Chains of Discovery in Astronomy

An Astrostatistical Perspective

Tom Loredo
Dept. of Astronomy, Cornell University
http://www.astro.cornell.edu/staff/loredo/

Banff Discovery Workshop — 12 July 2010

Outline

1. Discovering periodic phenomena: Exoplanets & pulsars
   Detecting signals with (partially) known signatures

2. Discovering anomalies in the CMB
   Detecting signals with unknown signatures

3. Discovering sources of ultra-high energy cosmic rays
   Assessing coincidences (similarity amidst measurement error)

4. Recurring statistical themes

Outline

1. Discovering periodic phenomena: Exoplanets & pulsars
   Detecting signals with (partially) known signatures

2. Discovering anomalies in the CMB
   Detecting signals with unknown signatures

3. Discovering sources of ultra-high energy cosmic rays
   Assessing coincidences (similarity amidst measurement error)

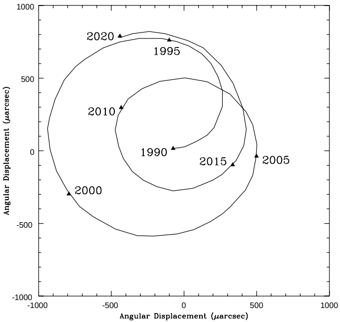
4. Recurring statistical themes

Finding Exoplanets via Stellar Reflex Motion

All bodies in a planetary system orbit wrt the system's center of mass, *including the host star*:

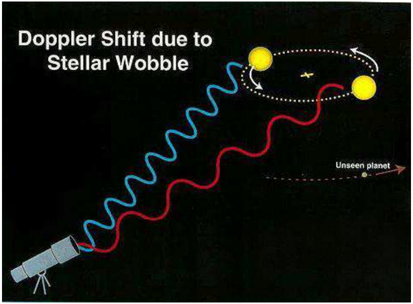
Astrometric Method

Sun's Astrometric Wobble from 10 pc



Doppler Radial Velocity (RV) Method

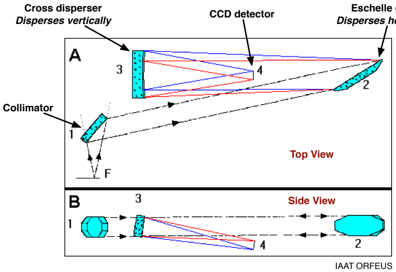
Doppler Shift Along Line-of-Sight



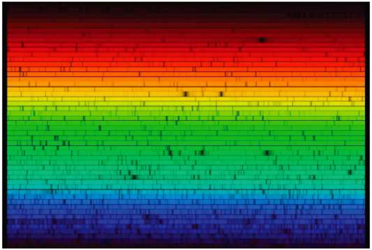
≈ 430 *of* ≈ 460 *currently known exoplanets found using RV method*

RV Data Via Precision Spectroscopy

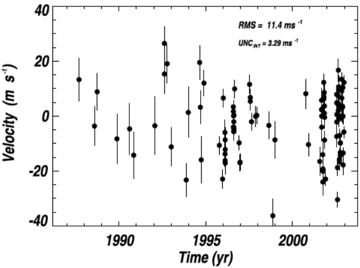
Eschelle Spectrometer



Millipixel spectroscopy



Meter-per-second velocities

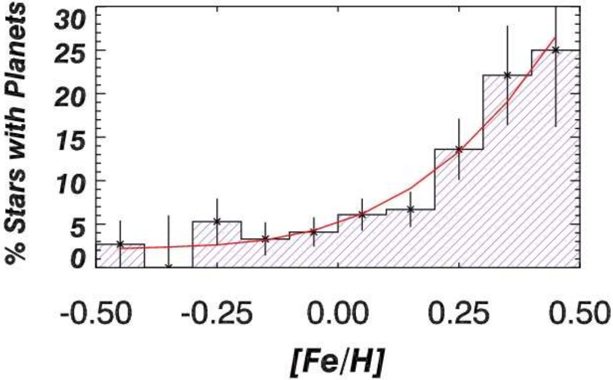


This step involves nontrivial analysis that should be investigated!

Population Properties

California-Carnegie search of ∼1300 FGKM stars

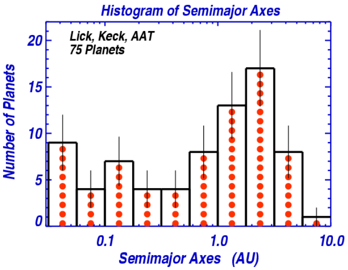
How many stars have planets?

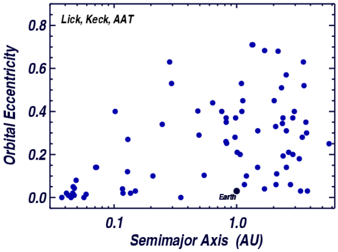


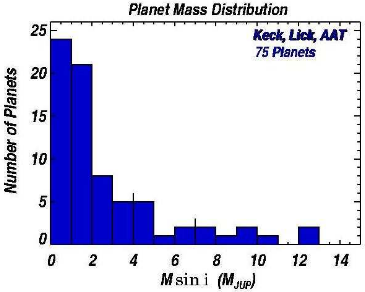
What about the non-detections?

Now that we know this, should it influence detection criteria?

What are the orbits like?



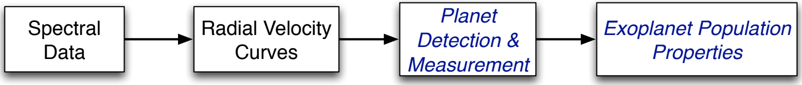




Plots ignore uncertainties & strong selection effects!

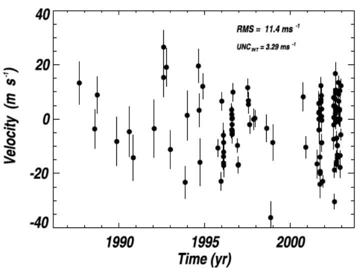
A Variety of Related Statistical Tasks

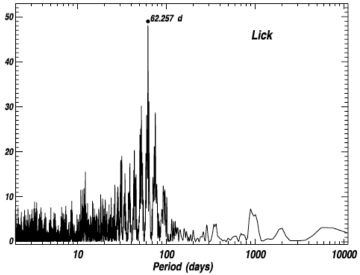
- *Planet detection* — Is there a planet present? Are multiple planets present?
- *Orbit estimation* — What are the orbital parameters? Are planets in multiple systems interacting?
- *Orbit prediction* — What planets will be best positioned for follow-up observations?
- *Population analysis* — What types of stars harbor planets? With what frequency? What is the distribution of planetary system properties?
- *Optimal scheduling* — How may astronomers best use limited, expensive observing resources to address these goals?

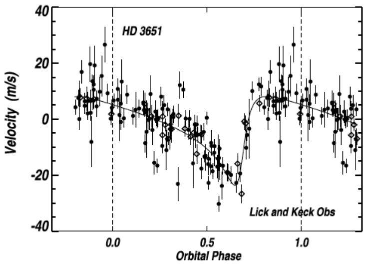


Conventional RV Orbit Fitting

Analysis method: Identify best candidate period via periodogram; fit parameters with nonlinear least squares/min χ^2







System: HD 3651

P = 62.23 d
e = 0.63
m sin i = 0.20 M_J
a = 0.28 AU

Fisher et al. 2003

Schuster Periodogram (1898)

Setting

Find "hidden periodicities" in terrestrial phenomena: weather, earthquakes, magnetic storms

Describe periodicity via:

- Period τ (e.g., seconds, days)
- (Natural) frequency $\nu = 1/\tau$ (e.g., Hz, cycles/day)
- Angular frequency $\omega = 2\pi\nu$ (e.g., rad/s, rad/day)

Data are N *uniformly sampled* measurements contaminated by additive noise:

$$d_i = f(t_i) + \epsilon_i, \qquad t_i = i\delta t$$

Data spacing is δt

Data span a total duration $T = t_N - t_1$ Noise std dev'n σ

Periodogram

"It is convenient to have a word for some representation of a variable quantity which shall correspond to the 'spectrum' of a luminous radiation. I propose the word periodogram*..."*

By analogy with Fourier analysis of a signal:

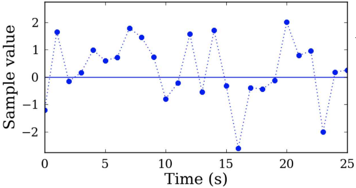
$$\mathcal{P}(\omega) \equiv \frac{1}{N}\left[C^2(\omega) + S^2(\omega)\right],$$

where

$$C(\omega) = \sum_i \frac{d_i}{\sigma}\cos(\omega t_i), \qquad S(\omega) = \sum_i \frac{d_i}{\sigma}\sin(\omega t_i).$$

Use to test compatibility of data with null hypothesis of just noise; report p-value

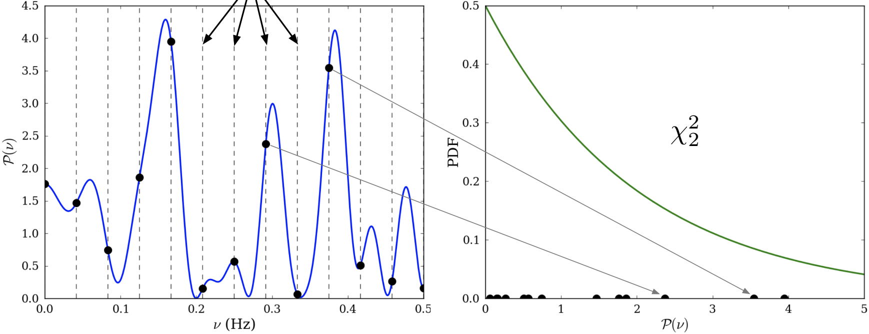
Frequency spacing



26 samples from H_0
Sampling interval $\delta t = 1$ s
Duration $T = 25$ s

Fourier spacing $\delta\nu = 1/N\delta t \approx 1/T$

$N/2$ Fourier frequencies



Connection to (parametric) harmonic analysis (Lomb, Scargle)

Fit a *single sinusoid* model via least squares/max likelihood:

$$\begin{aligned} d_i &= A\cos(\omega t_i + \phi) + \epsilon_i \\ &= A_1 \cos\omega t_i + A_2 \sin\omega t_i \epsilon_i \end{aligned}$$

Sum-squared residuals ($\propto$ log likelihood):

$$\chi^2(\omega, A, \phi) = \sum_i \left[d_i - A\cos(\omega t_i - \phi)\right]^2$$

A (linearly) separable nonlinear model (linear in amplitudes)
$\rightarrow \hat{\mathbf{A}}(\omega) = [\hat{A}_1(\omega), \hat{A}_2(\omega)]$ via linear least squares

Profile likelihood $\mathcal{L}_p(\omega) \propto e^{-\chi^2_p(\omega)/2}$, with

$$\begin{aligned} \chi^2_p(\omega) &\equiv \chi^2(\omega, \hat{\mathbf{A}}(\omega)) \\ &= \chi^2(A=0) - \mathcal{P}(\omega) \quad = \quad \Delta\chi^2(\omega) \end{aligned}$$

Use this to define periodogram for non-uniformly sampled data $\rightarrow$ *Lomb-Scargle periodogram*

Bayesian counterpart (Jaynes, Bretthorst)

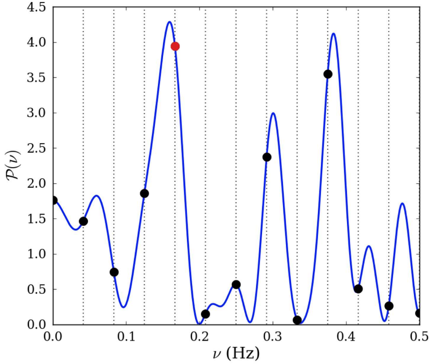
Marginal likelihood for frequency (flat or broad conjugate prior):

$$\begin{aligned}\mathcal{L}_m(\omega) &\equiv \int dA_1 \int dA_2 \, \pi(A_1, A_2) \, \mathcal{L}(A_1, A_2, \omega) \\ &\approx \exp[\mathcal{P}(\omega)/2]\end{aligned}$$

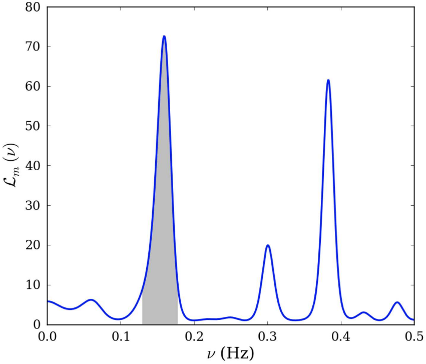
Normality + linearity $\rightarrow$ extremizing and marginalizing lead to similar inferences for *estimation* of ω

Frequentist vs. Bayesian detection

Frequentist approach:
Maximize $\mathcal{P}(\nu)$



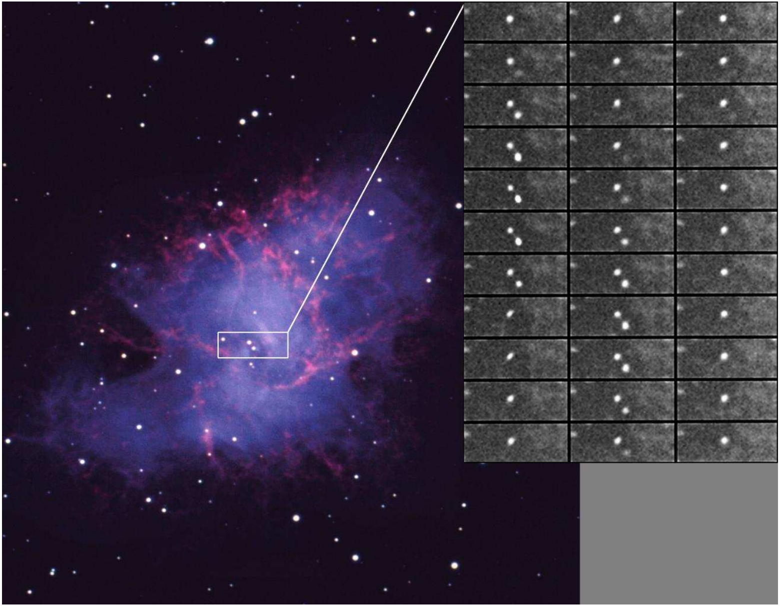
Bayesian approach:
Integrate $\exp[\mathcal{P}(\nu)]$



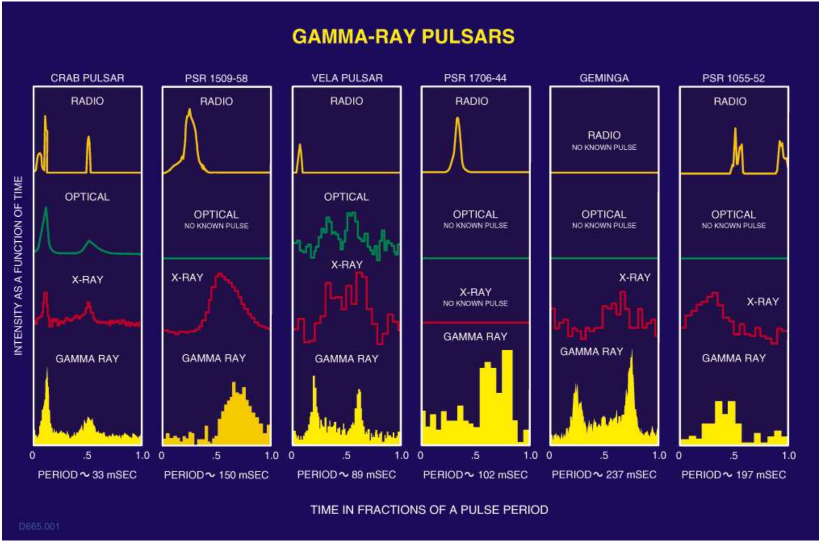
Frequentist p-value must adjust for # of frequencies examined; $p \approx N\hat{p}$
Change sinusoid to *Keplerian RV curve* $\rightarrow$ *Kepler periodogram*

Note: There is no fundamental requirement for a frequentist approach to maximize; Bickel^{+} (2006) *integrate* a statistic derived from the score function, over phase and frequency.

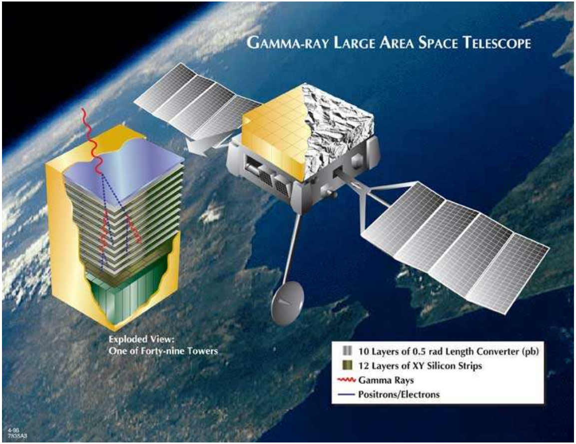
The Crab Pulsar



Pulsars from Radio to Gamma Rays



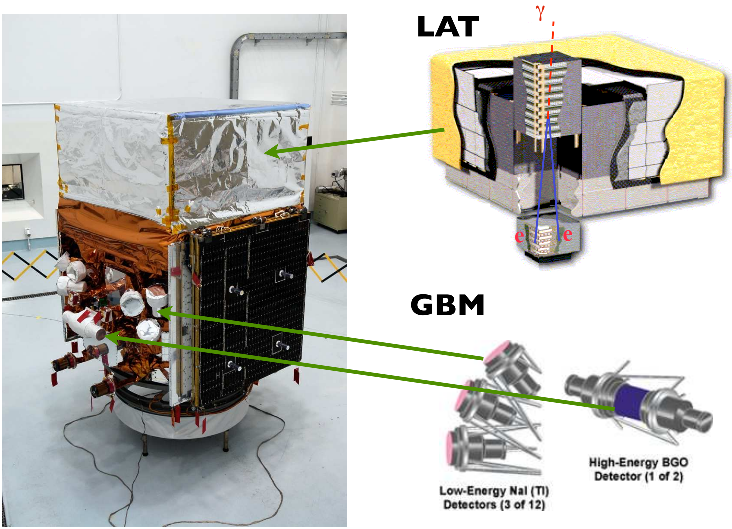
Getting the Gammas: Fermi



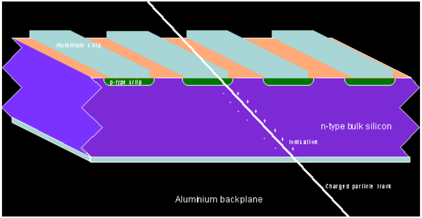
Launched June 2008

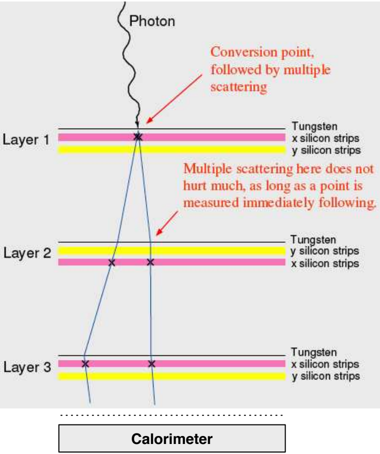
Fermi's Large Area Telescope

Particle physics in space!



Silicon Tracker: Complex Data

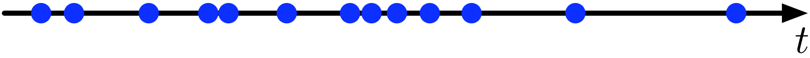




LAT Tracker Data → Gamma Ray Properties → *Pulsar Detection & Measurement*

Time, direction, energy

Gamma Ray Data: Photon Arrival Times



Data D: $\{t_i\}$ for $i = 1$ to N

Non-homogeneous Poisson point process sampling distribution, rate $r(t;\theta)$:

$$p(D|\theta, M) = \exp\left[-\int_T dt\, r(t)\right] \prod_{i=1}^{N} r(t_i)$$

Conventional Approaches

Try to reject a "null" (constant rate) hypothesis with an omnibus test for candidate ω; report p-value adjusted for # trials.

- Rayleigh statistic:

$$R^2(\omega) = \frac{1}{N}\left[\left(\sum_{i=1}^{N}\cos\omega t_i\right)^2 + \left(\sum_{i=1}^{N}\sin\omega t_i\right)^2\right]$$

- Z_n^2 statistic:

$$Z_n^2(\omega) = \sum_{j=1}^{n} R^2(j\omega)$$

- χ^2-Epoch folding:
  - Fold data with trial period $\rightarrow$ phases $\theta_i = \omega t_i \mod 2\pi$; bin $\rightarrow n_j$, $j = 1$ to M
  - Calculate Pearson's $\chi^2(\omega)$ vs. $n_j = N/M$; *average over phase*

Needle in a Huge Haystack

Recall the frequency scale: $\delta\nu \approx 1/T$.

For gamma-ray pulsars $T \sim$ weeks to months for $N \sim 10^3$ to 10^5:

$$1/T \sim .1 \text{ to } 1 \ \mu\text{Hz} \qquad\qquad \nu_{\max} \sim 3 \text{ kHz}$$

$\rightarrow \sim 10^9$ frequencies to examine

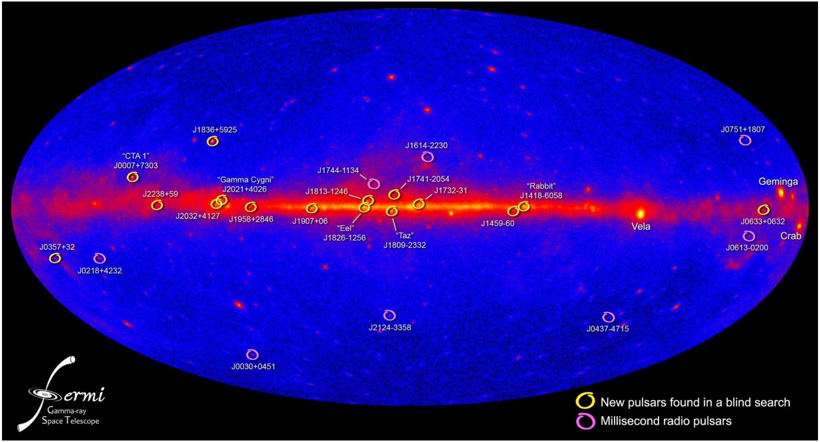
Actually much worse: pulsars *spin down* $\rightarrow$ need $\dot{\nu}$ parameter;
$\sim 10^3$ $\dot{\nu}$ values to explore

Clever tricks reduce this burden by a few orders of magnitude:

- Tapered time difference FFTs (Atwood$^+$ 2006)
- Tapering + dynamic programming (Meinshausen$^+$ 2008)

Still need *simple* yet sensitive methods
What statistic is "best"? What role is there for Bayesian models?

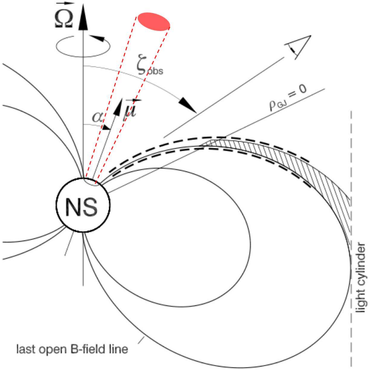
Fermi's Pulsars



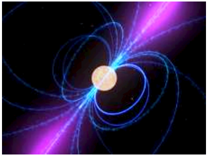
Detected 46 (54?) gamma ray pulsars to date, 16 (24?) unknown at other wavelengths

Population-Level Pulsar Science

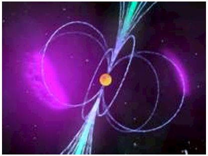
Emission physics



Radio & Gamma From Cap



Gamma From Gap



GSFC

Fermi data point to separate regions for radio and gamma emission

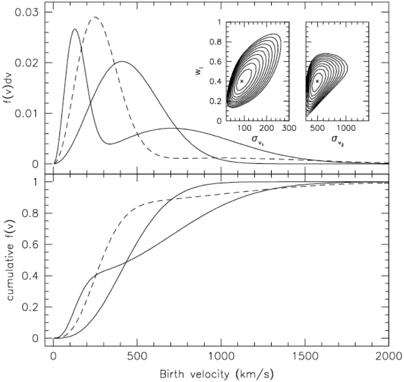
LAT Tracker Data → Gamma Ray Properties → *Pulsar Detection & Measurement* → *Pulsar Population Properties*

Neutron star formation (radio pulsars)

Arzoumanian+ (2002):

- Data = Locations, velocities, spins, luminosities of radio pulsars
- Model birth, motion, beam geometry, lifetime
- Bayesian multilevel model accounts for selection & uncertainties
- Compare rival models with Bayes factors

Dist'n of Pulsar Birth Kicks

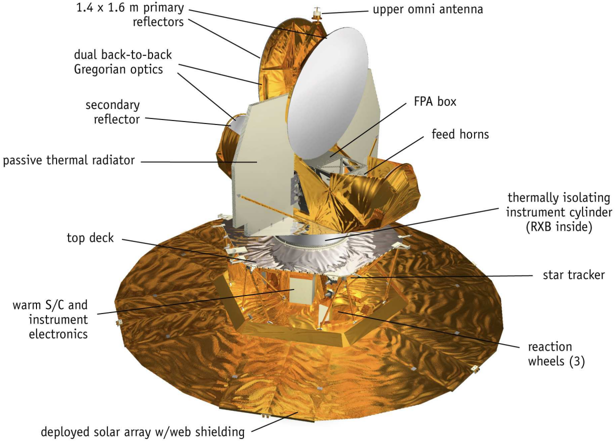


Outline

1. Discovering periodic phenomena: Exoplanets & pulsars
   Detecting signals with (partially) known signatures
2. Discovering anomalies in the CMB
   Detecting signals with unknown signatures
3. Discovering sources of ultra-high energy cosmic rays
   Assessing coincidences (similarity amidst measurement error)
4. Recurring statistical themes

Measuring the Cosmic Microwave Background

From space: WMAP

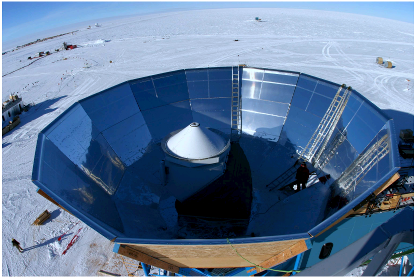


Measures the whole sky

From the South Pole

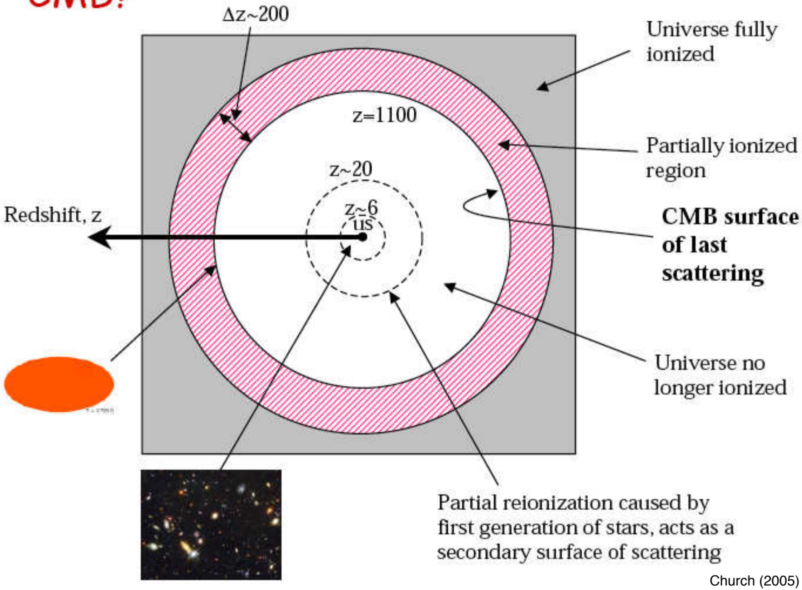
BOOMERanG

QUaD



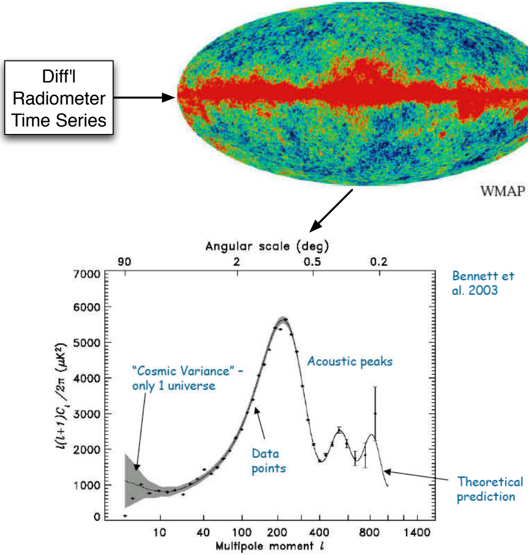
Measure small patches at high resolution to learn about small angular scales

What are we seeing when we look at the CMB?



Church (2005)

CMB Measurements



Also get polarization maps and spectra

Modeling the CMB

Expand temperature fluctuations in spherical harmonics:

$$\begin{aligned}
\frac{\Delta T(\mathbf{n})}{\bar{T}} &\equiv \frac{T(\mathbf{n}) - \bar{T}}{\bar{T}} \\
&= \sum_{l=1}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\mathbf{n})
\end{aligned}$$

Make two assumptions (approximations?):

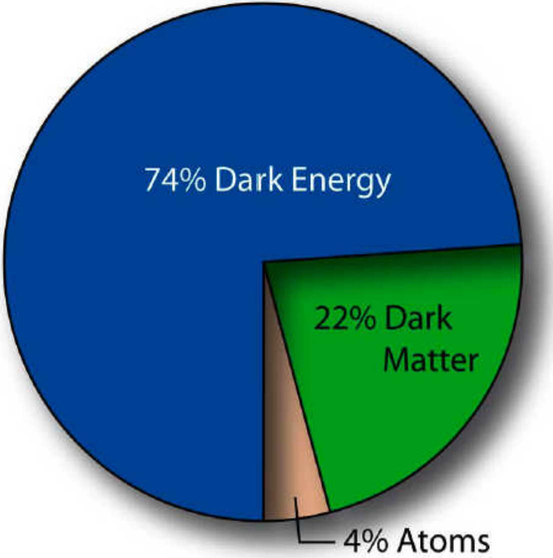
- Large-scale *isotropy*
- *Gaussianity*

Then a theory with parameters θ predicts *variances* (power spectrum) $C_l(\theta)$ such that:

$$p(a_{lm}|C_l) = \frac{1}{\sqrt{2\pi C_l}} \exp\left(-\frac{1}{2}\frac{|a_{lm}|^2}{C_l}\right)$$

Note: Only $2l+1$ measurements constrain $C_l \rightarrow$ "cosmic variance"

What the Universe is Made Of

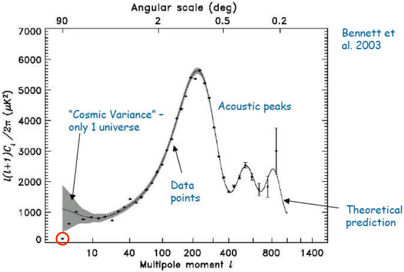


Coles 2009

Testing Assumptions

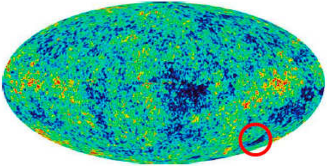
Look for non-Gaussianity, unusual anisotropy (e.g., from nontrivial topology):

- Low quadrupole?
- Patterns in *m*-dependence?
- Hot/cold spots?
- Alignment of multipoles?
- ...



Methods use wavelets, needlets, higher-order corr'n functions...
Issues:

- What can spatial statistics offer?
- When you look so hard for something unusual in voluminous data, aren't you bound to find something?



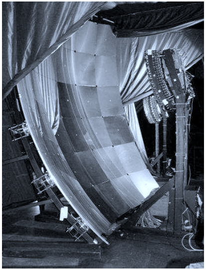
Outline

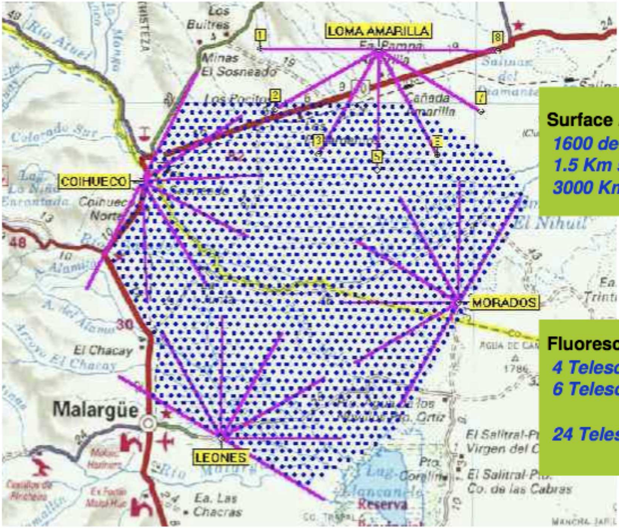
1. Discovering periodic phenomena: Exoplanets & pulsars
   Detecting signals with (partially) known signatures
2. Discovering anomalies in the CMB
   Detecting signals with unknown signatures
3. Discovering sources of ultra-high energy cosmic rays
   Assessing coincidences (similarity amidst measurement error)
4. Recurring statistical themes

The Highest Energy Particle Physics

Pierre Auger Observatory

- 3000 sq km sites, Argentina
- SDs: 1600 H2O Cerenkov tanks/site
- FDs: 4 flourescence tel/station
- Northern (CO): R&D funded; plan 4000 improved SDs over 20k sq km





Surface Array

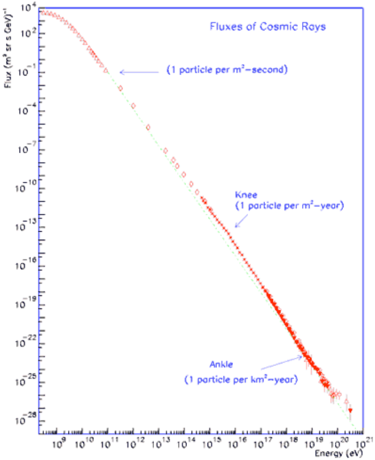
1600 detector stations
1.5 Km spacing
3000 Km²

Fluorescence Detectors

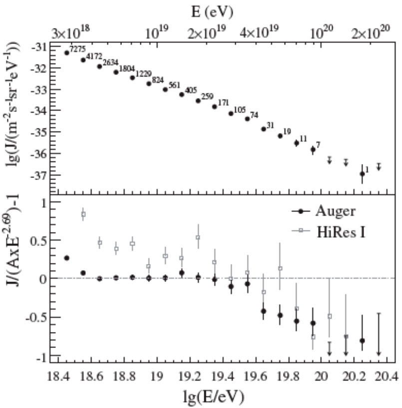
4 Telescope enclosures
6 Telescopes per enclosure
24 Telescopes total

Ultra-High Energy Cosmic Rays

CRs with $E > 10^{19.5}$ eV are suppressed by interaction with the CMB.



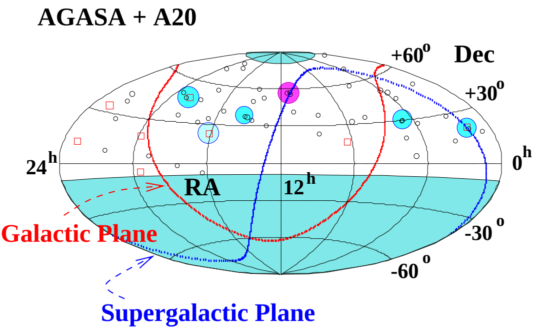
GZK Cutoff!



Auger Collab'n 2008

Coincidences Among UHE Cosmic Rays?

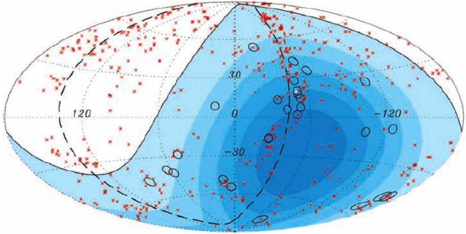
AGASA data above GZK cutoff (Hayashida et al. 2000)



- 58 events with $E > 4 \times 10^{19}$ eV
- Energy-dependent direction uncertainty $\sim 2^\circ$
- Significance test — Search for coincidences $< 2.5^\circ$:
  - 6 pairs; $\lesssim$1% significance
  - 1 triplet; $\lesssim$1% significance

Auger UHE CR Data: AGN Association?

Auger data above GZK cutoff (Nov 2007)



- 27 events with $E > 5.7 \times 10^{19}$ eV
- Energy-dependent direction uncertainty $\lesssim 1^\circ$
- Crosses = 472 AGN with distance $D < 75$ Mpc
- Significance test of correlation with AGN:
  - Tune E, D, angle cutoffs with early events
  - Apply to 13 new events → p-value 1.7×10^{-3}

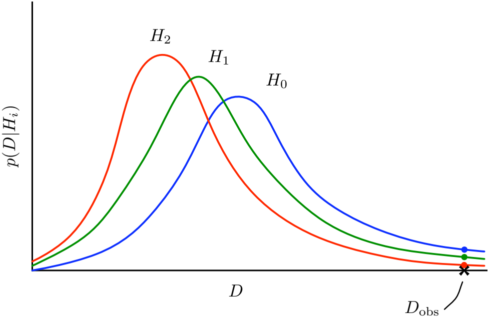
Issues In Coincidence Assessment

Directional, spatial, spatio-temporal

- **Directional uncertainties** — *"Hard" box/annulus, spherical "normal" distributions; extended sources*
- **Choice of statistic** — How to measure "close"? *Nearest neighbor distance, correlation functions, Mahalanobis distance, likelihood ratio, Bayes factor*
- **Proximity criterion** — How close is "close enough"? *Significance level/p-value, power, odds/Bayes factor*
- **Multiple testing** — How to account for number of candidates? *Bonferroni, FDR, Bayes; exploration algorithms*

Bayesian Model Comparison

- Calculate $p(D|H_i)$ for each H_i
- Favor the hypothesis that makes the observed data most probable (up to a prior factor)

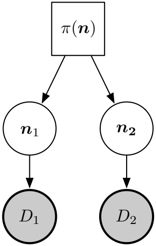


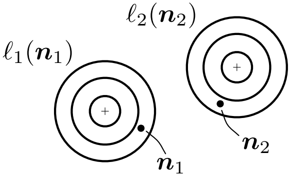
If the data are improbable under H_0, the hypothesis *may* be wrong, *or* a rare event may have occured. Significance tests reject the latter possibility at the outset.

Bayesian Coincidence Assessment

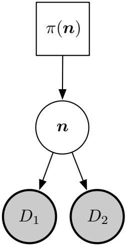
Two-Source Case

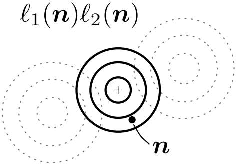
Not associated





Associated





Direction uncertainties accounted for via likelihoods for object directions:

$$\ell_i(\mathbf{n}) = p(d_i|\mathbf{n}), \quad \text{normalized w.r.t. } \mathbf{n} \text{ (convention)}$$

E.g., Fisher distribution for azimuthally symmetric errors

H_0: No repetition

$$\begin{aligned}
p(d_1, d_2|H_0) &= \int d\mathbf{n}_1 \, p(\mathbf{n}_1|H_0)\, \ell_1(\mathbf{n}_1) \quad \times \int d\mathbf{n}_2 \cdots \\
&= \frac{1}{4\pi} \int d\mathbf{n}_1 \, \ell_1(\mathbf{n}_1) \quad \times \frac{1}{4\pi} \int d\mathbf{n}_2 \cdots \\
&= \frac{1}{(4\pi)^2}
\end{aligned}$$

H_1: Associated (same direction!)

$$p(d_1, d_2|H_0) \quad = \quad \int d\mathbf{n}\; p(\mathbf{n}|H_0)\, \ell_1(\mathbf{n})\, \ell_2(\mathbf{n})$$

Odds favoring association:

$$\begin{aligned} O &= 4\pi \int d\mathbf{n}\; \ell_1(\mathbf{n})\, \ell_2(\mathbf{n}) \\ &\approx \frac{2C}{\sigma_{12}^2} \exp\left[-\frac{C\theta_{12}^2}{2\sigma_{12}^2}\right]; \quad \sigma_{12}^2 = \sigma_1^2 + \sigma_2^2 \end{aligned}$$

	Odds O	
Angular error	$\theta_{12} = 26^\circ$	$\theta_{12} = 0^\circ$
$\sigma_1 = \sigma_2 = 10^\circ$	≈ 1.5	≈ 75
$\sigma_1 = \sigma_2 = 25^\circ$	≈ 7	≈ 12

Challenge: Large hypothesis spaces

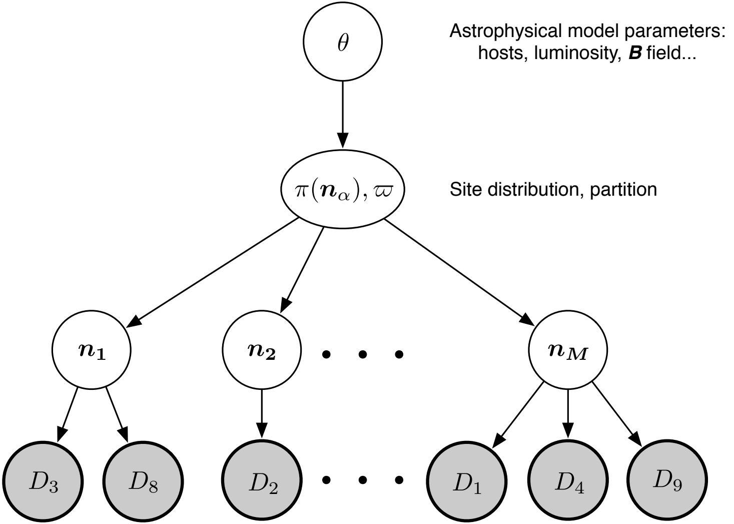
For $N = 2$ events, there was a single coincidence hypothesis, M_1 above.

For $N = 3$ events:

- Three doublets: $1 + 2$, $1 + 3$, or $2 + 3$
- One triplet

The number of alternatives (partitions, ϖ) grows combinatorially:

- Must assign sensible priors to partitions
- Must deal with computational challenge of summing over them


θ
Astrophysical model parameters:
hosts, luminosity, $\boldsymbol{B}$ field...
$\pi(\boldsymbol{n}_\alpha), \varpi$
Site distribution, partition
$\boldsymbol{n_1}$
$\boldsymbol{n_2}$
$\boldsymbol{n_M}$
D_3
D_8
D_2
D_1
D_4
D_9

Small-N Brute Force Example

Bayesian Coincidence Assessment for AGASA UHECRs

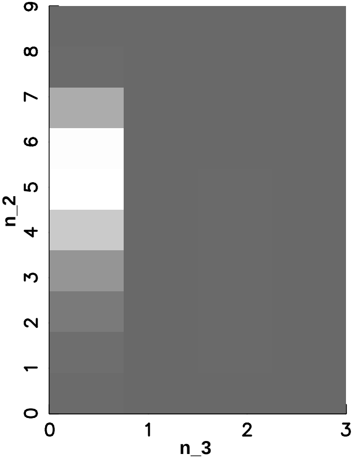
$N = 58$ directions; search for coincidences

n_2	n_3	$\mathcal{N}$
1	0	1653
2	0	1,272,810
3	0	607,130,370
0	1	30,856
0	2	404,753,580

Method:

- Identify all pairs (13) and triplets (3) with multiplet Bayes factors > 1
- Generate & sum over all partitions including those multiplets (gives lower bound)
- Use flat prior over all possible (n_2, n_3)

Odds for repetition: 1.4 (i.e., no significant evidence)



Advantages of Bayesian Approach

Several "tuning" issues in conventional approaches are addressed by *averaging* over choices:

- No angular cuts: Introduce latent $\mathbf{n}_i$ and marginalize
- Energy cut may be similarly dealt with by introducing distances and B field parameters and marginalizing
- Inter/intra-galactic B field scatter similarly dealt with via parameters controlling inflation of uncertainties

Remaining issue: Choice of candidate population(s) to associate with.

Is there any non-subjective way to handle this?

Outline

1. Discovering periodic phenomena: Exoplanets & pulsars
   Detecting signals with (partially) known signatures

2. Discovering anomalies in the CMB
   Detecting signals with unknown signatures

3. Discovering sources of ultra-high energy cosmic rays
   Assessing coincidences (similarity amidst measurement error)

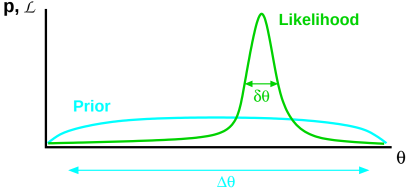
4. Recurring statistical themes

Some Recurring Statistical Themes

- Accounting for hypothesis multiplicity ("look elsewhere?")
- Choosing a test statistic
- Building chains of discovery & measurement

Penalizing for Searching

Bayes factors and "Occam factors"



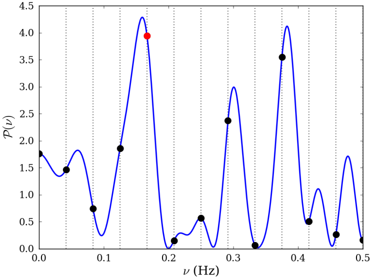
$$
\begin{aligned}
p(D|M_i) &= \int d\theta_i \; p(\theta_i|M) \; \mathcal{L}(\theta_i) \approx p(\hat{\theta}_i|M)\mathcal{L}(\hat{\theta}_i)\delta\theta_i \\
&\approx \mathcal{L}(\hat{\theta}_i)\frac{\delta\theta_i}{\Delta\theta_i} \\
&= \text{Maximum Likelihood} \times \text{Occam Factor}
\end{aligned}
$$

Models with more parameters often make the data more probable — *for the best fit*

Occam factor penalizes models for "wasted" volume of parameter space

Quantifies intuition that models shouldn't require fine-tuning

Frequentist multiplicity corrections



Frequentist periodicity searches test the null at N statistically independent frequencies. If $\hat{p}$ is the largest p-value, the reported p-value is calculated with a *Bonferonni correction* for test multiplicity:

$$p = 1 - (1 - \hat{p})^N \approx N\hat{p} \quad \text{for small } \hat{p}$$

This controls the "family-wise error rate" (FWER) at level p; we seek to not have a single false rejection among any of the N tests.

Questions

Since $N \propto$ the frequency range, this behaves similarly to the Occam factor. Does this ameliorate the Jeffreys-Lindley "paradox" in this setting (in favor of Bayes-like behavior)?

The periodogram is just a log MLR (for periodic vs. constant alternatives). Why is period uncertainty treated differently here than in the usual Wilks's theorem setting ($\Delta\chi^2$ with DOF $= \#$ new params)?

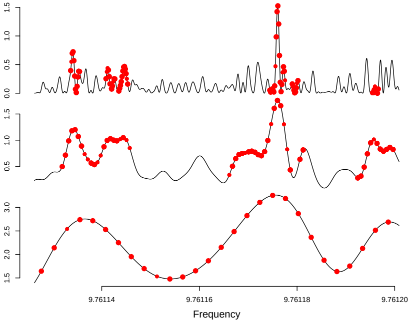
Put differently: Bayesian model assessment *always* takes into account the *range* of the space searched; MLR tests usually only account for the *dimension*. What conditions require range-dependent corrections (from frequentist POV)?

Optimal Searching Under Cost Constraints

The number of periodogram bumps grows linearly with size & duration of data; the size of period-drift search space grows quadratically. How to efficiently search?

Meinshausen, Bickel & Rice (2009):
"Recursive coarsening"

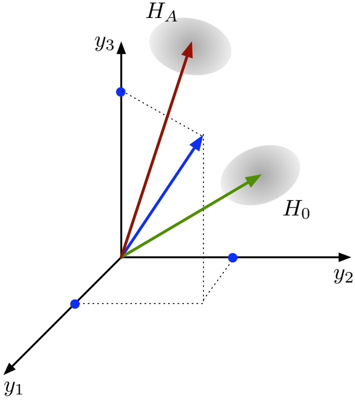
- Throw out some information in the data (e.g., tapered time series)
  - Reduces sensitivity to small signals (decreases power)
  - But also reduces computational cost for searching
- Use search with coarsened data to *focus* subsequent search with less-coarsened data
- Use *dynamic programming* to optimally search subject to cost constraints



Alternatives and Choice of Test Statistic

Setting

- Construct a test with *small* Type I error rate α for H_0
- Seek *large* power $\beta(H_A)$ against alternatives
- Focus on *local alternatives*: Distance from $H_0 \sim C/\sqrt{n}$



Limitations of Omnibus Goodness-of-Fit Tests

Theorem (Janssen 2000; Lehman & Romano 2005)

- $\beta \approx \alpha$ for all H_A except those along a finite set of directions (independent of n)
- The number of directions grows with α

"The results are not surprising. Every statistician knows that it is impossible to separate an infinite sequence of different parameters simultaneously if only a finite number of observations is available."

Freedman's Theorems (3 of 5)

"Diagnostics cannot have much power against general alternatives" (Freedman 2009 & forthcoming book)

- Consider a GoF test for smoothness of data pdf (no δ-functions):
  There are pdfs with large δ's for which $\beta \approx \alpha$
- Consider a GoF test for 2-D IID hypothesis, $\alpha < 1/2$:
  There are some alternatives with $\rho \approx 1$ for which $\beta \approx 0$
- Consider a GoF test of H_0 against all H_A:
  There are some *remote* H_A (various metrics) with $\beta(H_A) \to 0$

Outlook

- "Specification error is extremely difficult to evaluate using internal evidence."
- Diagnostics should be performed more often; they can pick up gross modeling errors.
- "Skepticism about diagnostics is warranted . . . a model can pass diagnostics with flying colors yet be ill-suited for the task at hand."
- "A proper choice of test must be based on some knowledge of the possible set of alternatives for a given experiment." (L&R 2005)

Pulsar searching: Bickel, Kleijn & Rice (2007)

No matter how clever you are, no matter how rich the dictionary from which you adaptively compose a detection statistic, no matter how multilayered your hierarchical prior, your procedure will not be globally optimal.

- Uses Bickel^{+} (2006) framework to concentrate power in set of a priori specified orthogonal directions
- Uses a finite Fourier basis
- Uses score (derivative of log-$\mathcal{L}$ at null) rather than $\mathcal{L}$ ratio
- Handles frequency uncertainty via averaging
- Handles frequency derivative

Score Tests: Lazy Scientist's MLR?

Maximum likelihood ratio (MLR) test

Consider a model with parameter θ, and two (simple) hypotheses:

$$H_0 : \theta = \theta_0 \qquad\qquad H_1 : \theta = \theta_1$$

The most powerful test of H_1 vs. H_0 rejects H_0 if likelihood ratio exceeds a critical value:

$$\frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} \equiv \frac{p(D|\theta_1)}{p(D|\theta_0)} > C$$

If H_1 is composite, the MLR test plugs in the MLE for θ_1,

$$\frac{\mathcal{L}(\hat{\theta}_1)}{\mathcal{L}(\theta_0)} > C$$

Requires fitting both null and alternative

Score test

Consider log-likelihood ratio test for $\theta_1 = \theta_0 + \epsilon$:

$$\begin{aligned} L(\theta_0 + \epsilon) - L(\theta_0) &> \log C \qquad \text{for } L(\theta) \equiv \log \mathcal{L}(\theta) \\ \rightarrow \epsilon \frac{dL}{d\theta}\bigg|_{\theta_0} &> \log C \end{aligned}$$

Define the *score function* $S(\theta) \equiv dL/d\theta$; build a test statistic from $S(\theta_0)$.

Can show that:

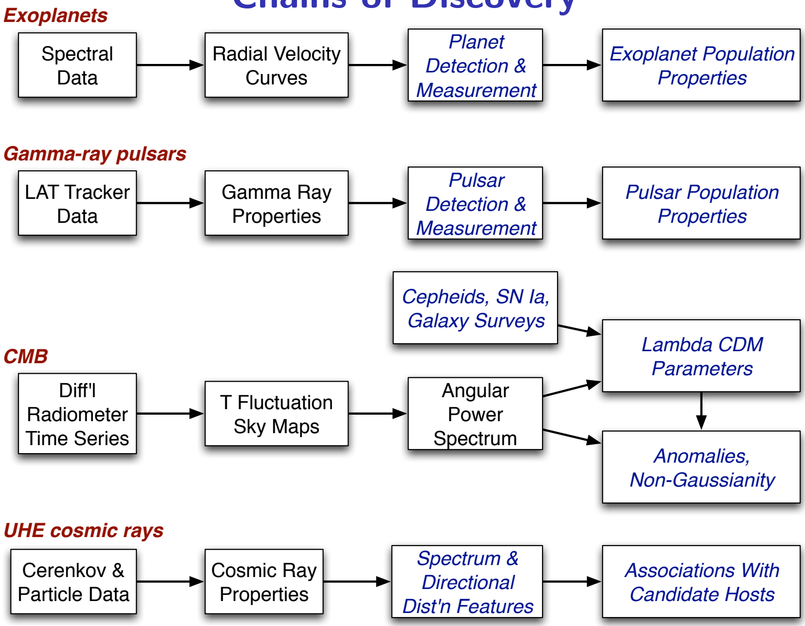
$$\mathbb{E}[S(\theta)|\theta] = 0, \qquad \mathbb{E}[S^2(\theta)|\theta] = \mathcal{I}(\theta) \quad \text{(Fisher info)}$$

So test the null using $Q = \frac{S^2(\theta_0)}{\mathcal{I}(\theta_0)} \sim \chi^2_1$ (asymp.)

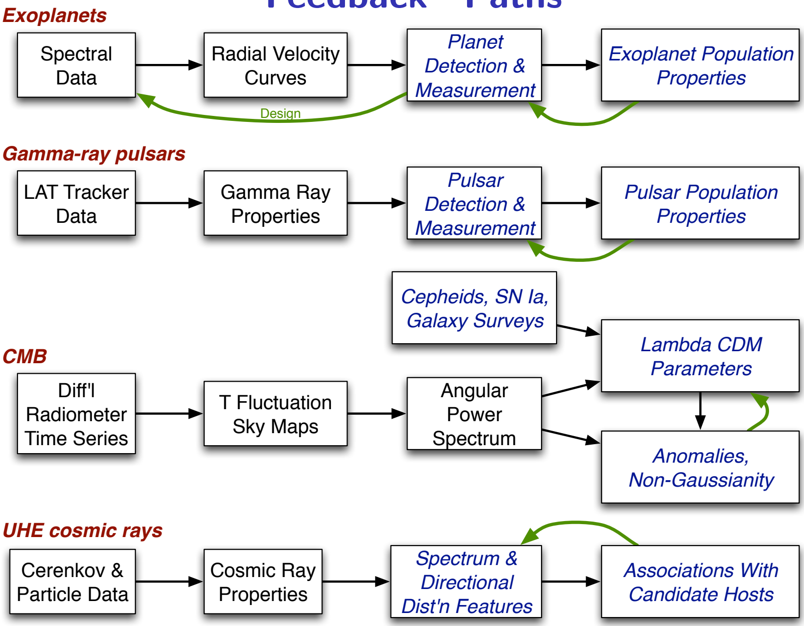
- *Does not require any fitting of* H_1
- Approximately the MLR test for $\theta_1 = \theta_0 + \epsilon$

Bickel$^+$ (2006) use this for pulsar searching, testing against alternatives in a Fourier family.

Chains of Discovery



“Feedback” Paths



Implementing Chains of Discovery

Discoveries must be communicated to facilitate reuse:

- Bayes: report *marginal* likelihoods for interesting parameters
  But priors must be chosen carefully if many possibly related parameters are marginalized
- Frequentist: report *profile* likelihoods
  But we know that profiling can be bad, tragically so in measurement error problems
- Frequentist: InCA group's method for propagating uncertainty by inverting tests

How to implement "feedback"/adaptivity?

Feedback Example: Adaptive Threshold vs. Multilevel Modeling

Setting: Counting sources (real vs. spurious)

Measure $N = 100$ objects with additive Gaussian noise, $\sigma = 1$:

- 30 have $A = 2.2$
- 70 have $A = 0$

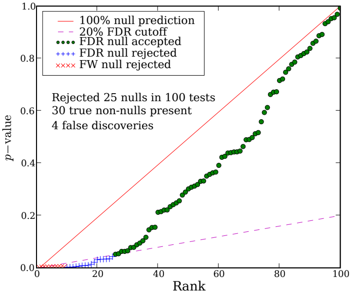
Detect via 100 tests of $H_0 : A = 0$

Source Present	Detection Result: *Negative*	*Positive*	Total
H_0: *No*	T_-	F_+	ν_0
H_1: *Yes*	F_-	T_+	ν_1
Total	N_-	N_+	N

Thresholding Controlling FWER and FDR

Threshold criteria:

- Control family-wise error rate at level α: accept objects with p-values$p = \alpha/N$, aiming to not make a single false discovery $\rightarrow$ 9 (accurate) discoveries for FWER $= 20\%$
- Control *false discovery rate*, $\langle F_+/N_+\rangle = 20\%$ via Benjamini-Hochberg $\rightarrow$ 25 discoveries (4 false)
- Other choices possible



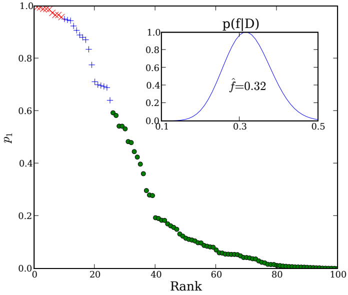
Issue with FRD control: Astronomers will use detections to infer distributions; will be biased for dim sources

Multilevel Model Approach

Let f = fraction of objects with $A = 2.2$.

If f were known, it would the prior probability for a Bayesian odds calculation.

Treat f as *unknown* (flat prior); infer it from the data:



One can say there are about 30 sources present, without being able to say for sure whether many of the candidates are sources or not.
Caution: The "upper level" prior needs some care in more complex settings (Scott & Berger 2008; MLM literature)

Final Provocation

Thesis: Important data analyses are often used sequentially

- Sequential experimentation/exploration
- Chains of discovery (individual $\rightarrow$ population)

Herman Chernoff on sequential analysis (1996):

> *I became interested in the notion of experimental design in a much broader context, namely: what's the nature of scientic inference and how do people do science? The thought was not all that unique that it is a sequential procedure...*
>
> *Although I regard myself as non-Bayesian, I feel in sequential problems it is rather dangerous to play around with non-Bayesian procedures.... Optimality is, of course, implicit in the Bayesian approach.*