# *Alanna Connors and the Origins of Principled Data Analysis*

Jeff Scargle (NASA/Ames Research Center)

***Alanna was a key pioneer in using statistically sound principles for analysis of astronomical data. The corresponding viewpoint and algorithms provide a robust toolkit for high-energy astrophysics as we enter the Age of Digital Astronomy.***

131 Astrostatistics in High Energy Astrophysics  Special Session in Memory of Alanna Connors

8 Apr 2013  High-Energy Astrophysics Division

First paper:



30.11 A Broad Band X-Ray Telescope, P.J. SERLEMITSOS, C. GLASSER, R. PETRE*, and A. CONNORS**, NASA/GSFC. NASA's OSS-2 mission includes the Broad Band X-ray Telescope (BBXRT) experiment of the Goddard X-Ray Group. The instrument consists of two co-aligned grazing incidence mirrors with cooled Si(Li) detectors at each focus. Its objectives and capabilities will be discussed with particular emphasis on the unique mirrors. These represent a significant development because they offer a number of attractive features including large throughput, broad energy response, light weight and low cost. Because image quality will largely determine future applications of this type of mirror, we will discuss our progress in attaining the arc minute image inherent in the mirror design.

*NAS/NRC Research Associate
**Also Dept. Physics & Astronomy, Univ. of Maryland

- 98 COMPTEL
- 43 GRB Papers
- 3 Sparse Bayesian Blocks
- 21 Bayesian papers
- 32 spectroscopy
- 42 first author

Total papers:172

Last paper:

**Accounting for Calibration Uncertainties in X-ray Analysis: Effective Areas in Spectral Fitting**
Hyunsook Lee, Vinay L. Kashyap, David A. van Dyk, Alanna Connors, Jeremy J. Drake, Rima Izem, Xiao-Li Meng, Shandong Min, Taeyoung Park, Pete Ratzlaff, Aneta Siemiginowska, Andreas Zezas

# Top 10

The next slide contains my choice for the top 10 statistic methods that have seen important use in astronomy. I purposely did not link the pictures with the moments, to give you a chance to have some fun making these identifications ... and also coming up with your nominations for other events.

1763 Bayes — Probability

1795 Gauss — Least-Squares

1810 Fourier — Harmonic analysis

1898 Schuster — Periodogram

1940 Bellman — Dynamic Programming

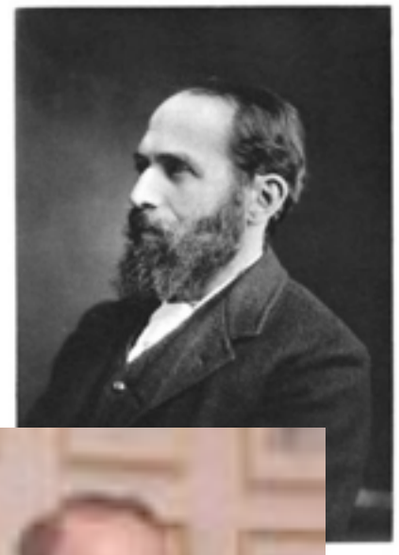
1953 Metropolis, Rosenbluth[2] Teller[2] — MCMC

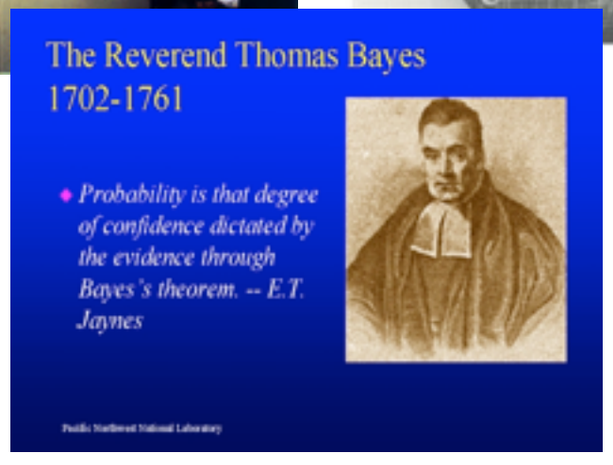
1960  — ARPANET

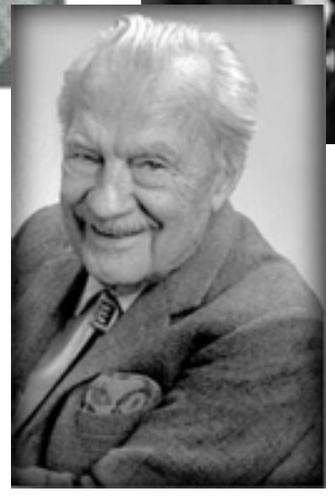
1965 Cooley and Tukey — Fast Fourier Transform

1979 Efron — Bootstrap

1986 Daubechies — Wavelets





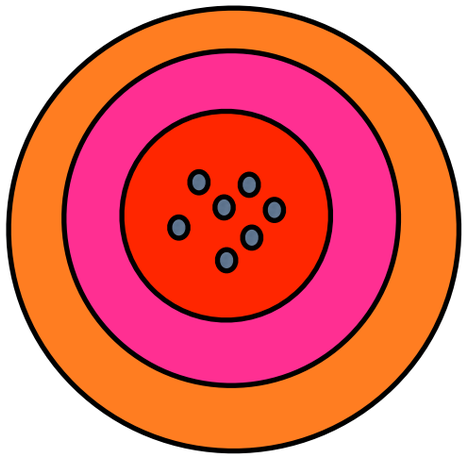
Alanna was a fan (inventor?) of the concept of principled methods. That is to say data analysis procedures based on sound statistical foundations, and opposed to ad hoc methods. The next few slides deal with a few that I think raise important issues.

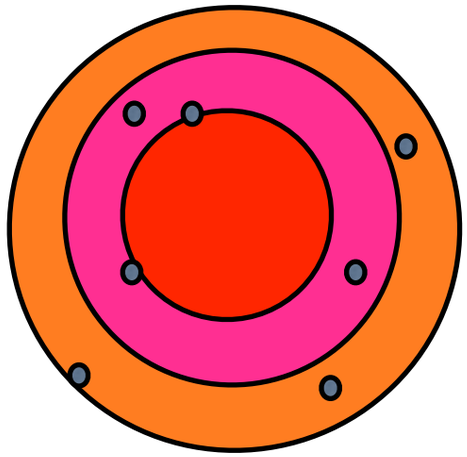
- Bias-Variance Tradeoff & Random vs. Systematic Errors (3 slides)
- Experimental Design (2 slides, adapted from Tom Loredo's work)
- Concern for Error Distributions (1 slide)
- Sampling Bias (1 slide)

Low Variance

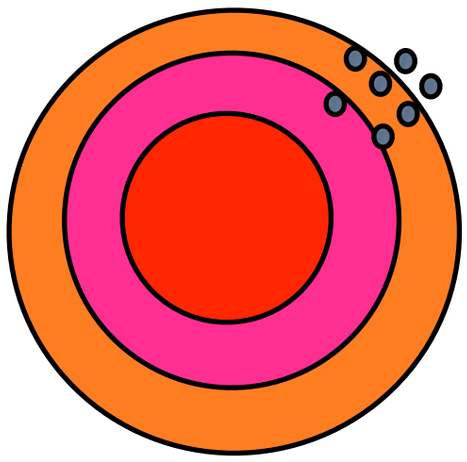
High Variance

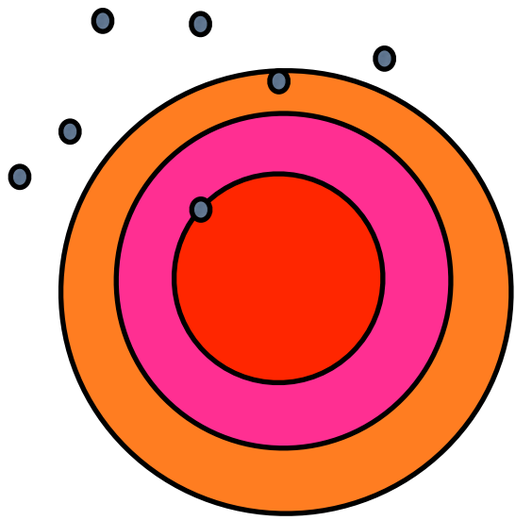
Low Bias





High Bias

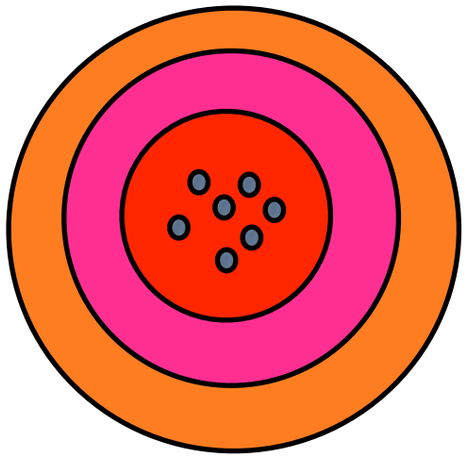


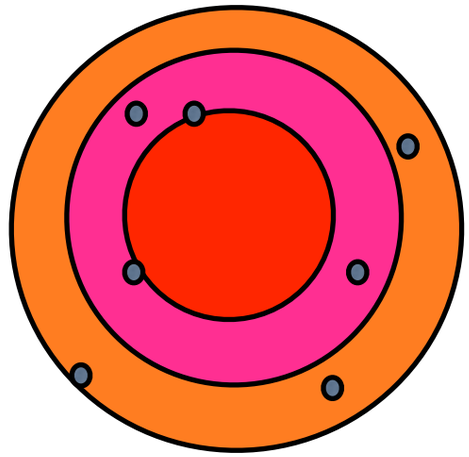


Low Variance

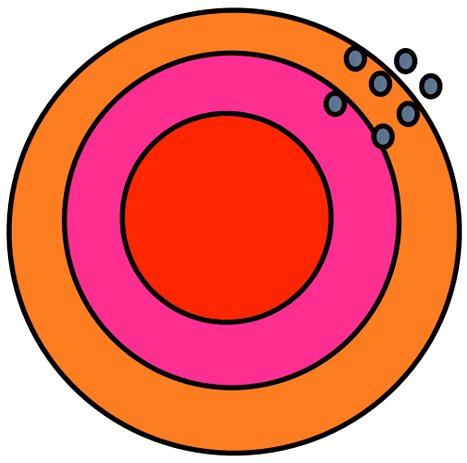
High Variance

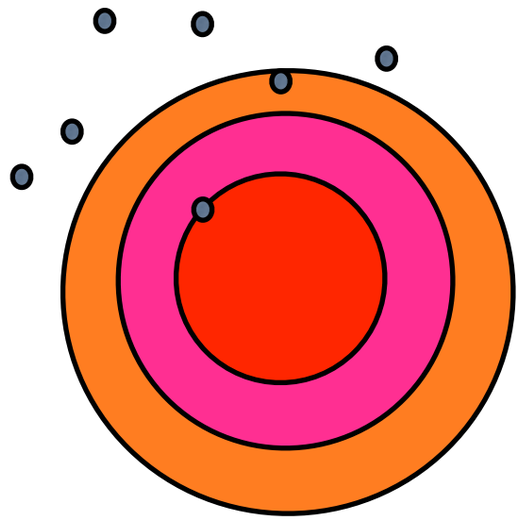
Low Bias





High Bias



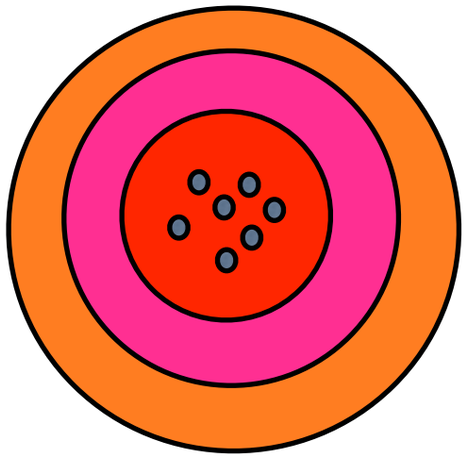


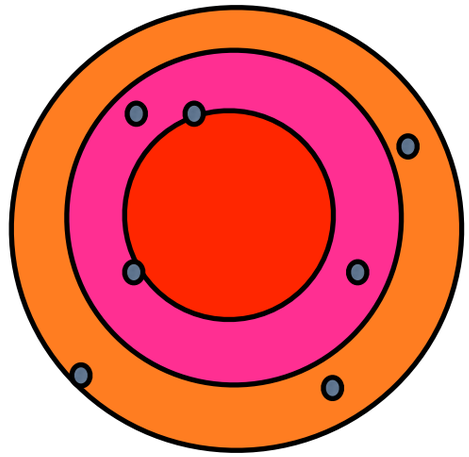
# Bias Variance Tradeoff
# Systematic vs. Random Errors

Low Variance

High Variance

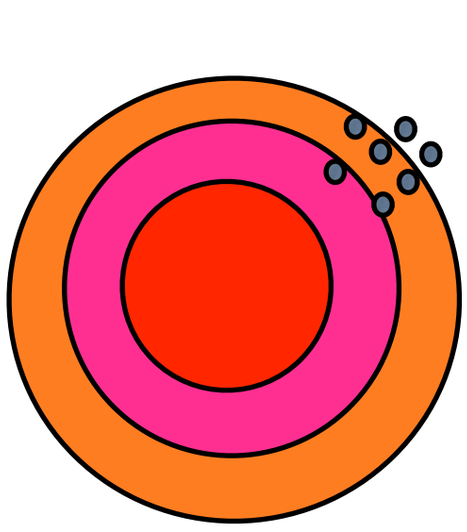
Low Bias

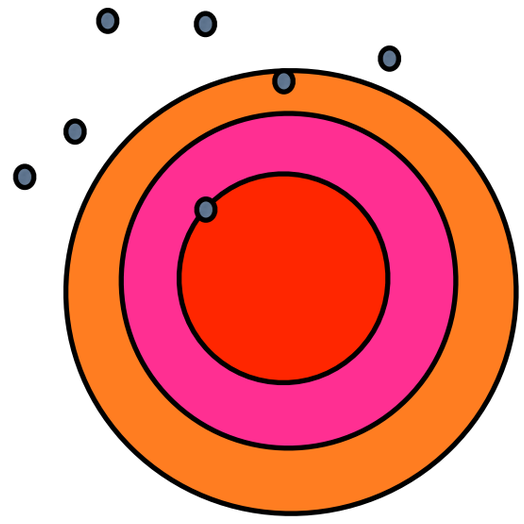




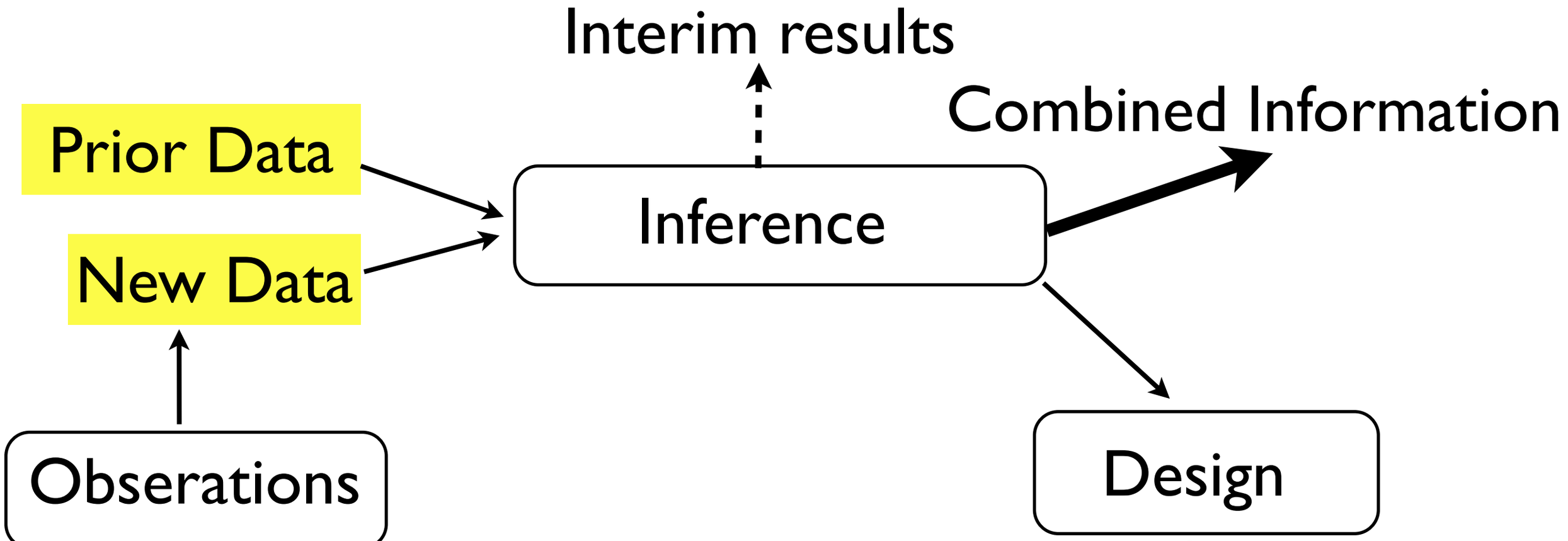
**Accounting for Calibration Uncertainties in X-ray Analysis: Effective Areas in Spectral Fitting**
Hyunsook Lee, Vinay L. Kashyap, David A. van Dyk, Alanna Connors, Jeremy J. Drake, Rima Izem, Xiao-Li Meng, Shandong Min, Taeyoung Park, Pete Ratzlaff, Aneta Siemiginowska, Andreas Zezas

High Bias



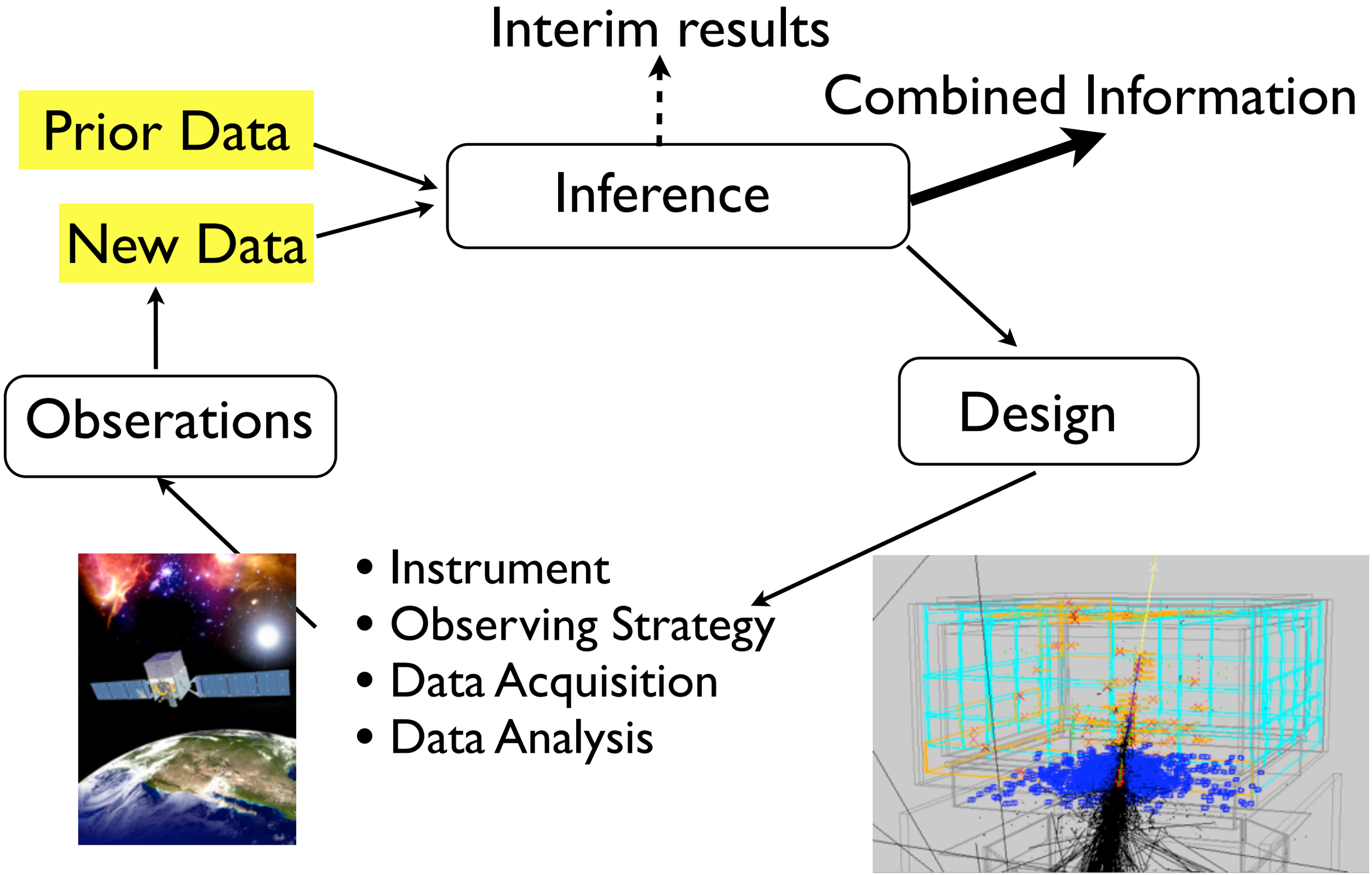


Bias Variance Tradeoff
Systematic vs. Random Errors

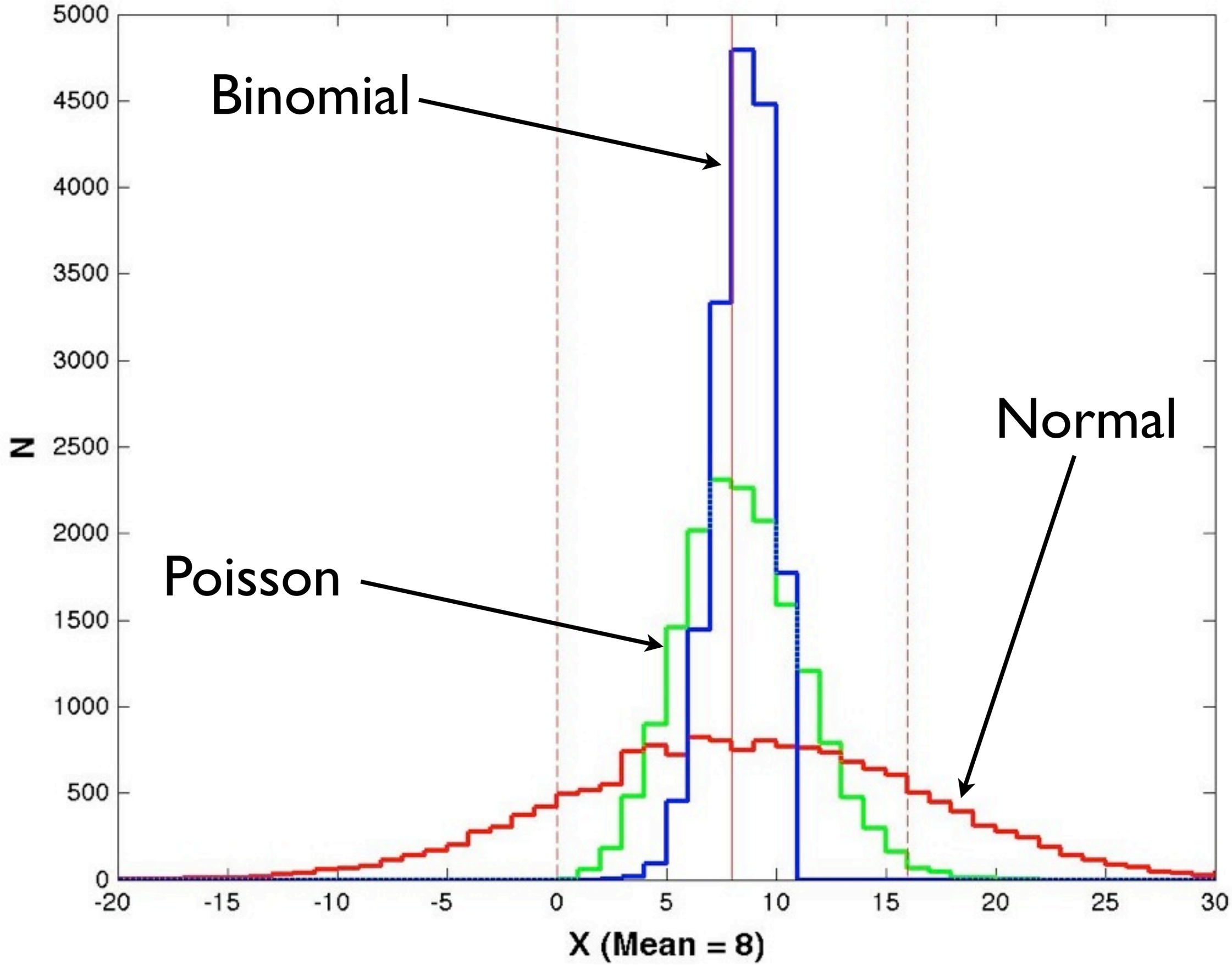


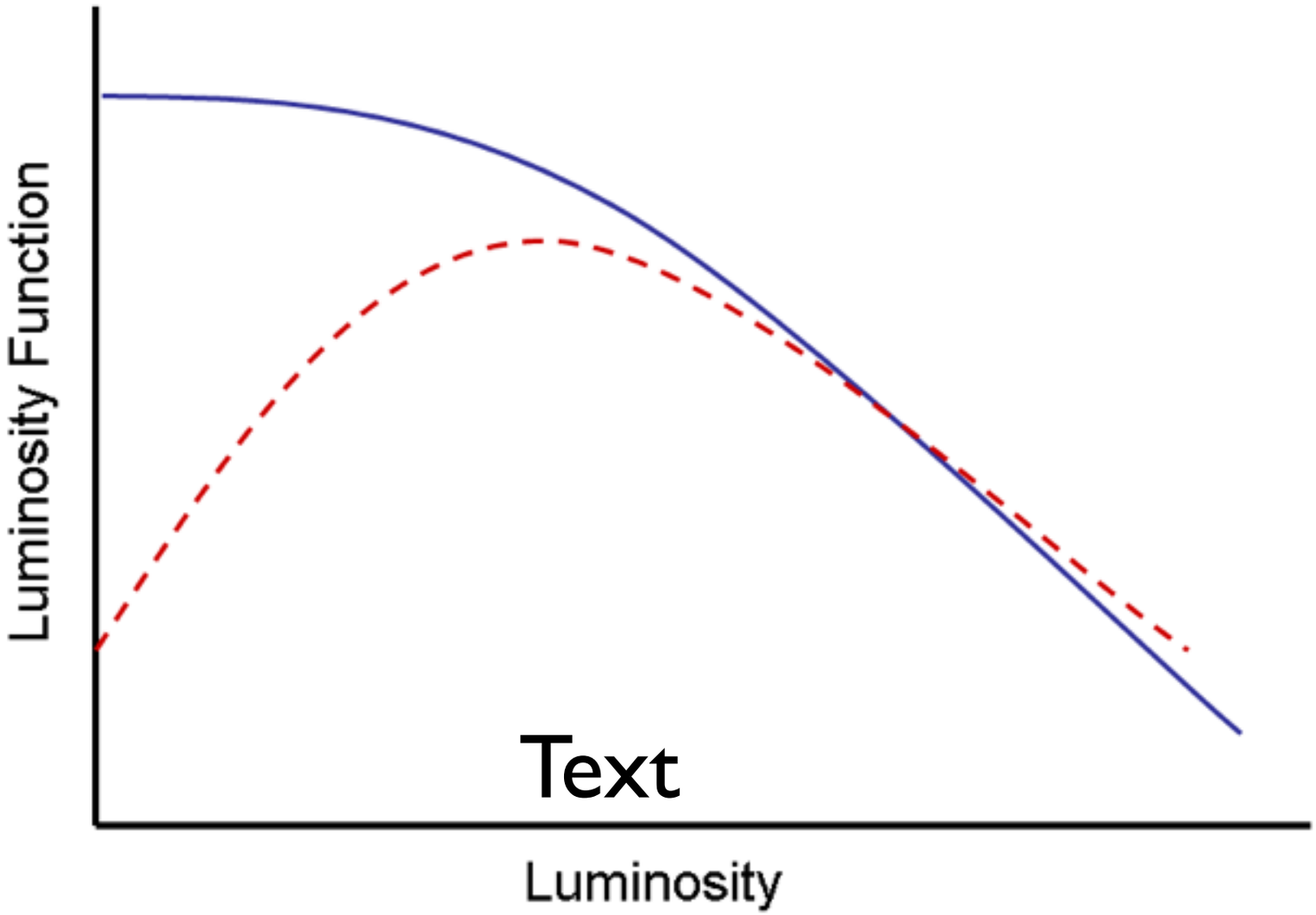

- Instrument
- Observing Strategy
- Data Acquisition & Analysis

***Adapted from Tom Loredo: Bayesian Adaptive Exploration***




***Adapted from Tom Loredo: Bayesian Adaptive Exploration***

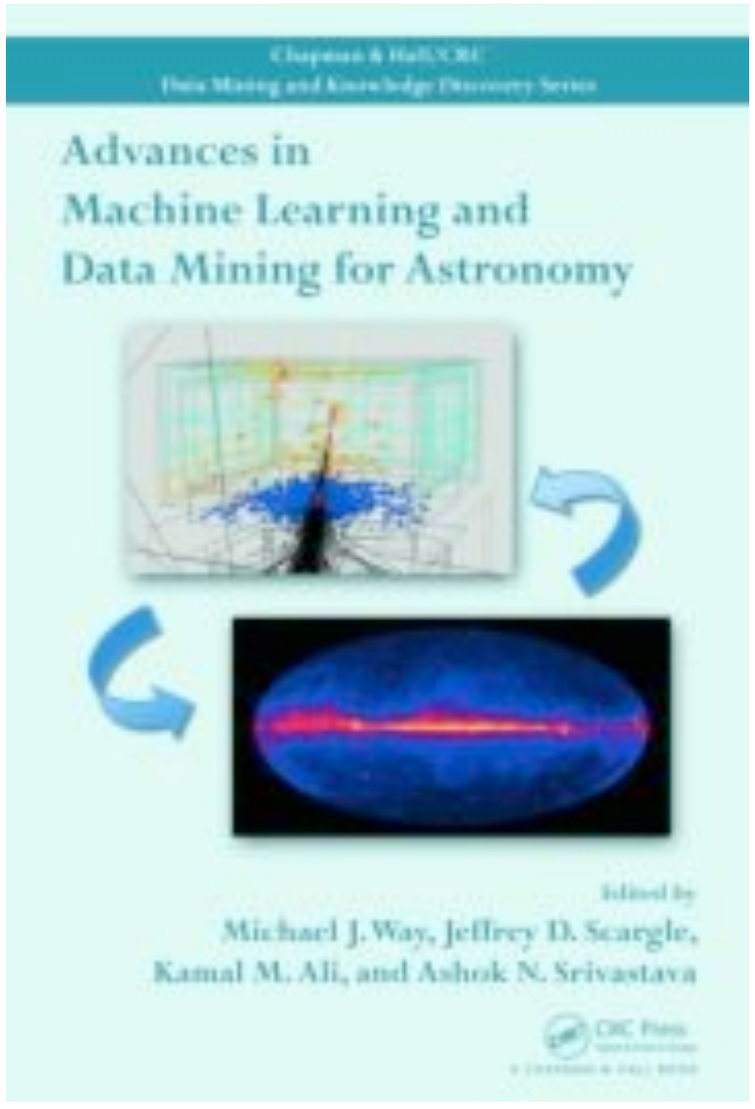

Binomial
Poisson
Normal
N
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
-20
-15
-10
-5
0
5
10
15
20
25
30
X (Mean = 8)




The dashed red line is an example luminosity function when the Malmquist bias is not corrected for. The more numerous low luminosity objects are underrepresented because of the apparent magnitude limit of the survey. The solid blue line is the properly corrected luminosity function using the volume-weighted correction method. (Wikipedia)

# Other Principles of Astrostatistics

Large scale observational programs are making automatic data analysis, machine learning and data mining necessary for scientific progress.



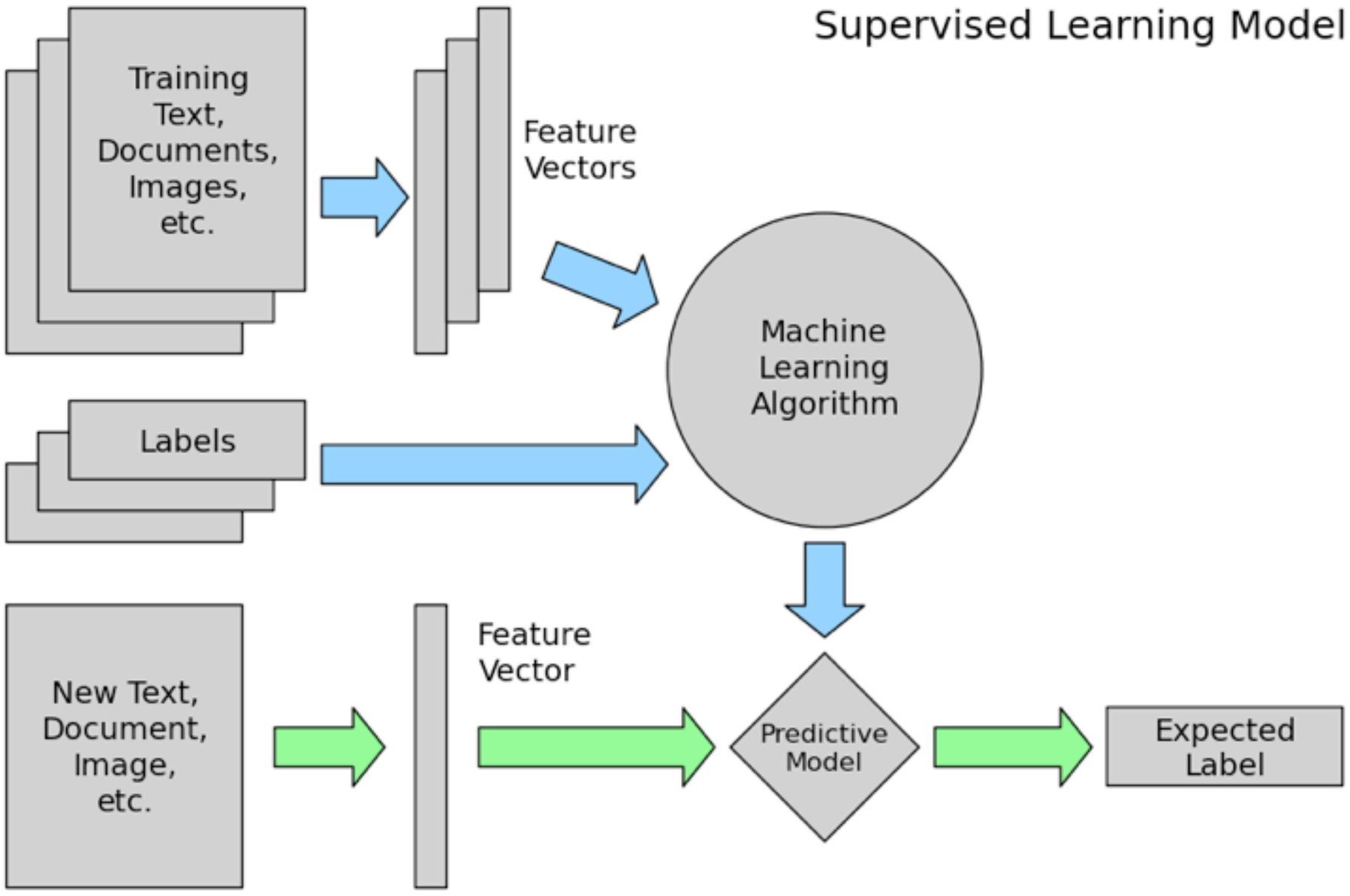
# Machine Learning for Astronomy with Scikit-learn

## http://astroml.github.io/sklearn_tutorial/

Jake Vanderplas, Oliver Grisel, Jaques Grobler, Gael Varoquaux

# AstroML: Machine Learning and Data Mining for Astronomy

## http://astroml.github.io



Simulations are data too, my friend.  Of course they are.

# Thank You Alanna