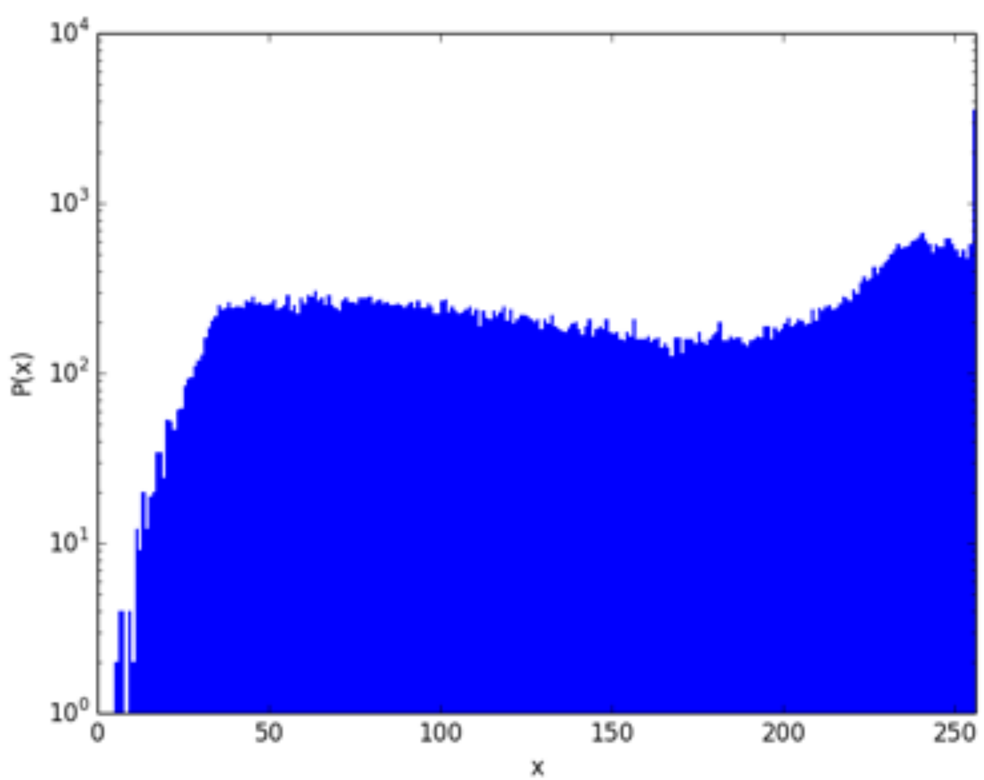
Statistics & Bayesian Inference Lecture 1

Joe Zuntz



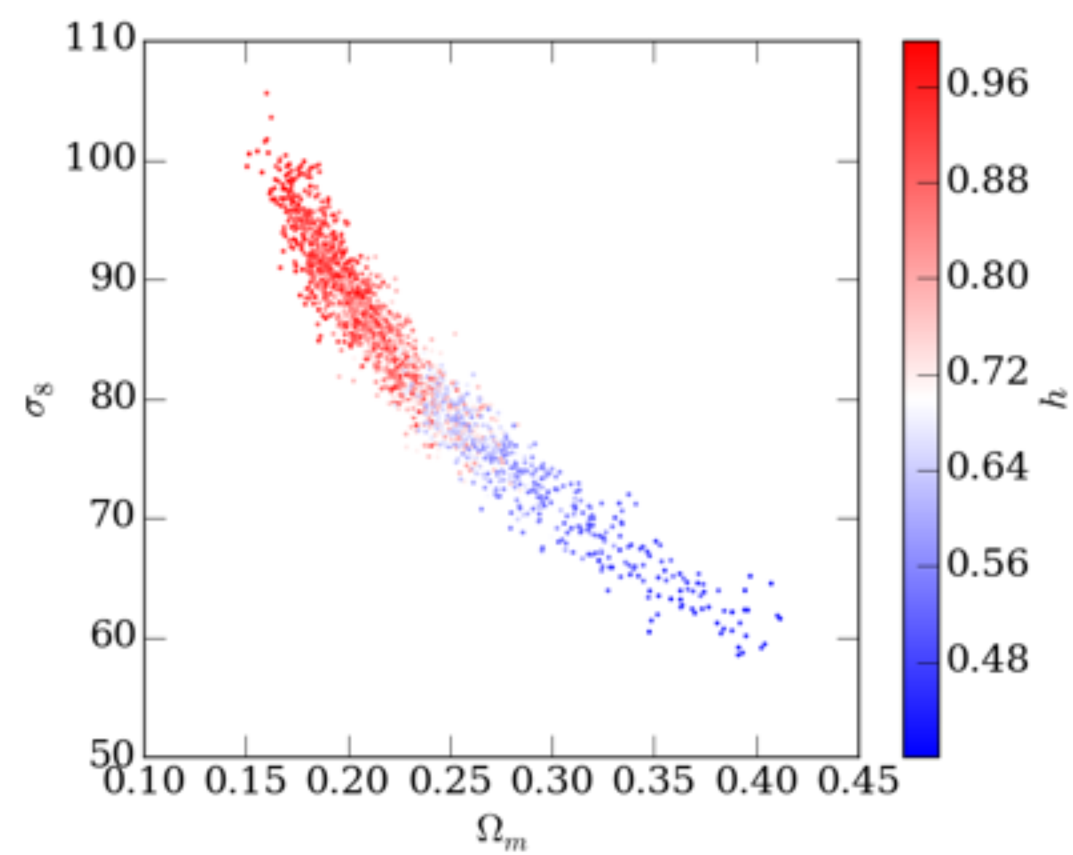
Lecture 1
Essentials of probability

- Motivations
- Definitions
- Probability Distributions
- Basic probability operations
- Some analytic distributions
- Bayes Theorem
- Models & Parameter Spaces
- How scientists can use probability

Motivations

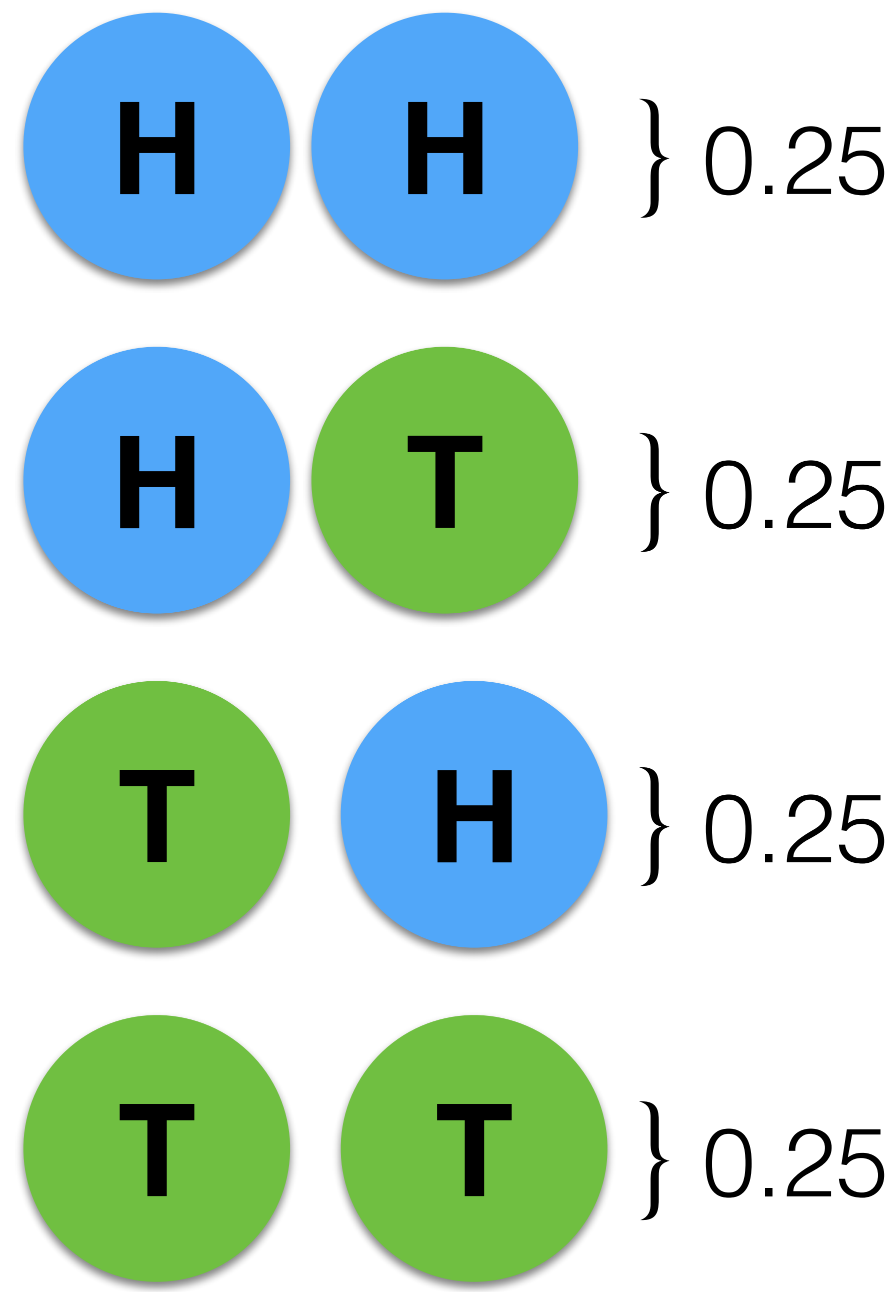
- Learn as much as possible from our (expensive) data
  - Constrain parameters in models
  - Test & compare models
- Characterize collections of numbers

$$H_0 = (72 \pm 8)\ \mathrm{km\,s^{-1}Mpc^{-1}}$$



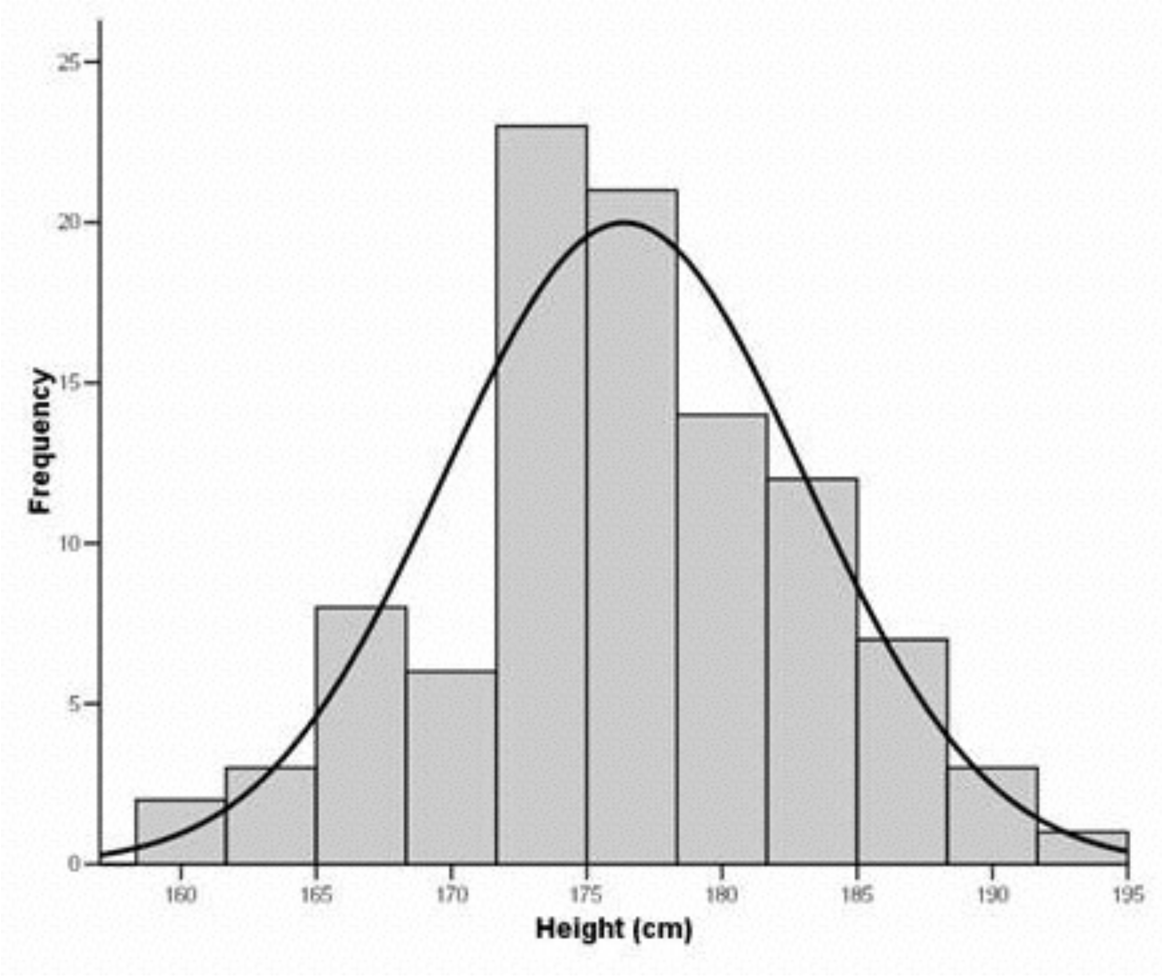
Probability Distributions: Definitions

- Assign real number $P \geq 0$ to each member of a *sample space* (discrete or continuous, finite or infinite)
- P=probability density function (PDF) or probability mass function (PMF)
- This set represents possible outcomes of an experiment/game/event/situation
- e.g. possible results tossing two coins, height of next person to walk through door



Probability Distributions: Definitions

- Assign real number $P \geq 0$ to each member of a *sample space* (discrete or continuous, finite or infinite)
- P=probability density function (PDF) or probability mass function (PMF)
- This set represents possible outcomes of an experiment/game/event/situation
- e.g. possible results tossing two coins, height of next person to walk through door



Probability Distributions: Definitions

- A random variable X is any value subject to randomness, e.g.:
    - was first toss heads?
      was the sequence Heads-Tails?
      were both tosses the same?
- Discrete X: P is a list of values
- Continuous X: P is a function, PDF, (which we have to integrate to answer questions)

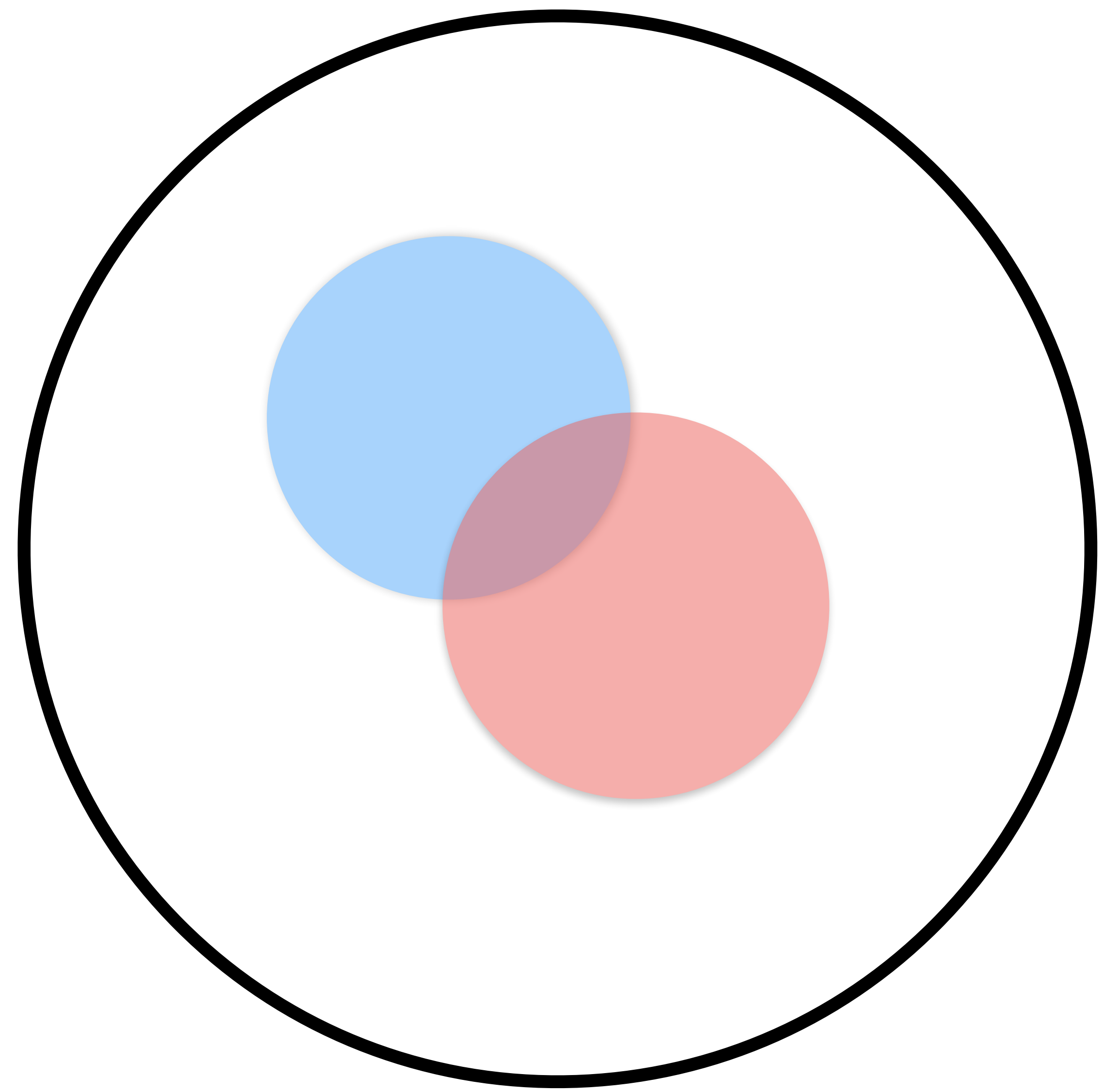
Probability Distributions: Basic properties

- Since X must have exactly one value:
  - Discrete: $$\sum_{x \in X} P(x) = 1$$
  - Continuous: $$\int_{x \in X} P(x) \mathrm{d}x = 1$$
- P(X=x) = f(x)
  Usually just write P(X) = f(x)
- 0 ≤ P(x) ≤ 1

Probability Distributions: Combining Probabilities

- Joint probability
  $P(XY)$
  $P(X=x \text{ and } Y=y)$
  $P(X \cap Y)$

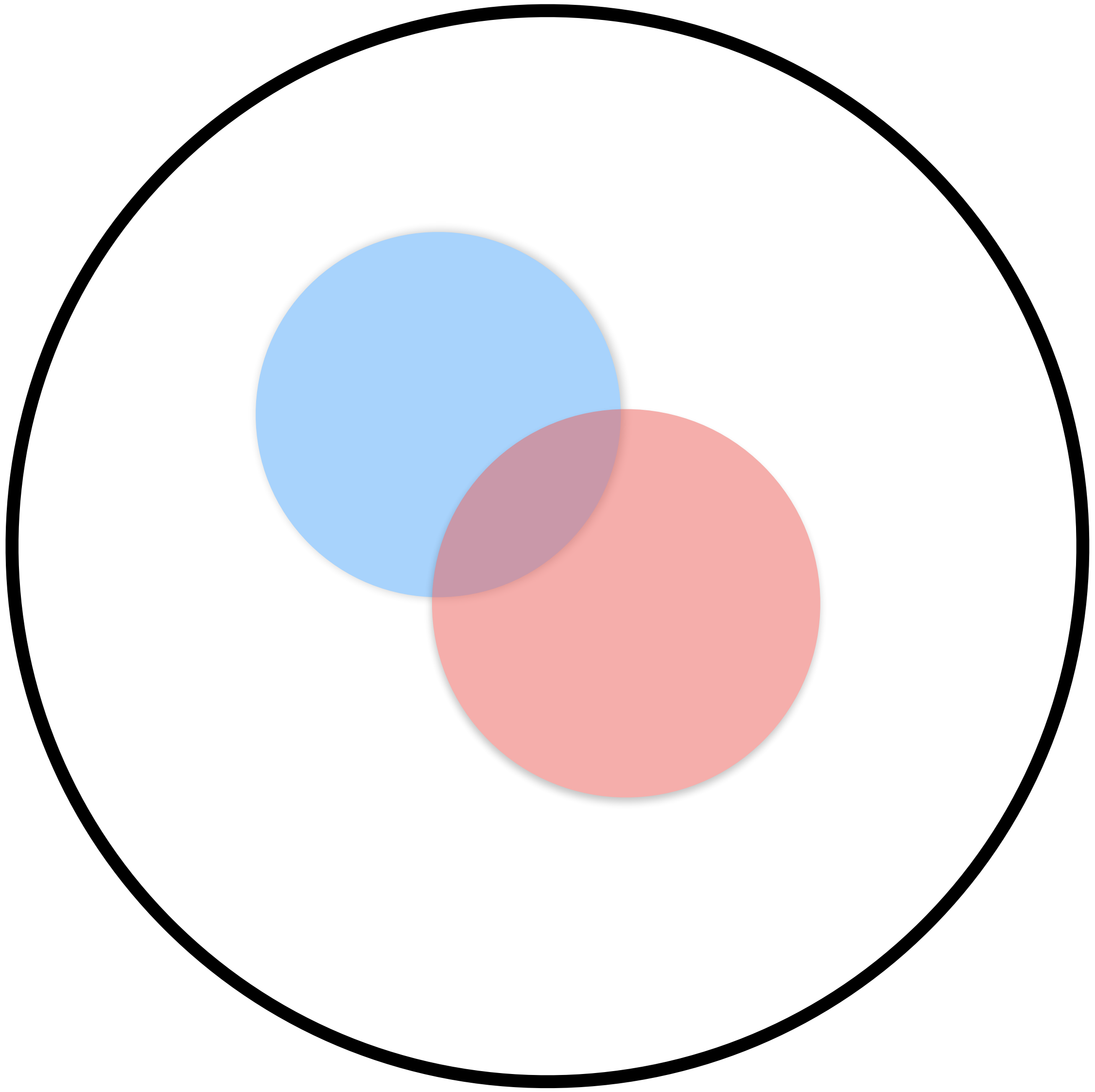
- Union
  $P(X=x \text{ or } Y=y)$
  $P(X \cup Y)$



Probability Distributions: Combining Probabilities

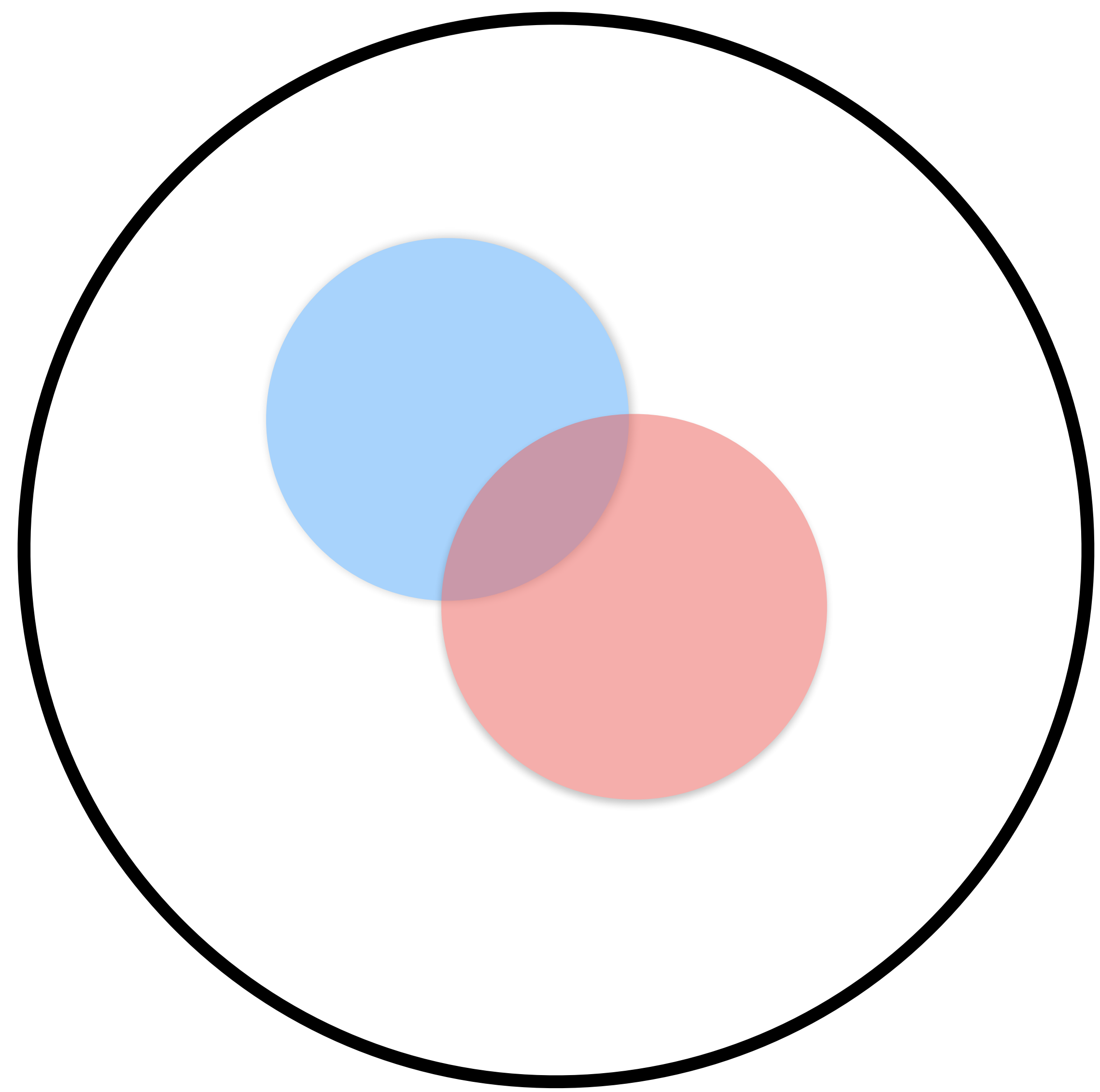
- Conditional
  P(X=x given Y=y)
  P(X|Y)

- Independence:
  - P(X|Y) = P(X)
  - X independent of Y



Probability Distributions: Identities

- $P(\text{not } X) = 1 - P(X)$
- $P(XY) = P(X|Y)\ P(Y)$
- $P(XY) = P(X) + P(Y) - P(X \cap Y)$



Probability Distributions: Expectations

- The expectation (or mean) of a random variable X is given by:

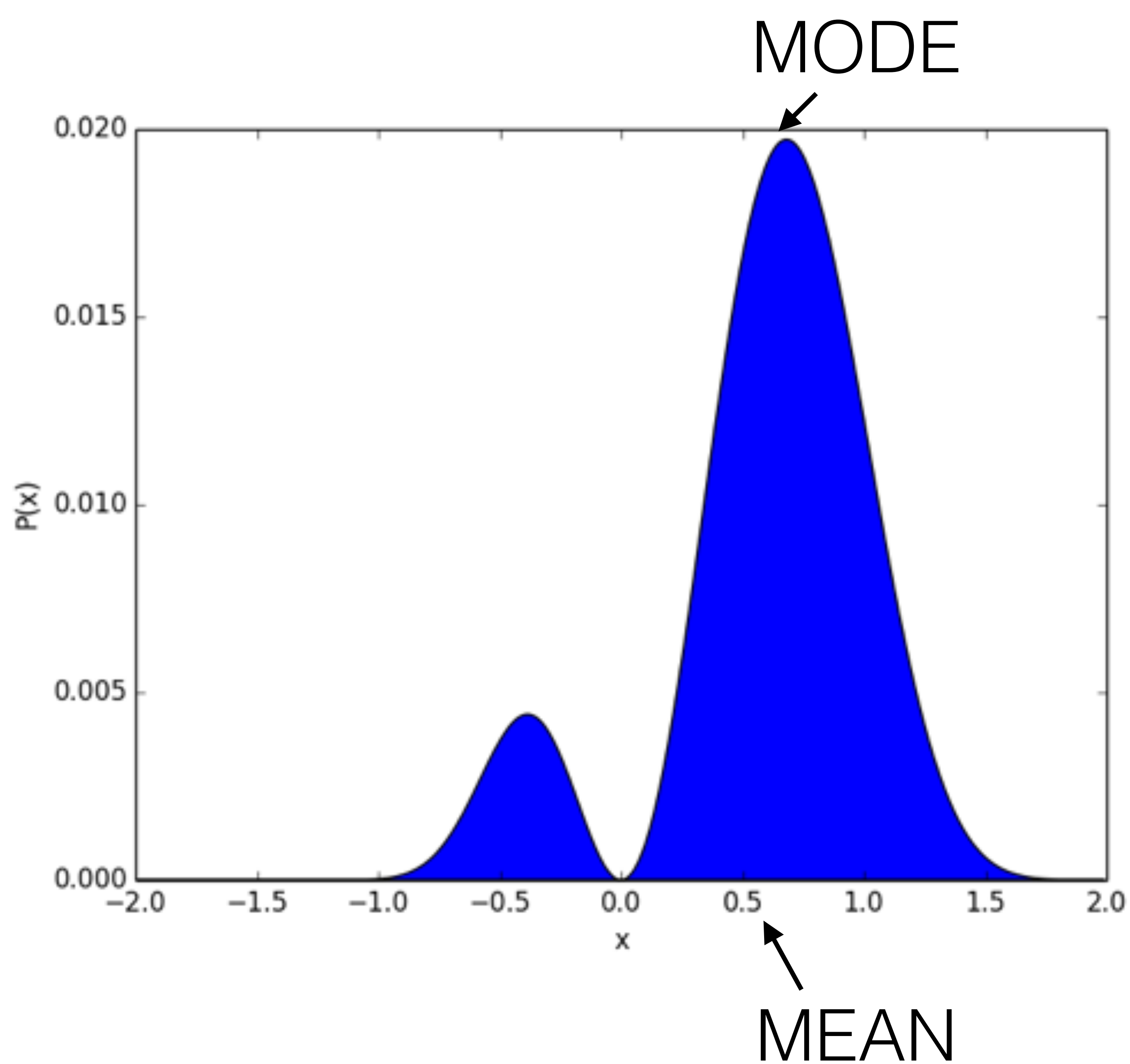
$$E(X) = \sum P(X)X \qquad E(X) = \int P(X)X\mathrm{d}X$$

- Or a function of it by:

$$E(f(X)) = \sum P(X)f(X) \qquad E(f(X)) = \int P(X)f(X)\mathrm{d}X$$

Probability Distributions: Expectations

- Expectations are one measure if centrality, and not always a good one.
- Mode and median also exist
- All just ways of reducing or characterizing a distribution



Probability Distributions: Marginalizing

- Discrete:

$$P(x) = \sum_i P(x|y_i)P(y_i)$$

- Continuous:

$$P(x) = \int P(x|y)P(y)\mathrm{d}y$$

- If you don't care about something, marginalize over it

Probability Distributions: Changing variables

- Probability mass must be conserved, not density
- Relate with a Jacobian
- Be especially careful in more dimensions

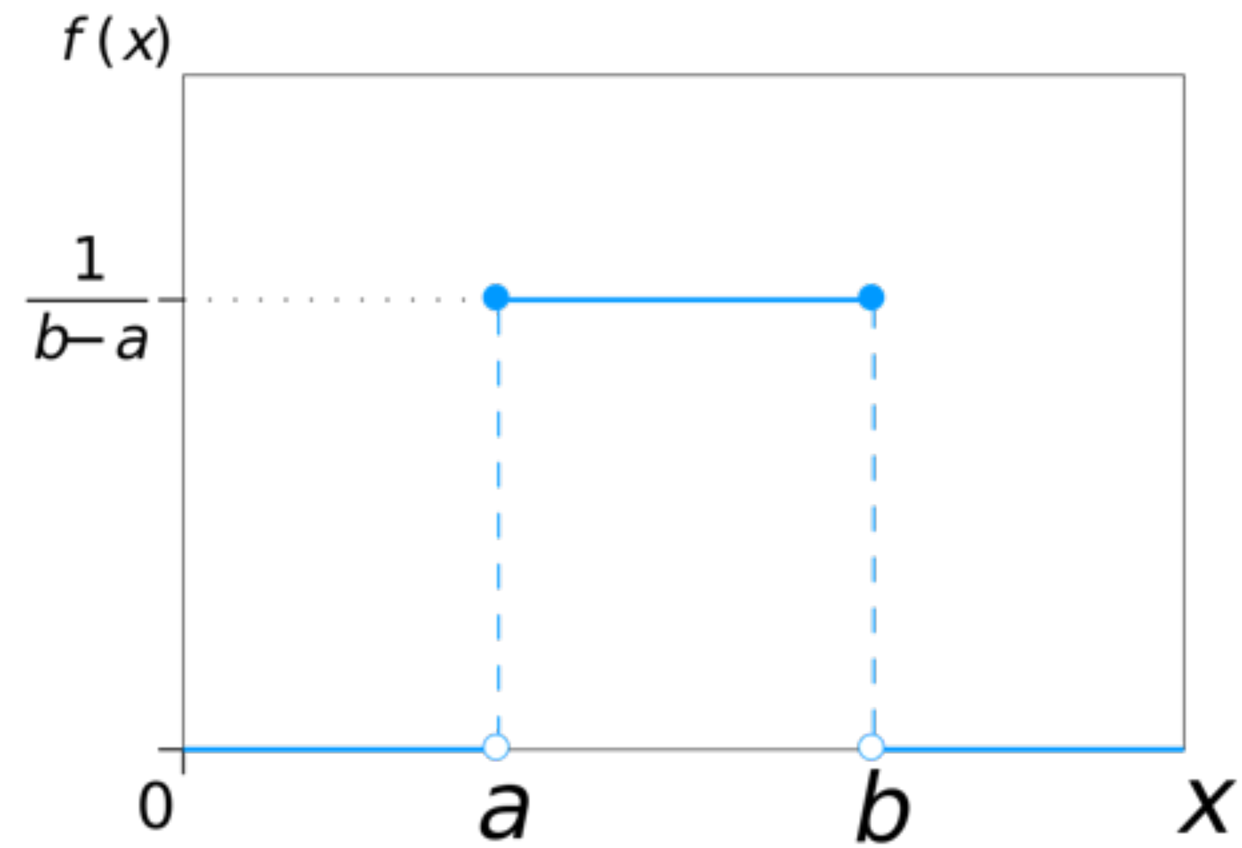
$$
\begin{aligned}
u &= f(x) \\
P(u)\mathrm{d}u &= P(x)\mathrm{d}x \\
P(u) &= P(x)\frac{\mathrm{d}x}{\mathrm{d}u} \\
&= P(x)/\frac{\mathrm{d}u}{\mathrm{d}x} \\
&= P(x)/f'(x)
\end{aligned}
$$

Probability Distributions: Drawing samples

- Generate values of X with probability specified by P(X)
- Draw enough samples: histogram looks like PDF
- See lecture 3

Probability Distributions: Analytic examples

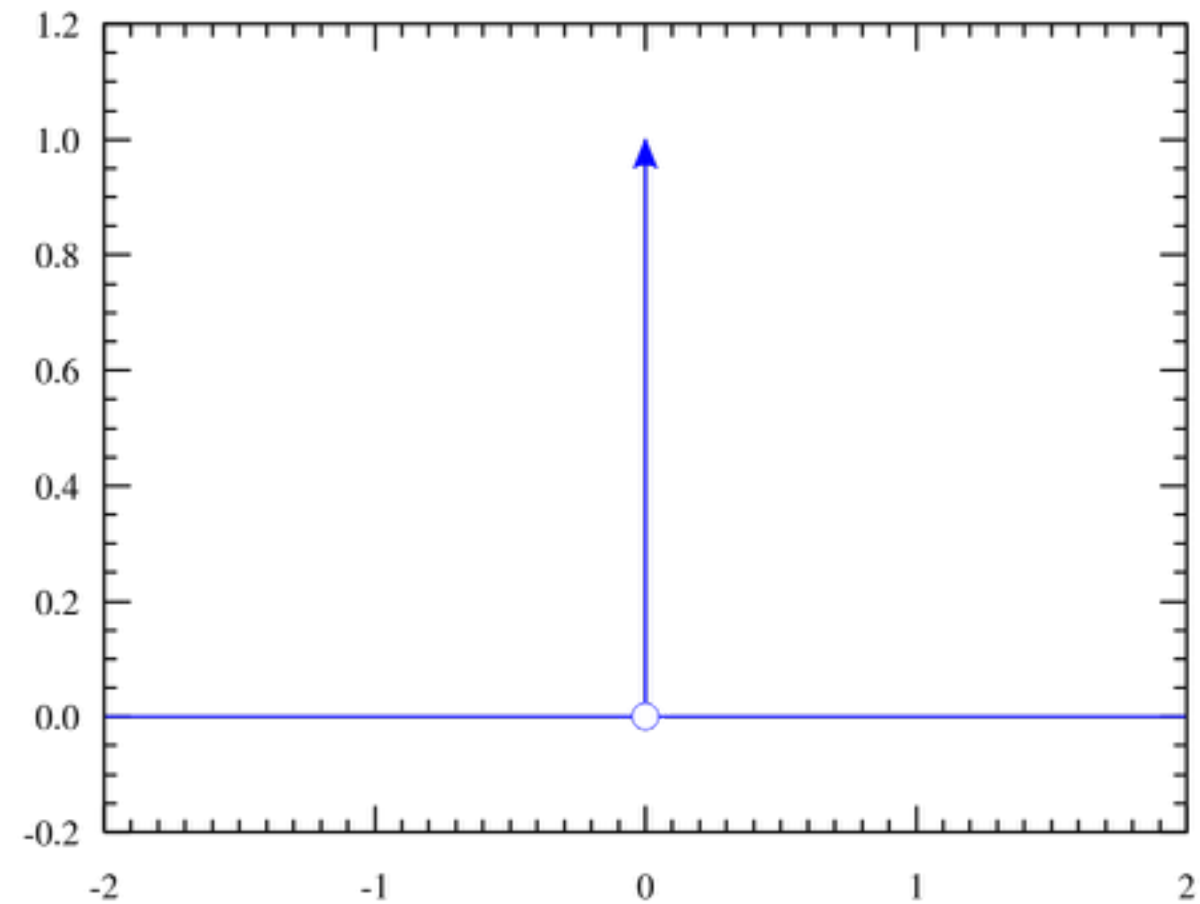
- Wikipedia is brilliant for this
- **Uniform**
- Delta function
- Gaussian (normal)
- Exponential
- Poisson



$$P(x) = \frac{1}{b-a}, \; x \in [a, b]$$

Probability Distributions: Analytic examples

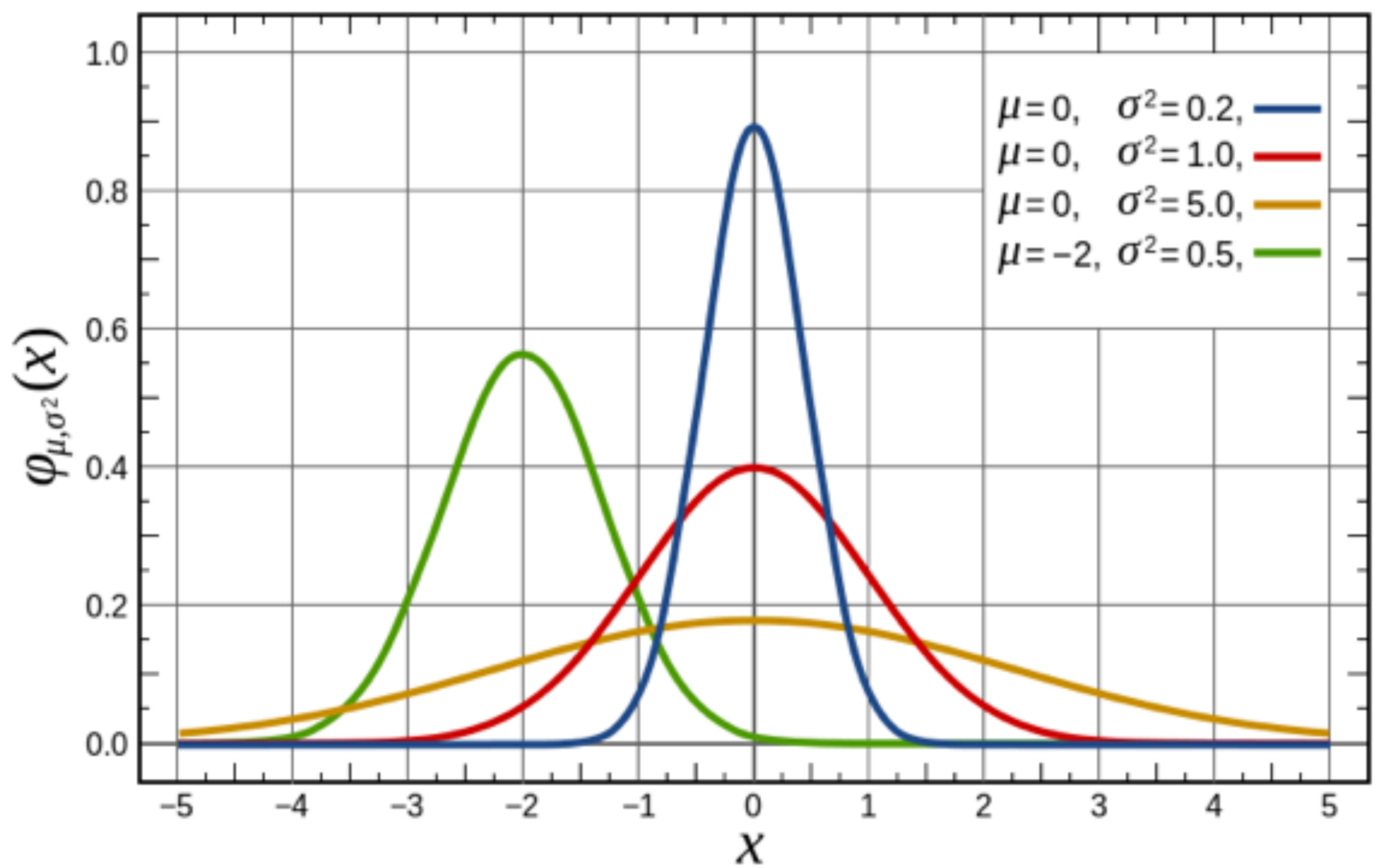
- Wikipedia is brilliant for this
- Uniform
- **Delta function**
- Gaussian (normal)
- Exponential
- Poisson



$$P(x) = \delta(x - x_0)$$

Probability Distributions: Analytic examples

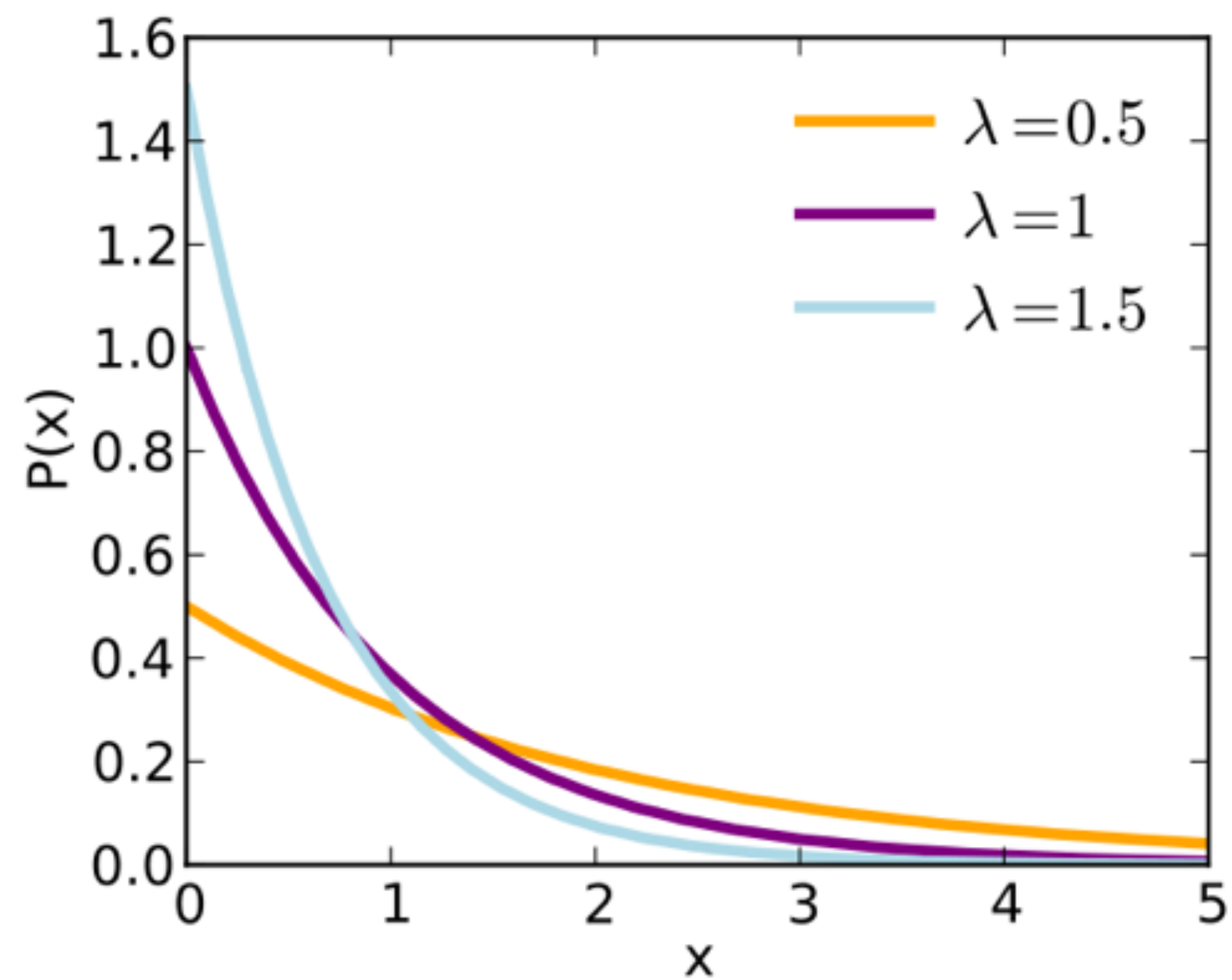
- Wikipedia is brilliant for this
- Uniform
- Delta function
- **Gaussian (normal)**
- Exponential
- Poisson



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

Probability Distributions: Analytic examples

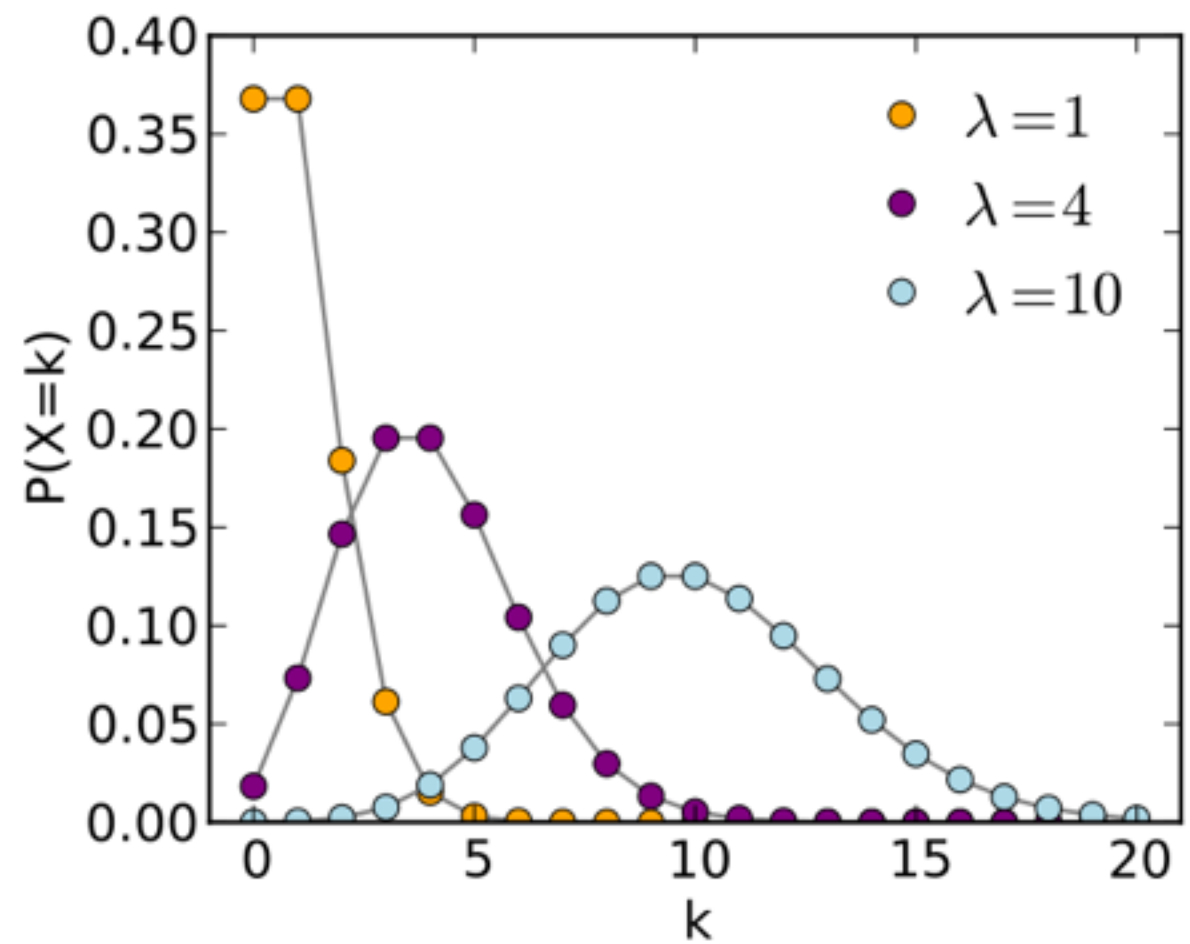
- Wikipedia is brilliant for this
- Uniform
- Delta function
- Gaussian (normal)
- **Exponential**
- Poisson



$$P(x) = \lambda e^{-\lambda x},\ x > 0$$

Probability Distributions: Analytic examples

- Wikipedia is brilliant for this
- Uniform
- Delta function
- Gaussian (normal)
- Exponential
- **Poisson**



$$P(n) = \frac{\lambda^n e^{-\lambda}}{n!}$$

Bayes Theorem and Inference

$$
\begin{aligned}
P(AB) &= P(A|B)P(B) \\
&= P(B|A)P(A)
\end{aligned}
$$

Bayes Theorem and Inference

$$\begin{aligned} P(AB) &= P(A|B)P(B) \\ &= P(B|A)P(A) \end{aligned}$$

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

Bayes Theorem and Inference

$$P(p|dM) = \frac{P(d|pM)P(p|M)}{P(d|M)}$$

Likelihood

Prior

$$\propto P(d|pM)P(p|M)$$

Model

Observed data

Parameters

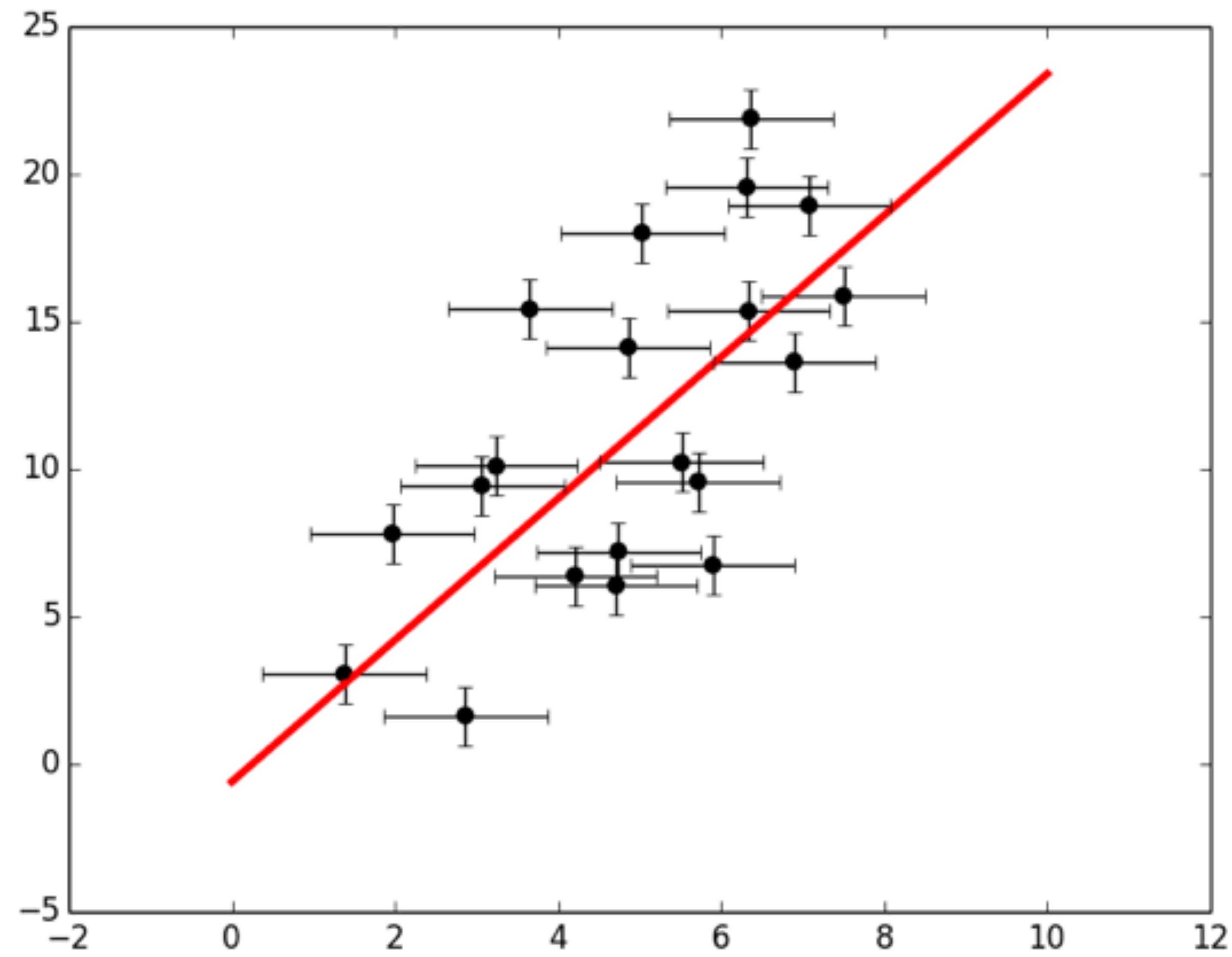
Bayes Theorem and Inference

What you know after looking at the data =

what you knew before
+ what the data told you

Models & Parameters

- A model is the mathematical theory that describes how your data arose.
    - It is **not** a theory of how what you **wanted** to measure arose.
- Non-trivial models include some deterministic and some stochastic parts.
    - Noise is one stochastic; many (most?) astrophysical models also have others too

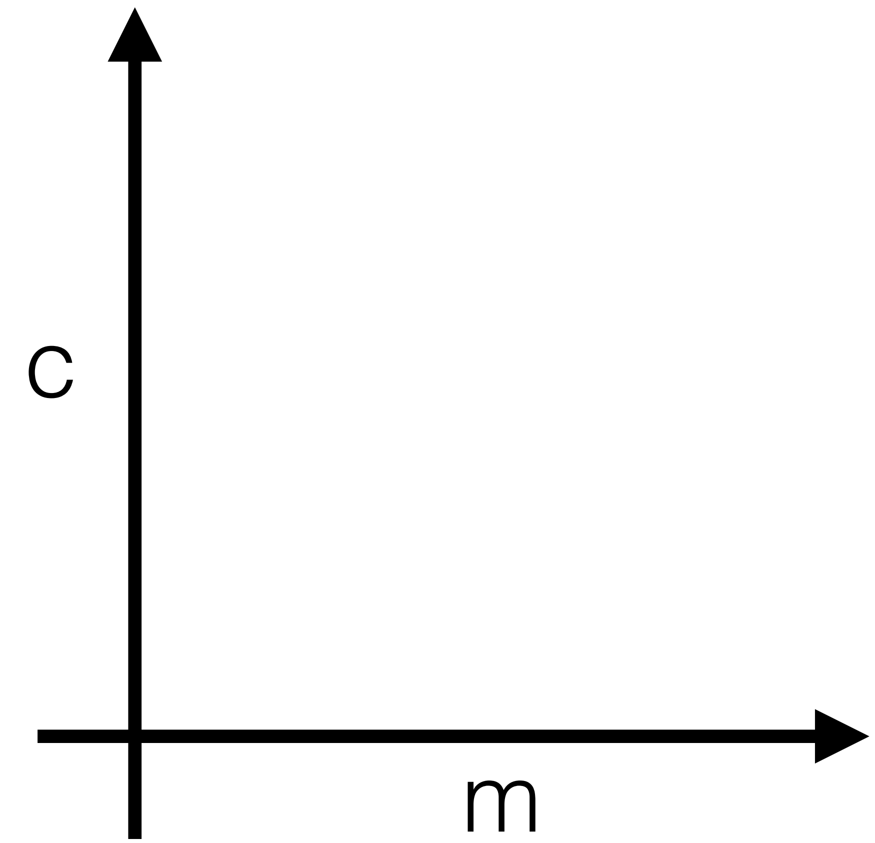


Models & Parameters

- Parameters are any unknown numerical values in your model
    - A parameter can have probability distributions
- You need (and have) some prior (background) information about all your parameters
    - This may be subjective!

Parameter Spaces

- Can use continuous parameters as dimensions in an abstract space
- Probabilities become functions of many variables: P(uvwxyz)
- As the dimension of this space increases your intuition becomes worse



Descriptive Statistics

- Reduce samples or distribution to set of characteristic numbers
    - In a analytic cases this is all you need to describe a distribution
- Statistics of samples
  = estimators/approximations to underlying distribution stats

Descriptive Statistics: Mean

- Distribution mean

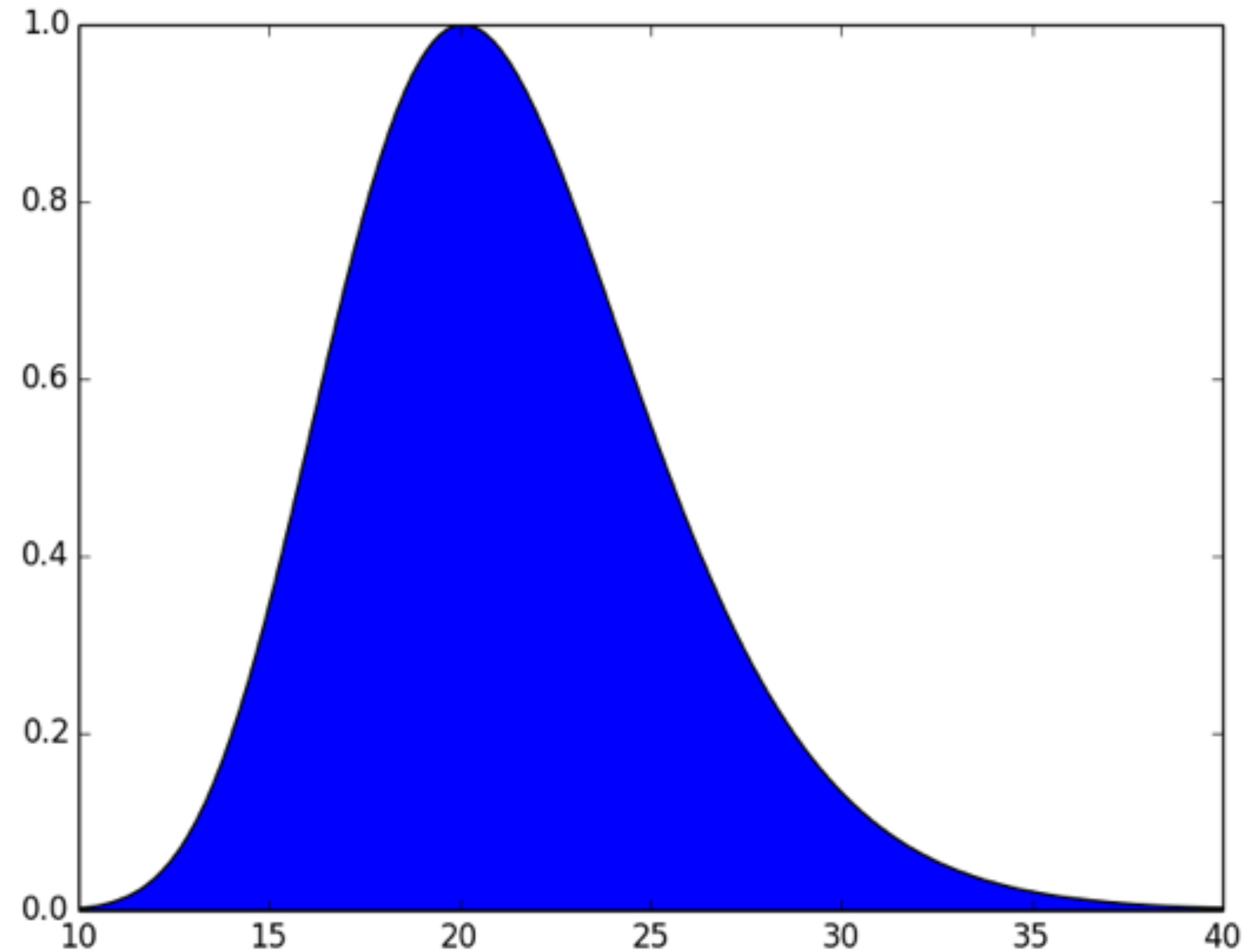
$$E[X] = \int X P(X) \mathrm{d}X$$

- Sample mean

$$\bar{X} = \frac{\sum X_i}{N}$$

Descriptive Statistics: Mean

- Means can be misleading!
- Most distributions are asymmetric



Descriptive Statistics: Variance

- Distribution variance

$$\mathrm{Var}(X) = E[(X - \bar{X})^2]$$
$$= \int (X - \bar{X})^2 P(X)\mathrm{d}X$$

- Sample variance

$$\sigma_X^2 = \frac{\sum (X_i - \bar{X})^2}{N}$$

- Population variance

$$s_X^2 = \frac{\sum (X_i - \bar{X})^2}{N - 1}$$

Descriptive Statistics: Covariance

$$\mathrm{Cov}(X, Y) = E[(X - \bar{X})(Y - \bar{Y})]$$

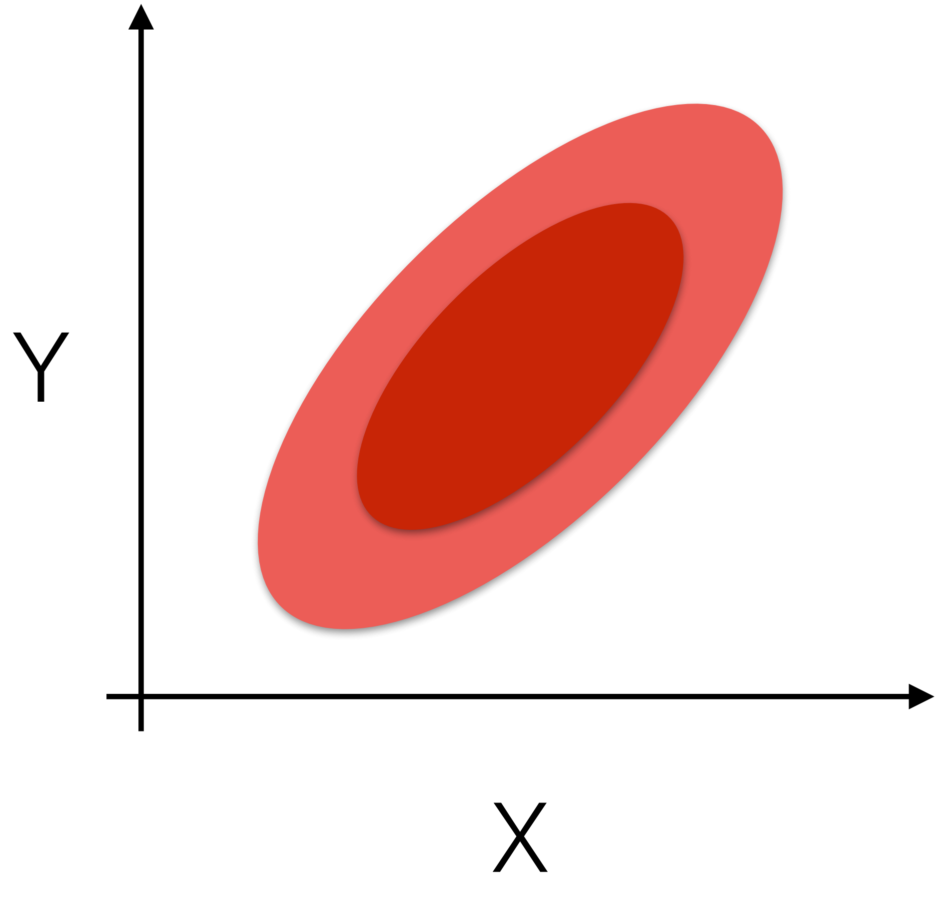
$$= \int (X - \bar{X})(Y - \bar{Y})P(XY)\mathrm{d}X\mathrm{d}Y$$

- Covariance

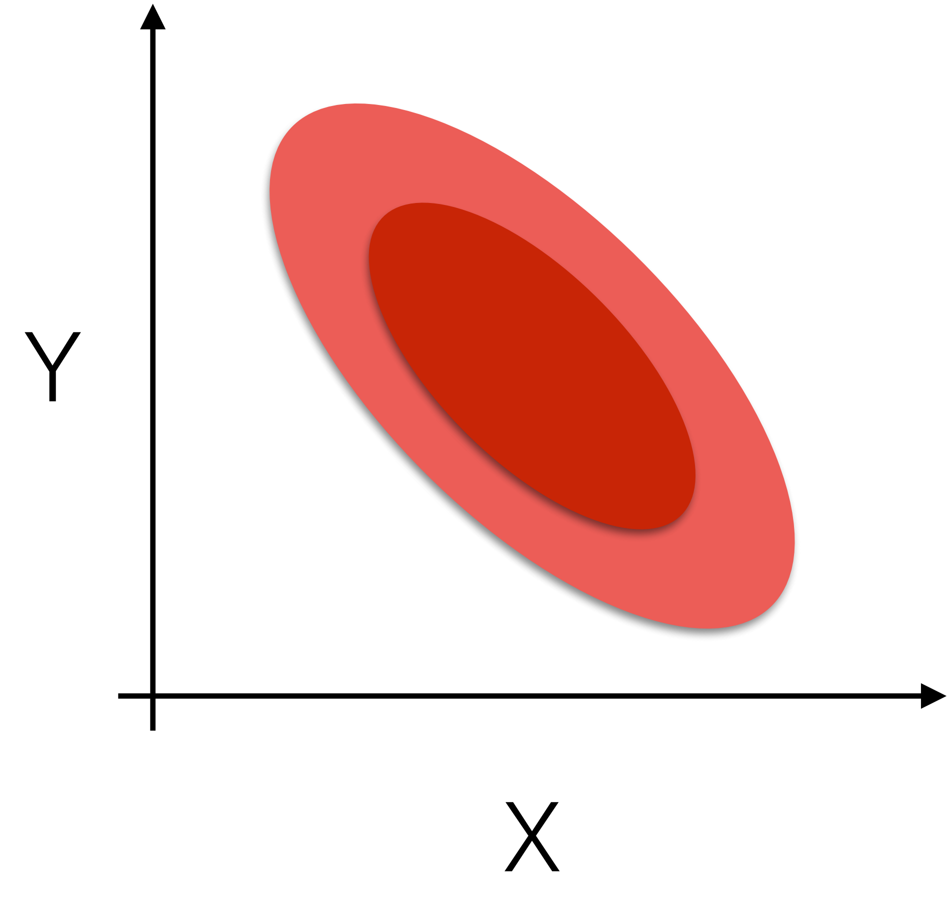
$$\sigma_{XY} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{N}$$

Descriptive Statistics: Covariance

$\sigma_{XY} > 0$



$\sigma_{XY} < 0$



Gaussians: The Basics

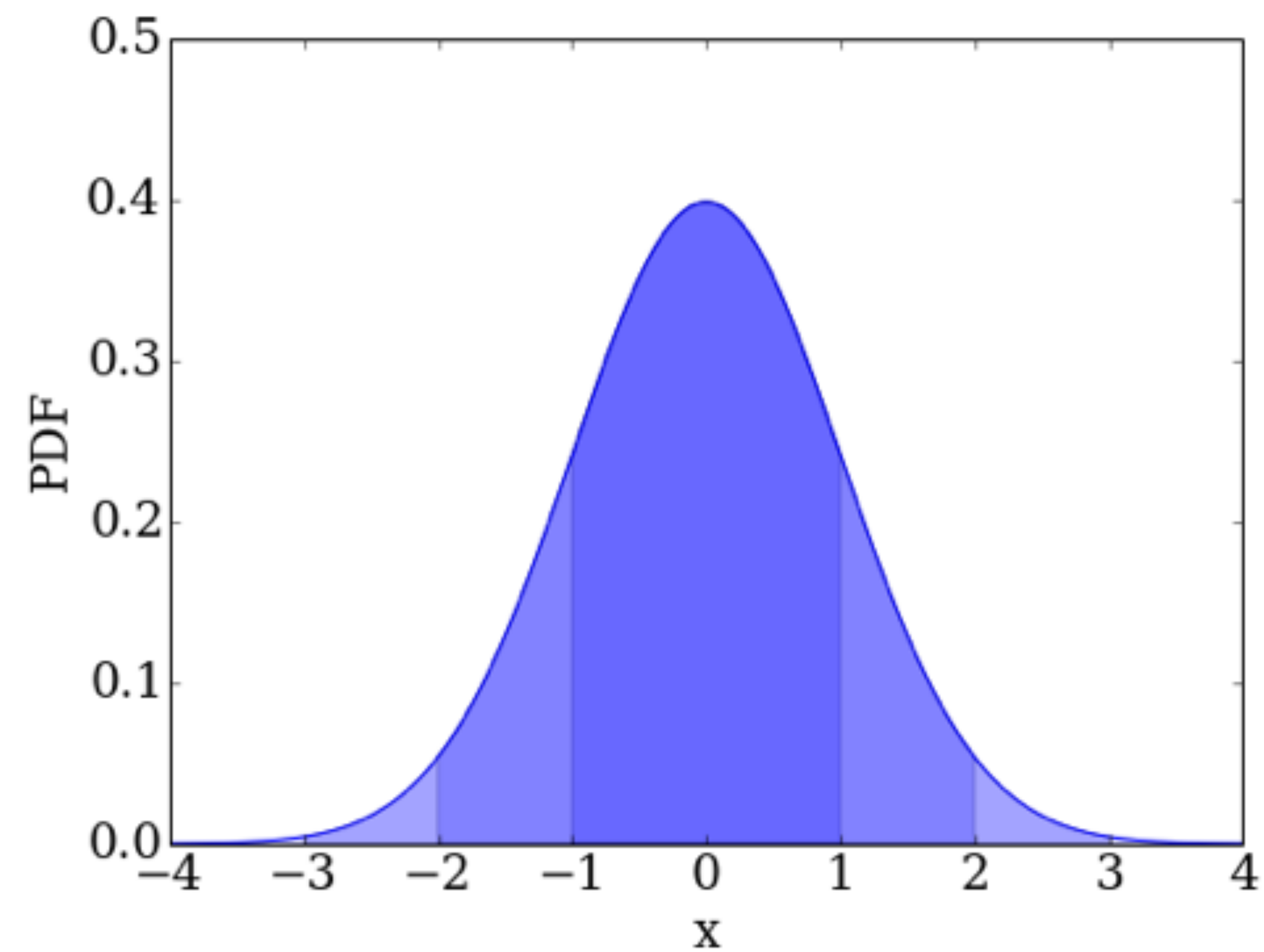
- One dimensional continuous PDF

$$P(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

- Two parameters:
  Mean μ
  Standard deviation σ

- Symmetric

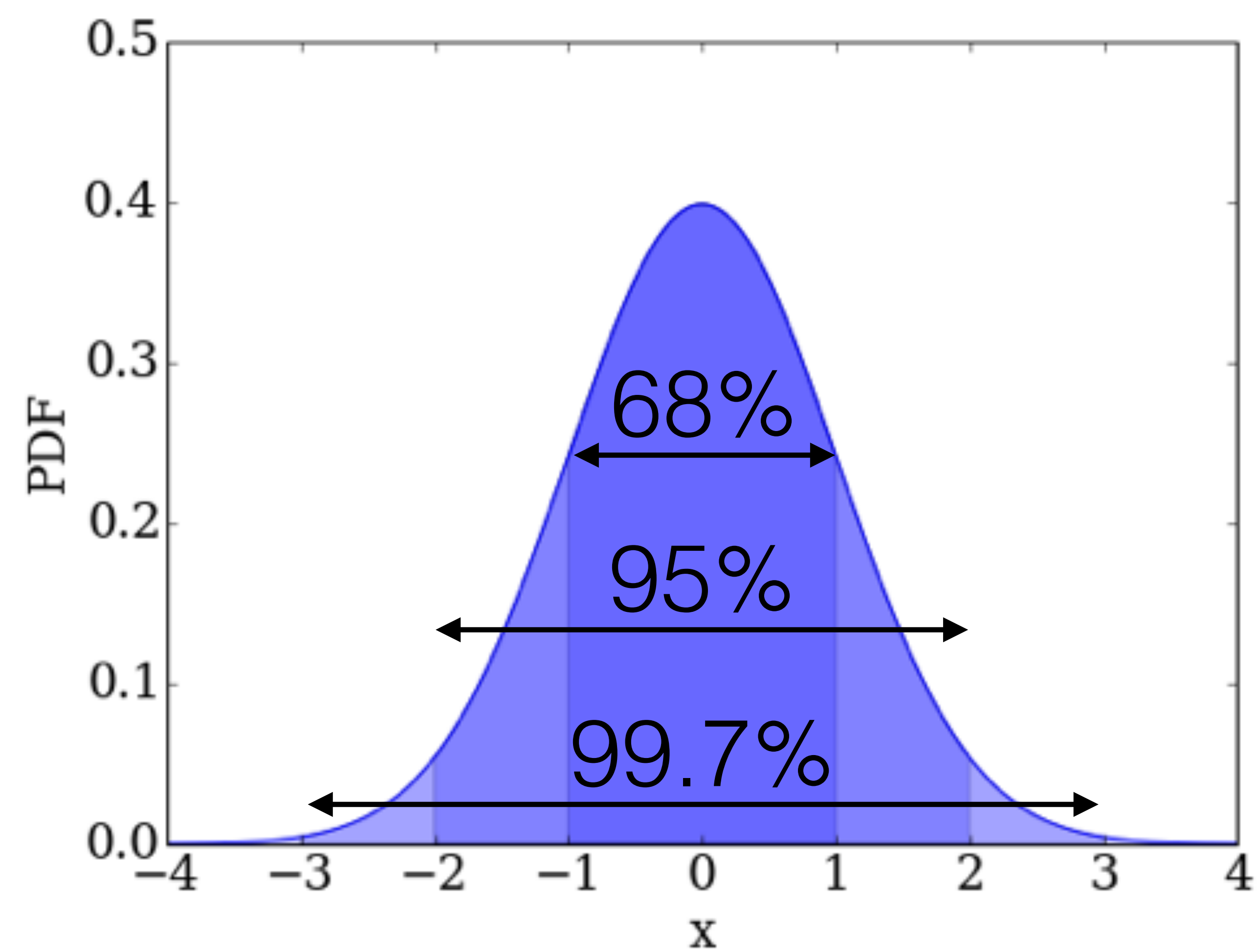
- Common! But often an over-simplification.



Gaussians: Sigma numbers

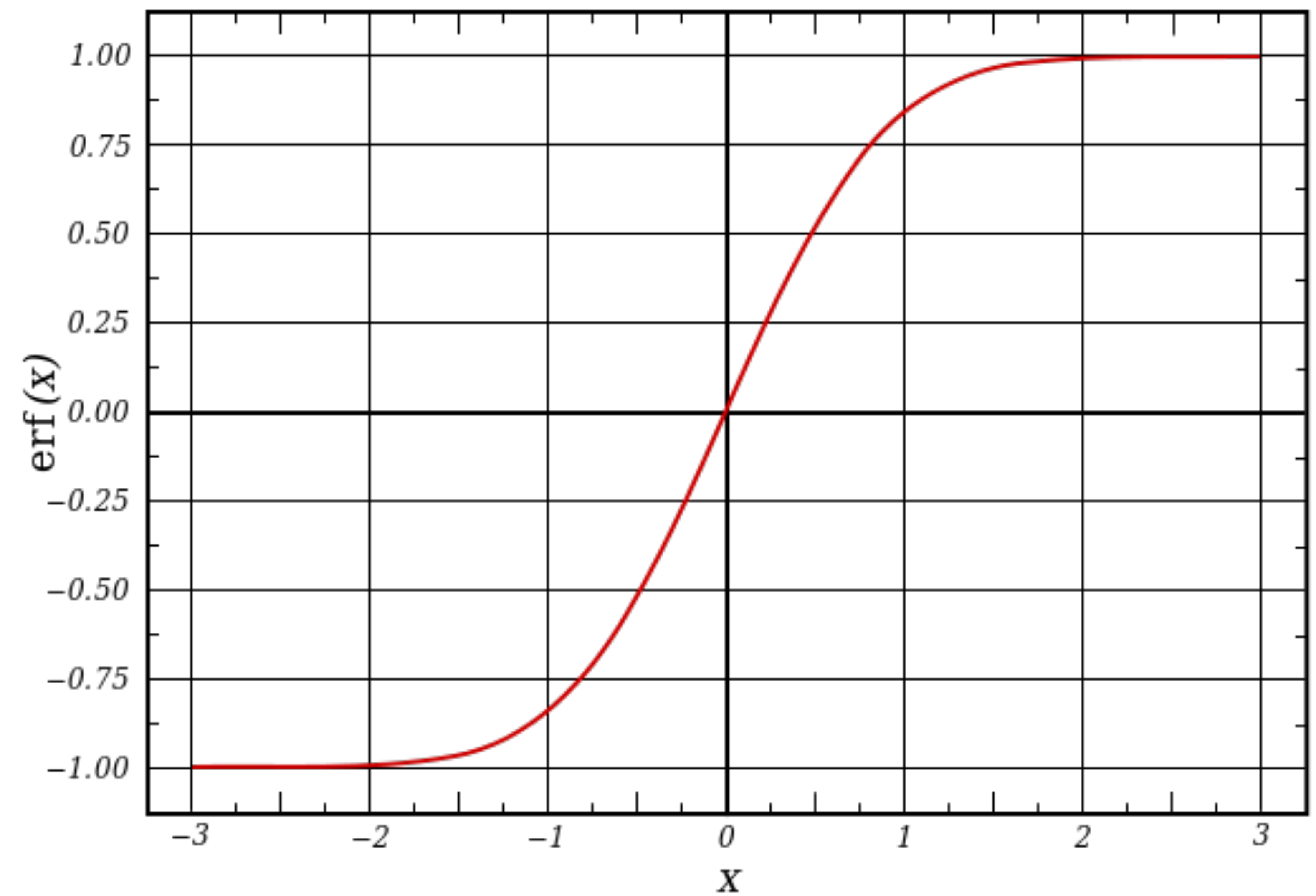
- Distance from mean defined in number of standard deviations sigma

- Probability mass:
  - 68% within 1σ
  - 95% within 2 σ
  - 99.7% within 3σ



Gaussians: Properties

- **Error function** is cumulative integral of Gaussian
- Sigma numbers can be read off



Gaussians: Properties

- Sum of Gaussians has simple form:

$$X \sim \mathcal{N}(\mu_x, \sigma_x^2)$$
$$Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$$
$$\implies X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

- Especially useful for sum of identical Gaussians, and leads to formula that error on the mean ~ $n^{1/2}$

Gaussians: Properties

- **Central limit theorem:**
  Given a collection of random variables X_i:

$$\frac{1}{s_n}\sum_{i=1}^{n}(X_i - \mu_i) \to \mathcal{N}(0,1)$$

$$s_n^2 = \sum_{i=1}^{n}\sigma_i^2$$

- Provided that:

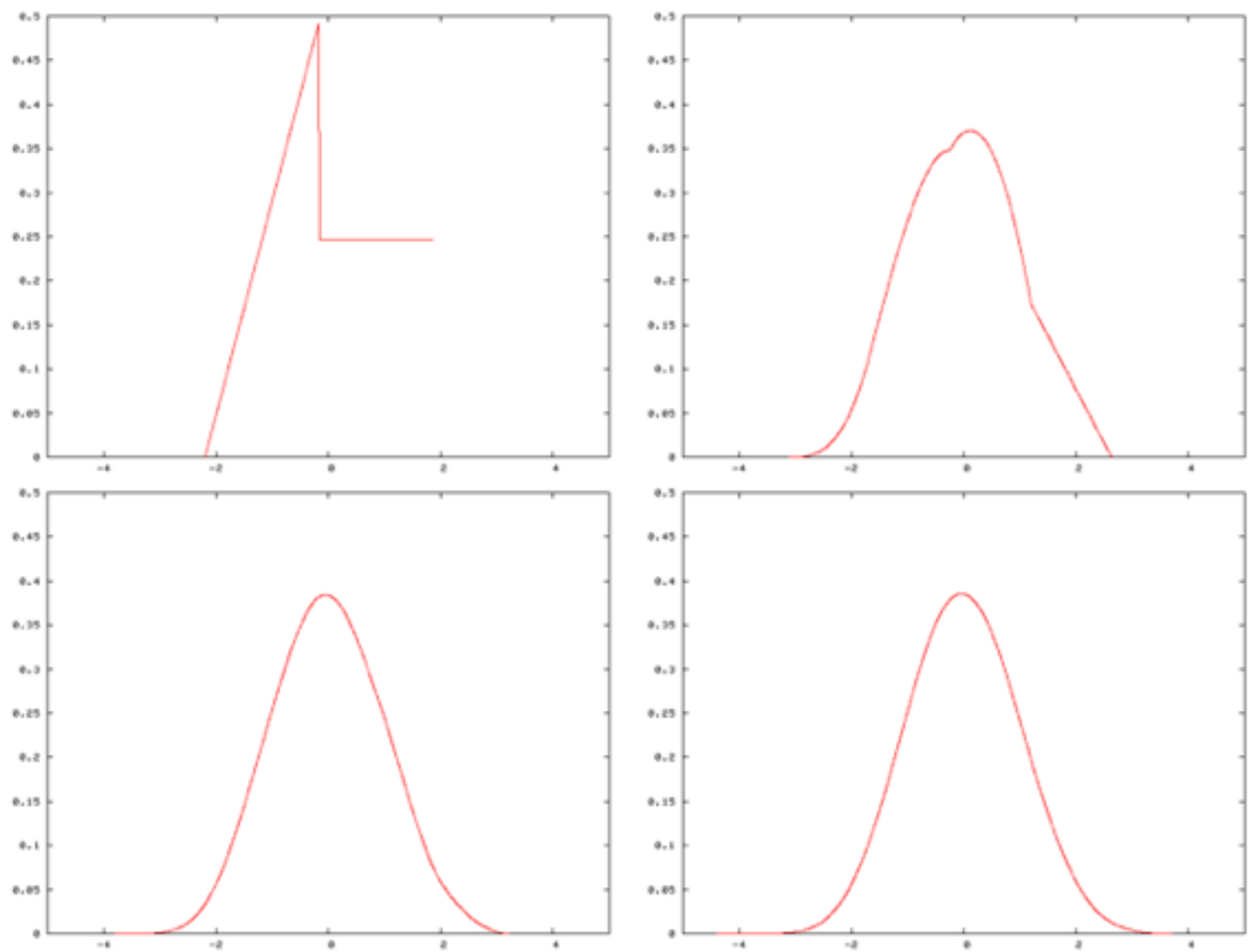
$$\frac{1}{s_n^2}\sum E\left[(X - \mu_i)^2\right] \to 0$$

Gaussians: Properties

- **Central limit theorem:**

Single distribution

Mean of 2



Mean of 3

Mean of 4

Gaussians: Multivariate

$$P(\underline{\boldsymbol{x}};\underline{\boldsymbol{\mu}},C)=\frac{1}{(2\pi)^{\frac{n}{2}}|C|}\exp\left[-\frac{1}{2}(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}})^{T}C^{-1}(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}})\right]$$

- C is the covariance matrix - describes correlations between quantities
    - For example: data points often have correlated errors

Interpretations of Probability

	Frequentists	Bayesians
Use probabilities to …	describe frequencies	quantify information
Think model parameters are …	fixed unknowns	random variables with probabilities
Think data is …	a repeatable random variable	observed and therefore fixed
Call their work …	“Statistics"	“Inference"
Make statements about …	intervals covering the truth x% of the time	constraints on model parameters
Have …	many approaches with lots of implicit choices	one approach with explicit choices

Why Bayesian probability for science?

- Answers the right question
    - We want facts about the world, not about hypothetical ensembles of experiments
- The ideal process is always clear
    - Practical implementations more difficult
- Problems and questions are more explicit

Interpretations of Probability

- Frequentist approach:
  - Construct an estimator, a single number derived from your data points
  - Simulate data under different models and hypotheses and see how often measured estimator value appears

Interpretations of Probability

- Bayesian approach:
    - Construct a probability of the parameters given the data
    - Compute that probability for various points in parameter space to see if they are good fits

Interpretations of Probability

- Most astronomy data analysis takes neither of these approaches
    - Make up a statistic using rules of thumb and things you half remember from undergrad

A few maxims

- Don’t model your data.
  Model the process that led to your data.

- Everything is a distribution.
  Distrust point estimates.

- You can’t learn anything without making assumptions.
  All probabilities are conditional.

Easy Questions

- Show that if X is independent of Y then Y is independent of X
- Linda is 31, single, outspoken, and very bright. She majored in philosophy in college. As a student, she was deeply concerned with racial discrimination and other social issues, and participated in anti-nuclear demonstrations. Estimate the probability of these things being true:

  (1) Linda is active in the feminist movement.
  (2) Linda is a bank teller.
  (3) Linda is a bank teller and active in the feminist movement.

- Show that $P(XY) \leq P(X)$ and $P(XY) \leq P(Y)$
- If a roll a twenty-sided dice and cube the number shown, what is the expectation of the result?

Medium Question

- Photons arrive at a detector with a Poisson distribution with λ =1 photon/s

  Each photon has an energy drawn from a Gaussian distribution with μ = 1000 eV and σ=100 eV.

  Plot the probability distribution of the amount of energy arriving per second.

  The energy of each photon is independent of the number that arrive.

Hard Question

- On my journey to work I can see the bus stop for the last 3 minutes of my walk towards it.

  On my first day I saw one bus go past it before I got there. How long did I think I would have to wait for the next bus?

- You can assume that buses obey Poisson statistics. This is reasonable for British buses.

- If you need to make any other assumptions then describe and justify them.