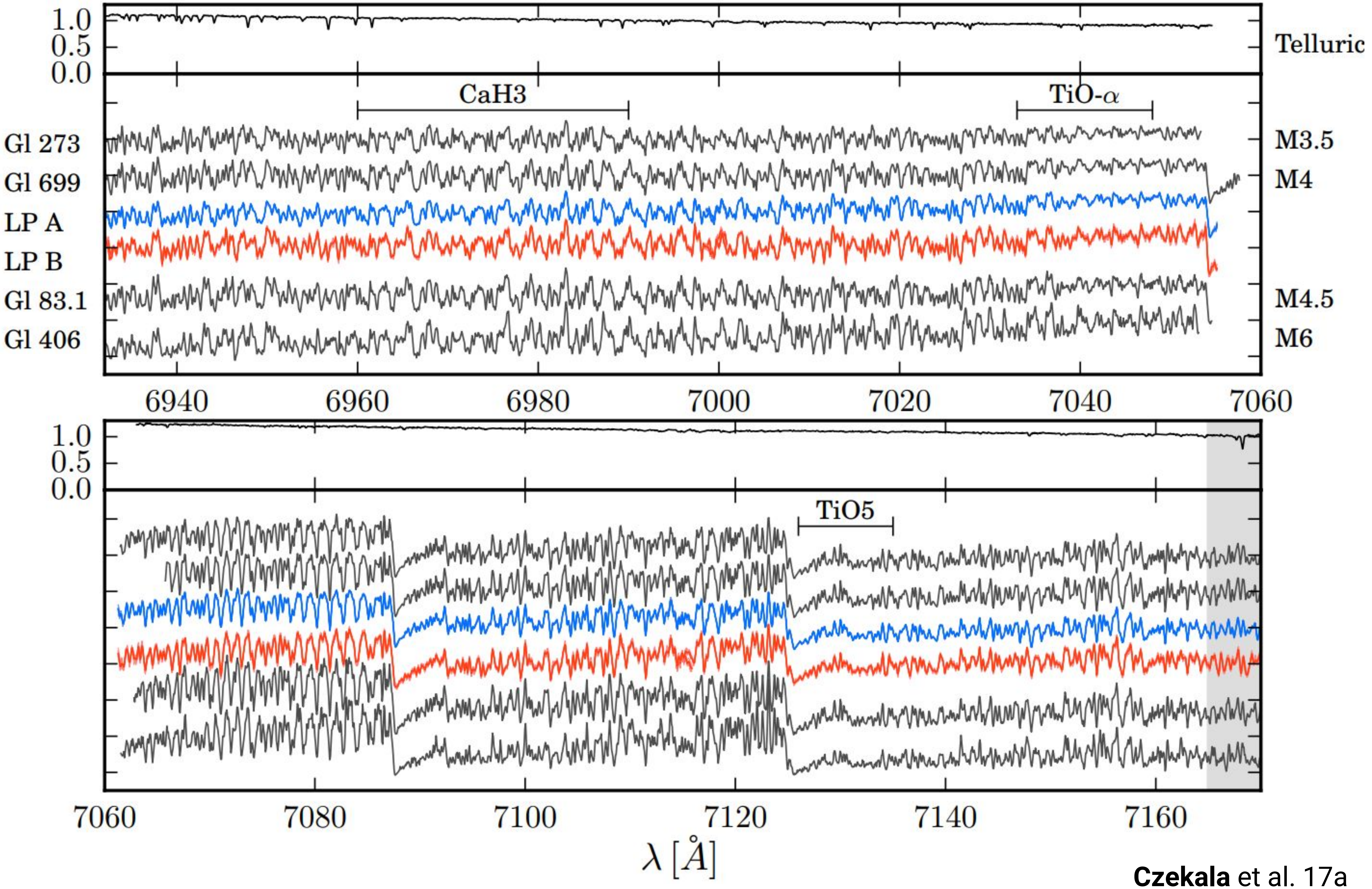
Posterior Inference of Component Spectra (**f**, **g**)

Compared to 10 epochs of observed spectra (**data**)

Model Checking! Checking Fit against Data

Model Checking!
Checking Fit against Domain Knowledge (astrophysics)!

Disentangled spectra match other single standard stars



Czekala et al. 17a

Astrostatistics Case Study: Disentangling Time Series Spectra with Gaussian Processes: Applications to Radial Velocity Analysis (Czekala et al. 2017, arXiv:1702.05652)

http://psoap.readthedocs.io/en/latest/

- Statistics:
  - Parametric Modelling (Stellar Orbit Parameters)
  - Nonparametric Modelling (Gaussian Process Spectrum)
  - Bayesian Inference (probability of unknowns given data)
  - Markov Chain Monte Carlo (computing posterior probability)
- Astronomy:
  - Applications to Radial Velocity Analysis of Stars/Exoplanets