Learning How To Count: Poisson Processes

(Lecture 3)

Tom Loredo

Dept. of Astronomy, Cornell University

Motivation/Terminology

We consider processes that produces discrete, isolated events in some interval, possibly multidimensional. We will make inferences about the event rates per unit interval. Examples:

- Arrival time series: $D = \{t_i\}$, rate $r(t) =$ events s^{-1}
- Photon # flux: $D = \{t_i, x_i, y_i\}$, flux $F(t, x, y) =$ photons cm^{-2} s^{-1}
- Spectrum: $D = \{\epsilon_i\}$, specific intensity $I_\epsilon(\epsilon) =$ cts keV^{-2}
- Population studies: $D = \{L_i\}$, luminosity function $n(L) =$ events/luminosity

If our measurements are coarse, we “bin” events and can only report the number of events in one or more finite intervals. Then the appropriate model is the *Poisson counting process*.

If our measurements have sufficient resolution for us to measure every individual event, the appropriate model is the *Poisson point process*. If the event characteristics are measured with error, it is a *point process with error*.

If the event rate is constant over the entire interval of interest, the process is *homogeneous*; otherwise it is *inhomogeneous*.

Today's Lecture

- Poisson Process Fundamentals
- Poisson counting processes—Photon counting
- Poisson point processes—Arrival time series
- Point processes with error:
  - Population studies—TNO size distribution
  - Spatio-temporal coincidences—GRBs, cosmic rays

Poisson Process Fundamentals

For simplicity we consider 1-d processes; for concreteness, consider time series.

Let $r(t)$ be the event rate per unit time.

Let $E =$ "An event occured in $[t, t + dt]$"

Let Q denote any kind of information about events occuring or not occuring in other intervals.

A Poisson process model results from two properties (M):

- Given the event rate $r(t)$, the probability for finding an event in a small interval $[t, t + dt]$ is proportional to the size of the interval:

$$p(E|r, M) = r(t)\,dt$$

- Information about what happened in other intervals is irrelevant if we know r; the probabilities for separate intervals are independent:

$$p(E|Q, r, M) = p(E|r, M) = r(t)\,dt$$

Homogeneous Poisson Counting Process

Basic datum: The number of events, n, in a given interval of duration T. We seek $p(n|r, M)$.

No event:

$$h(t) = P(\text{no event in } [0,t]|r,M); \qquad h(0) = 1$$

$$\begin{aligned} A = & \text{ "No event in } [0, t+dt]\text{"} \\ = & \text{ "No event in } [0,t]\text{" } \textbf{AND} \\ & \text{ "No event in } [t, t+dt]\text{"} \end{aligned}$$

$$\begin{aligned} P(A|r,M) &= h(t+dt) = h(t)[1 - r\,dt] \\ h(t) + dt\frac{dh}{dt} &= h(t) - r\,dt\,h(t) \\ \frac{dh}{dt} &= -r\,h(t) \\ \Rightarrow h(t) = e^{-rt} \end{aligned}$$

One event:

$$B = \text{“One event is seen in } [0,T] \text{ in } [t_1, t_1 + dt_1]\text{”}$$

$$P(B|r,M) = e^{-rt_1} \cdot (r\,dt_1) \cdot e^{-r(T-t_1)} = e^{-rT} r\,dt_1$$

$$p(n=1|r,M) = \int_0^T dt_1\; r\, e^{-rT} = (rT)e^{-rT}$$

Two events:

$C =$ "Two events are seen in $[0, T]$ at (t_1, t_2) in (dt_1, dt_2)"

$$\begin{aligned} P(C|r, M) &= e^{-rt_1} \cdot (r\,dt_1) \cdot e^{-r(t_2-t_1)} \cdot (r\,dt_2) \cdot e^{-r(T-t_2)} \\ &= e^{-rT} r^2\,dt_1\,dt_2 \end{aligned}$$

$$\begin{aligned} p(n = 2|r, M) &= \int_0^T dt_2 \int_0^{t_2} dt_1\; r^2\, e^{-rT} \\ &= r^2\, e^{-rT} \int_0^T dt_2 t_2 \\ &= \frac{(rT)^2}{2} e^{-rT} \end{aligned}$$

$$\Rightarrow \quad p(n|r, M) = \frac{(rT)^n}{n!} e^{-rT}$$

The *Poisson Distribution* for n.

Moments:

$$\begin{aligned}\langle n\rangle &\equiv \sum_{n=0}^{\infty} n\, p(n|r,M) \\ &= rT \equiv \bar{n}\end{aligned}$$

$$\left[\langle(n-\bar{n})^2\rangle\right]^{1/2} = \sqrt{\bar{n}}$$

$$p(n|\bar{n},M) = \frac{\bar{n}^n}{n!}e^{-\bar{n}}$$

$\bar{n}$ specifies both the mean and standard deviation.

Inferring a Rate from Counts

Problem: Observe n counts in T; infer r

Likelihood:

$$\mathcal{L}(r) \equiv p(n|r, M) = p(n|r, M) = \frac{(rT)^n}{n!} e^{-rT}$$

Prior: Two standard choices:

- r known to be nonzero; it is a scale parameter:

$$p(r|M) = \frac{1}{\ln(r_u/r_l)} \frac{1}{r}$$

- r may vanish; require $p(n|M) \sim$ Const:

$$p(r|M) = \frac{1}{r_u}$$

Predictive:

$$p(n|M) = \frac{1}{r_u}\frac{1}{n!}\int_0^{r_u} dr (rT)^n e^{-rT}$$
$$\approx \frac{1}{r_u T} \quad \text{for} \quad r_u \gg \frac{n}{T}$$

Posterior: A gamma distribution:

$$p(r|n, M) = \frac{T(rT)^n}{n!} e^{-rT}$$

Summaries:

- Mode $\hat{r} = \frac{n}{T}$; mean $\langle r \rangle = \frac{n+1}{T}$ (shift down 1 with $1/r$ prior)
- Std. dev'n $\sigma_r = \frac{\sqrt{n+1}}{T}$; credible regions found by integrating (can use incomplete gamma function)

The flat prior . . .

Bayes's justification: Not that ignorance of $r \rightarrow p(r|I) = C$

Require (discrete) predictive distribution to be flat:

$$p(n|I) = \int dr\; p(r|I)p(n|r, I) = C$$
$$\rightarrow \quad p(r|I) = C$$

A convention:

- Use a flat prior for a rate that may be zero
- Use a log-flat prior ($\propto 1/r$) for a nonzero scale parameter
- Use proper (normalized, bounded) priors
- Plot posterior with abscissa that makes prior flat

Inferring a Signal in a Known Background

Problem: As before, but $r = s + b$ with b known; infer s

$$p(s|n, b, M) = C \frac{T\left[(s+b)T\right]^n}{n!} e^{-(s+b)T}$$

$$\begin{aligned} C^{-1} &= \frac{e^{-bT}}{n!} \int_0^\infty d(sT)\ (s+b)^n T^n e^{-sT} \\ &= \sum_{i=0}^{n} \frac{(bT)^i}{i!} e^{-bT} \end{aligned}$$

A sum of Poisson probabilities for background events; it can be found using the incomplete gamma function.

The On/Off Problem

Basic problem:

- Look off-source; unknown background rate b
  Count $N_{\rm off}$ photons in interval $T_{\rm off}$
- Look on-source; rate is $r = s + b$ with unknown signal s
  Count $N_{\rm on}$ photons in interval $T_{\rm on}$
- Infer s

Conventional solution:

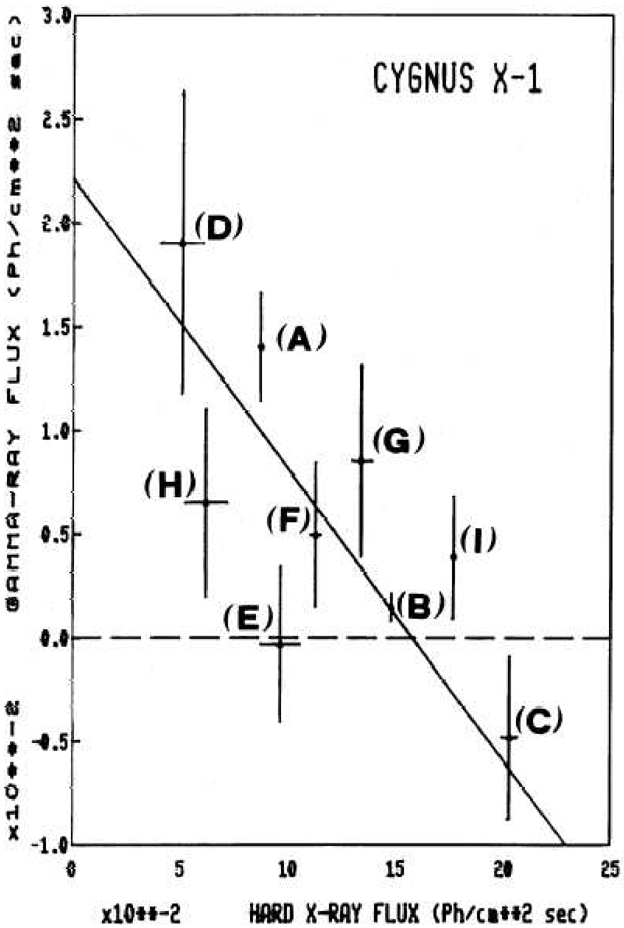
$$\begin{aligned}
\hat{b} &= N_{\rm off}/T_{\rm off}; \quad & \sigma_b &= \sqrt{N_{\rm off}}/T_{\rm off} \\
\hat{r} &= N_{\rm on}/T_{\rm on}; \quad & \sigma_r &= \sqrt{N_{\rm on}}/T_{\rm on} \\
\hat{s} &= \hat{r} - \hat{b}; \quad & \sigma_s &= \sqrt{\sigma_r^2 + \sigma_b^2}
\end{aligned}$$

But $\hat{s}$ can be negative!

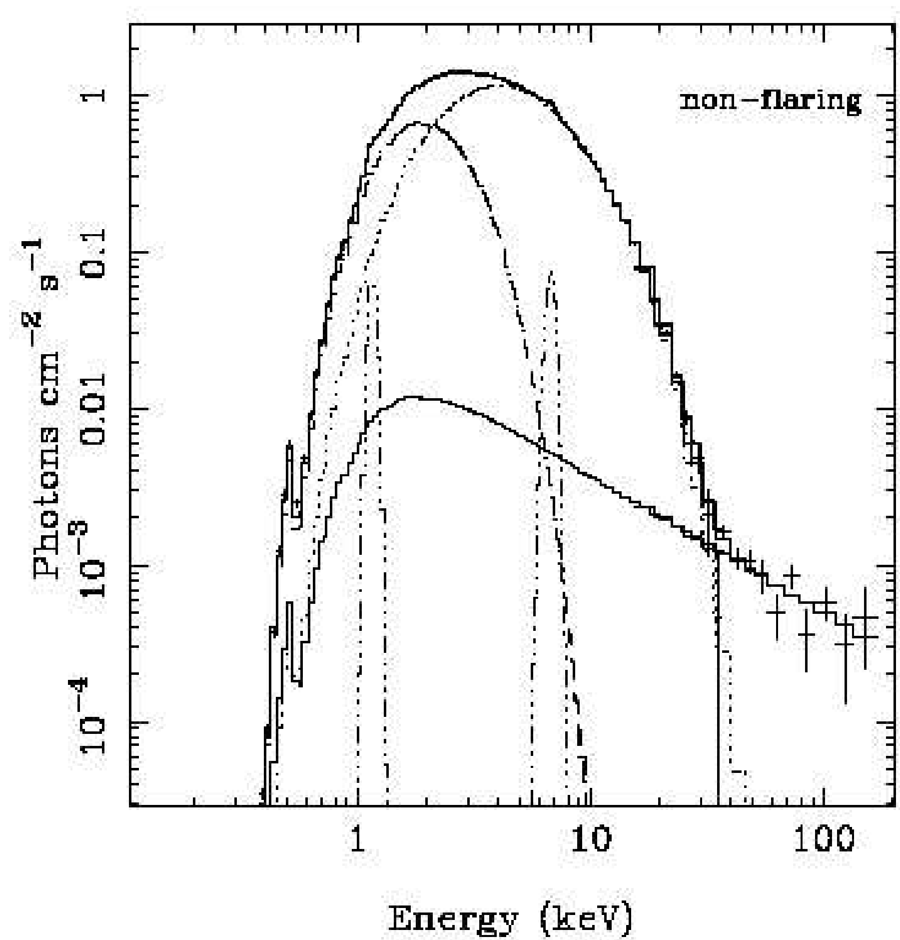
Examples

Spectra of X-Ray Sources

Bassani et al. 1989

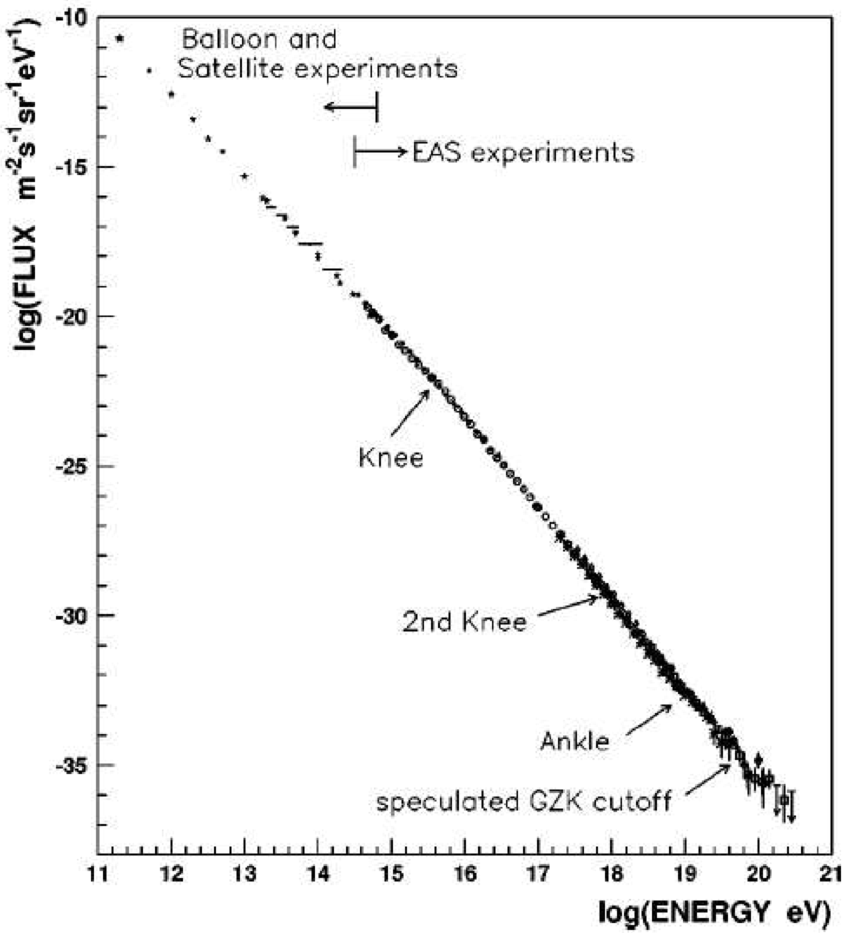


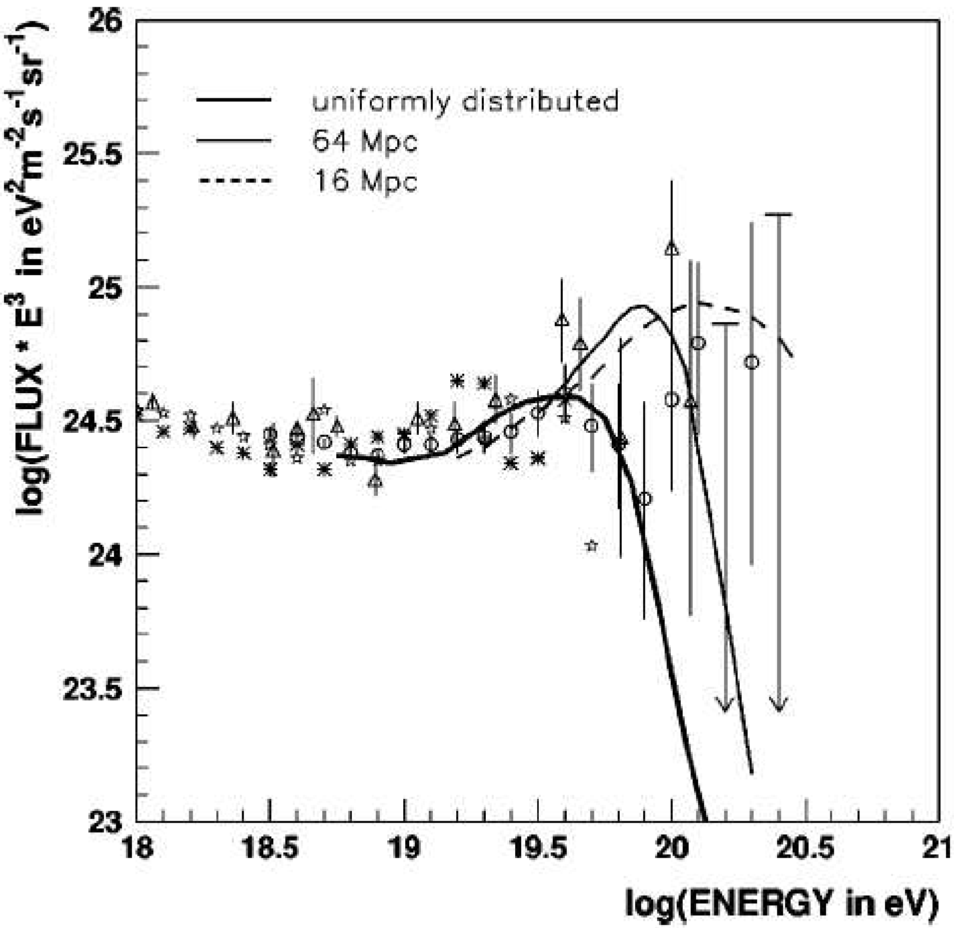
Di Salvo et al. 2001



Spectrum of Ultrahigh-Energy Cosmic Rays

Nagano & Watson 2000





“Advanced” solutions:

- Higher order approximation (Zhang and Ramsden 1990)
  But for $N_{\rm off} = 0$ and large $T_{\rm off}$, confidence region collapses to $s = 0$
- Likelihood-based methods
  Several incorrect attempts (interpret likelihood ratio as coverage; do not account for b uncertainty)

Backgrounds as Nuisance Parameters

Background marginalization with Gaussian noise:

Measure background rate $b = \hat{b} \pm \sigma_b$ with source off.

Measure total rate $r = \hat{r} \pm \sigma_r$ with source on.

Infer signal source strength s, where $r = s + b$.

With flat priors,

$$p(s, b|D, M) \propto \exp\left[-\frac{(b-\hat{b})^2}{2\sigma_b^2}\right] \times \exp\left[-\frac{(s+b-\hat{r})^2}{2\sigma_r^2}\right]$$

Marginalize b to summarize the results for s (complete the square to isolate b dependence; then do a simple Gaussian integral over b):

$$p(s|D, M) \propto \exp\left[-\frac{(s - \hat{s})^2}{2\sigma_s^2}\right] \qquad \begin{array}{l} \hat{s} = \hat{r} - \hat{b} \\ \sigma_s^2 = \sigma_r^2 + \sigma_b^2 \end{array}$$

Background *subtraction* is a special case of background *marginalization*.

Bayesian Solution to On/Off Problem

From off-source data:

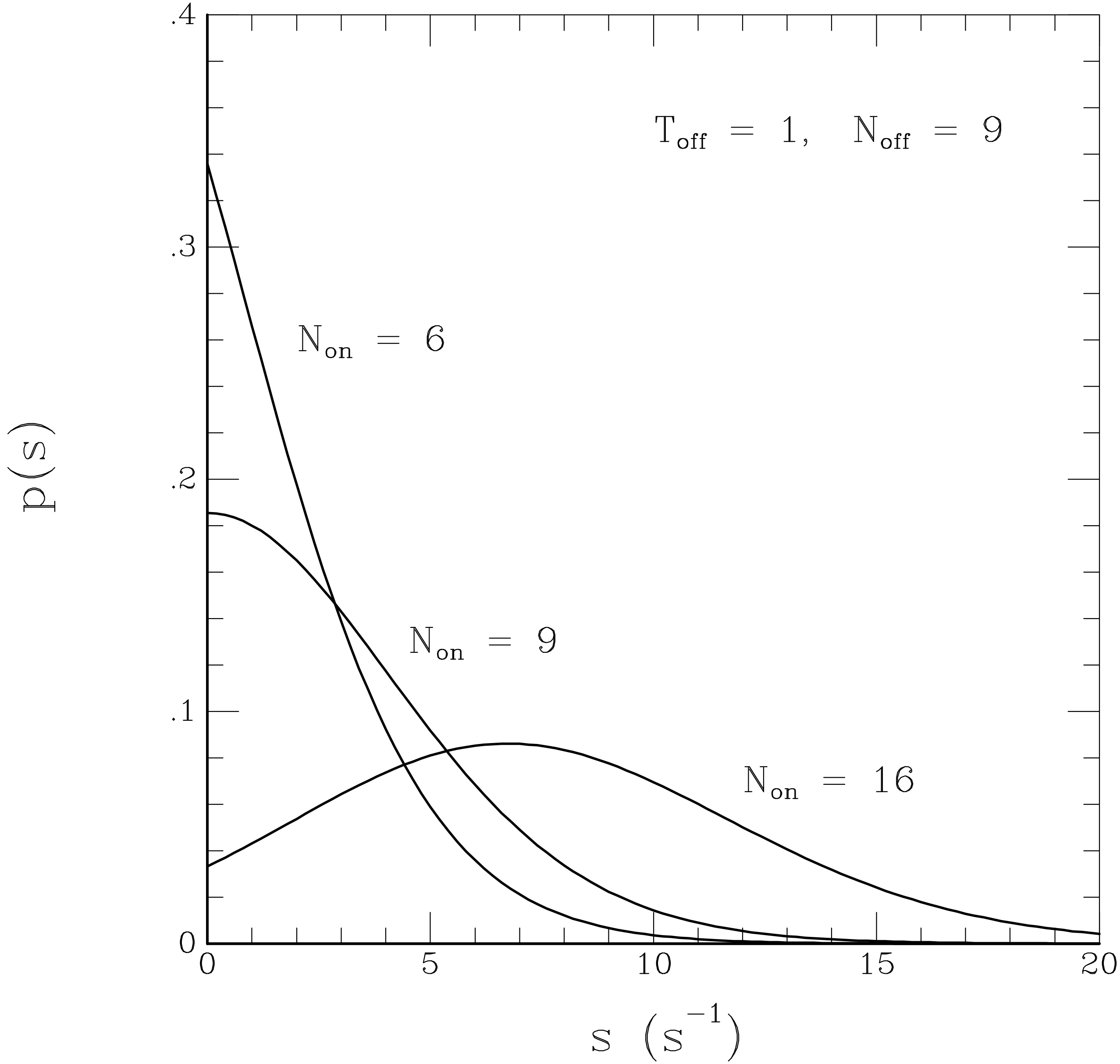
$$p(b|N_{\rm off}) = \frac{T_{\rm off}(bT_{\rm off})^{N_{\rm off}} e^{-bT_{\rm off}}}{N_{\rm off}!}$$

Use as a prior to analyze on-source data:

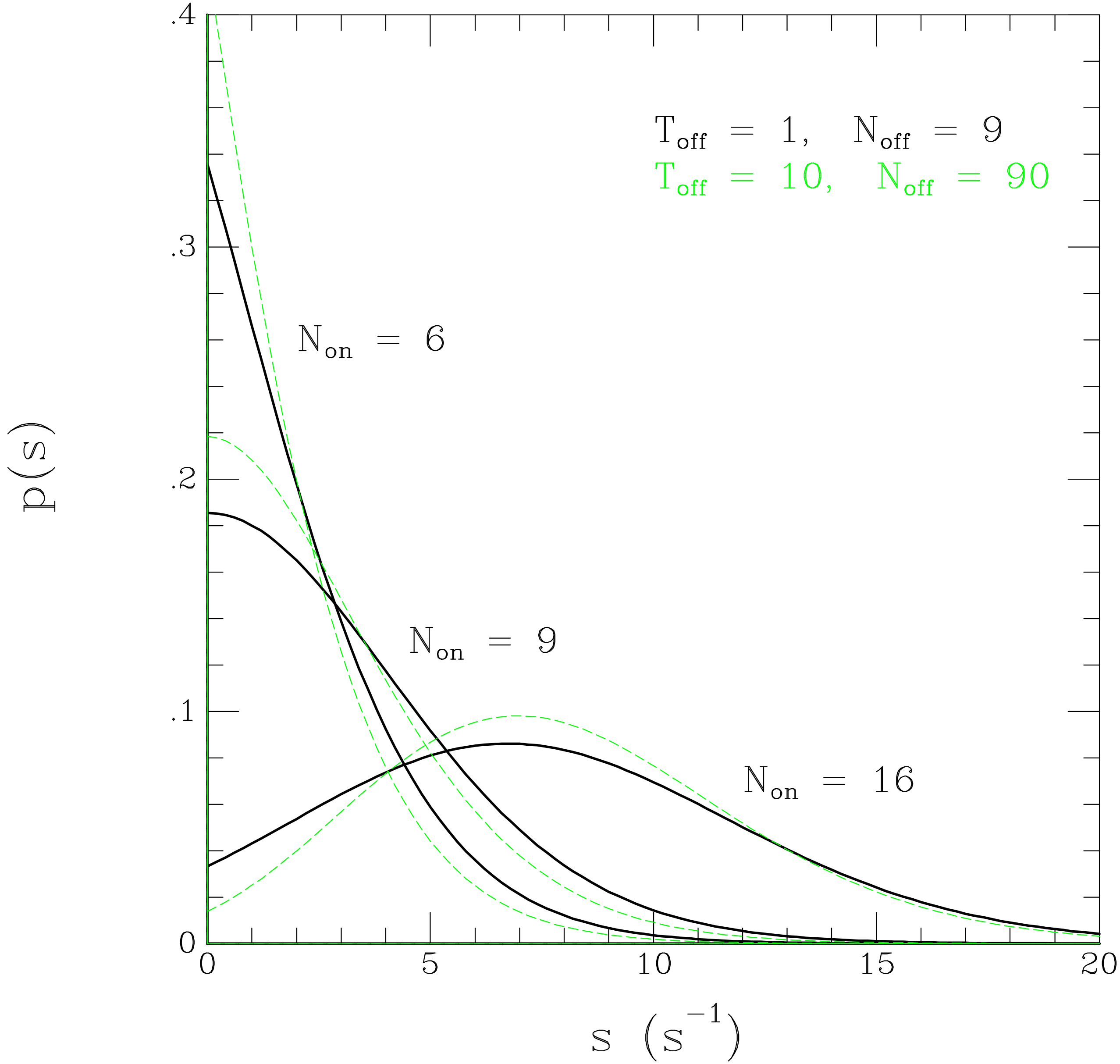
$$\begin{aligned}
p(s|N_{\rm on}, N_{\rm off}) &= \int db\, p(s, b \mid N_{\rm on}, N_{\rm off}) \\
&\propto \int db\, (s+b)^{N_{\rm on}} b^{N_{\rm off}} e^{-sT_{\rm on}} e^{-b(T_{\rm on}+T_{\rm off})} \\
&= \sum_{i=0}^{N_{\rm on}} C_i \frac{T_{\rm on}(sT_{\rm on})^i e^{-sT_{\rm on}}}{i!}
\end{aligned}$$

Can show that $C_i =$ probability that i on-source counts are indeed from the source.

Example On/Off Posteriors—Short Integrations



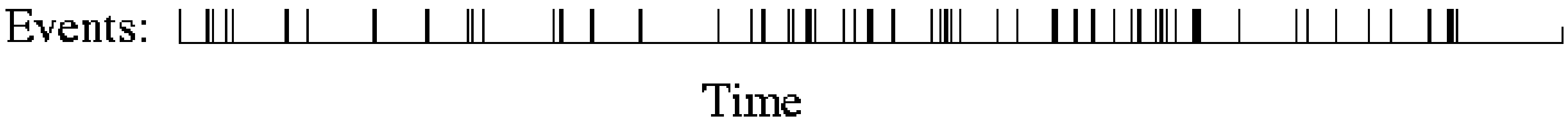
Example On/Off Posteriors—Long Background Integrations



Inhomogeneous Point Processes

Arrival Time Series

Data: Set of N arrival times $\{t_i\}$, known with small, finite resolution Δt; $N =$ dozens to millions



Goal: Detect periodicity, bursts, structure...

Conventional methods for period detection

- Binned FFT
- Rayleigh statistic

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

- Z_n^2 statistic

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

- Epoch folding
  - Fold data with trial period ($\phi_i = \omega t_i$); bin $\rightarrow n_j$, $j = 1$ to M
  - Calculate Pearson's $\chi^2(\omega)$ vs. $n_j = N/M$

Bayesian Approach

Likelihood:

$$
\begin{aligned}
p_0(t) &= P(\text{no event in } \Delta t \text{ at } t|\theta, M) \\
p_1(t) &= P(\text{one event in } \Delta t \text{ at } t|\theta, M)
\end{aligned}
$$

$$
\Rightarrow \quad p(D|\theta, M) = \prod_i p_1(t_i) \prod_{\text{empties}} p_0(t)
$$

From the Poisson dist'n,

$$
\begin{aligned}
p_0(t) &= e^{-r(t)\Delta t} \\
p_1(t) &= r(t)\Delta t \, e^{-r(t)\Delta t}
\end{aligned}
$$

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

Likelihood for periodic models:

Rate $=$ avg. rate A $\times$ periodic shape $\rho(\phi)$ (params $\mathcal{S}$)

$$r(t) = A\rho(\omega t - \phi; \mathcal{S})$$

Inhom. point process likelihood (for $T \gg$ period)

$$\mathcal{L}(A, \omega, \phi, \mathcal{S}) = \left[A^N e^{-AT}\right] \prod_i \rho(\omega t_i - \phi; \mathcal{S})$$

Marginal likelihood for ω, ϕ, $\mathcal{S}$

$$\mathcal{L}(\omega, \phi, \mathcal{S}) = \prod_i \rho(\omega t_i - \phi; \mathcal{S})$$

Example models:

- Log-Fourier models—analytic ϕ marginalization

$$\log \rho(\theta) \propto \kappa \cos(\theta) \quad \rightarrow \quad \mathcal{L} \propto I_0 \left[\kappa N R(\omega)\right] / I_0^N(\kappa)$$

$$\text{Harmonic sum} \quad \rightarrow \quad Z_n^2 + \text{interference terms}$$

- Piecewise constant models—analytic $\mathcal{S}$ marginalization

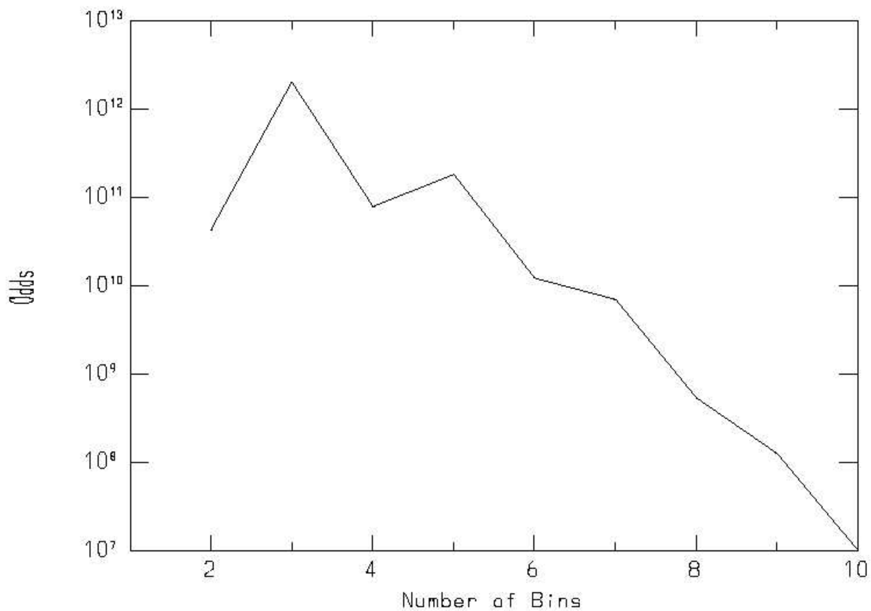
$$\rho \text{ flat in } M \text{ bins} \quad \rightarrow \quad \mathcal{L} \propto \frac{(M-1)!}{(N+M-1)!} \left[\frac{n_1!\, n_2!\, ...n_M!}{N!}\right]$$

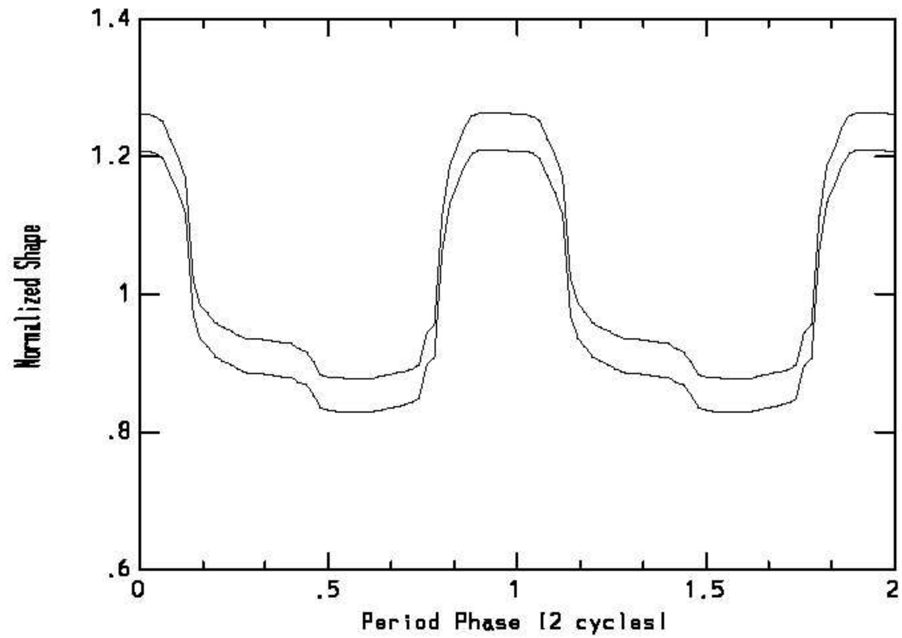
For signal detection, *integrate* over ω, rather than maximize over a grid. This removes ambiguity/subjectivity from conventional approach.

Piecewise Constant Modeling of X-Ray Pulsar

X-Ray Pulsar PSR 0540-693 (Gregory & Loredo 1996)

3300 events over 10^5 s, many gaps, FFT fails

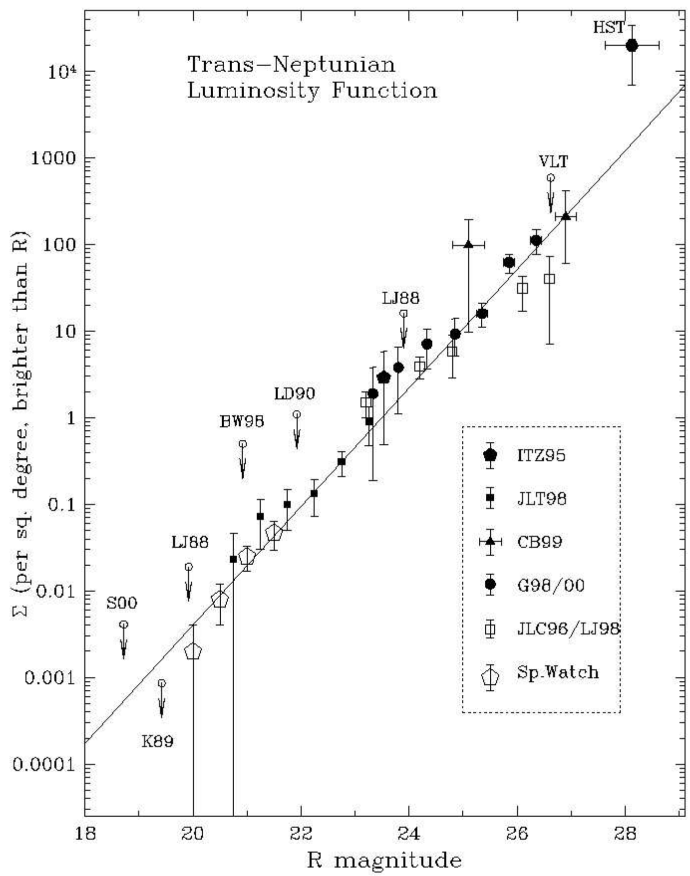




Point Processes With Error

Population Studies

Multiple searches for Trans-Neptunian Objects report $\{R_i, \sigma_i\}$ or non-detections.
What are the sizes of TNOs? How far out does the pop'n extend?



Phenomenology

Cumulative dist'n $\Sigma(R) = 10^{\alpha(R-R_0)}$, params α, R_0
Differential dist'n $\sigma(R) = d\Sigma/dR$

Physics

Size dist'n $f(D)$ and radial dist'n $n(r)$

Visible via reflection $\rightarrow$ calculate R from D^2/r^4 law

Conventional analyses

Least squares or χ^2 fit to binned *cumulative* dist'n

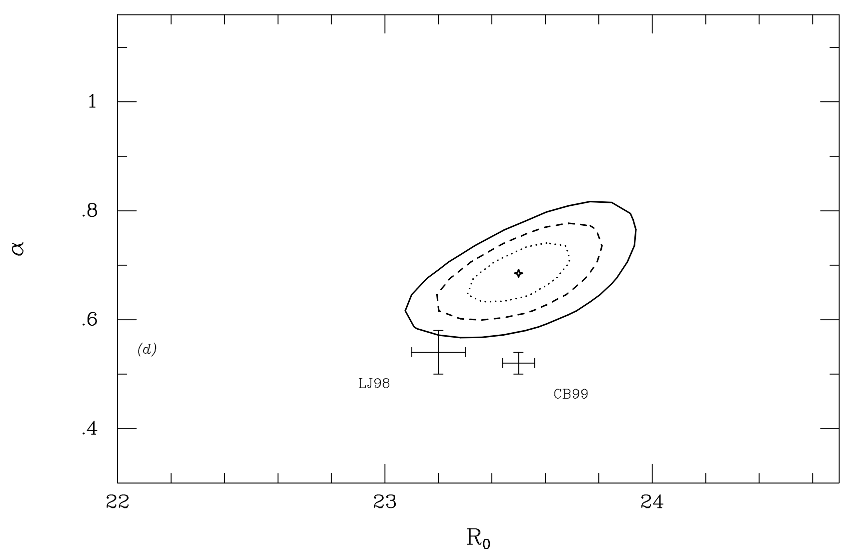
Ignores uncertainties; ambiguity in correcting for sampling; difficulty handling nondetections; difficulty combining disparate types of data; arbitrary, correlated bins

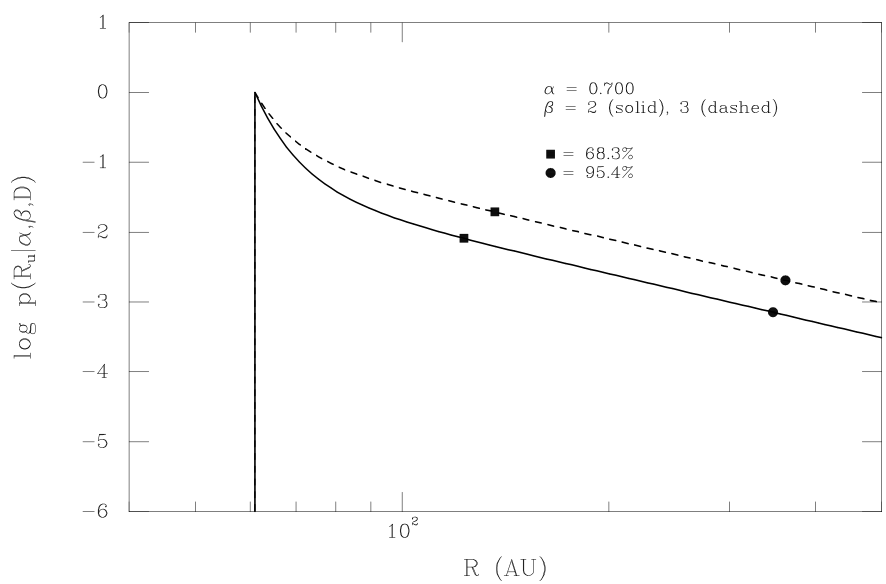
Bayesian approach

Multiply likelihoods for each survey modeled as point process with error,

$$\mathcal{L}(\theta) = \exp\left[-\Omega \int dR\, \eta(R)\sigma(R)\right] \prod_i \int dR\, \ell_i(R)\sigma(R)$$

A point process likelihood, including detection efficiency, $\eta(R)$, and object uncertainties, $\ell_i(R) = p(d_i|R)$.



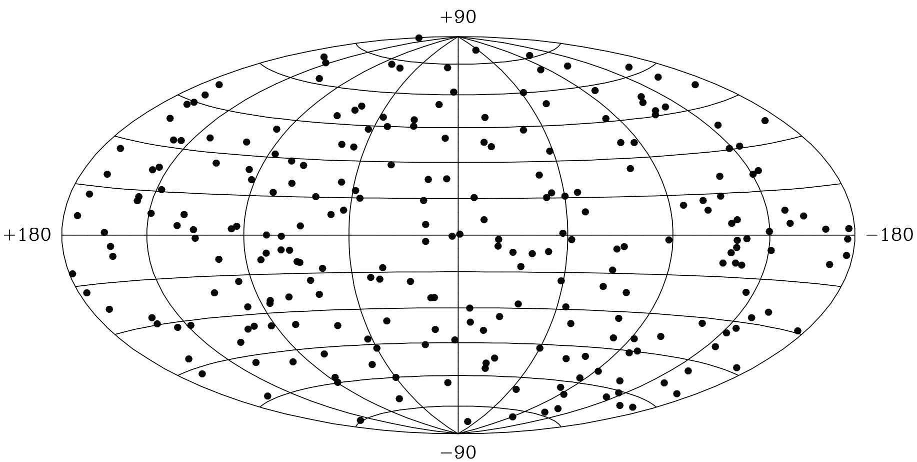


Gladman et al. 1998, 2001

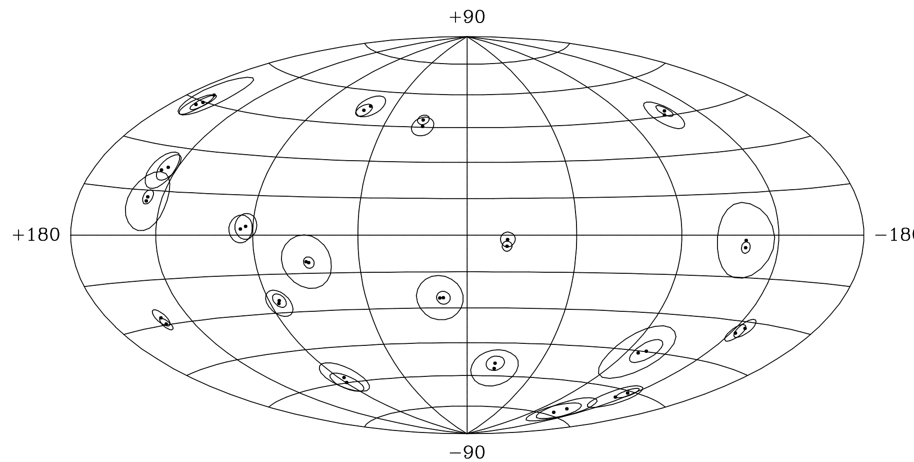
Spatio-Temporal Coincidences

Do GRB sources repeat?

250 GRB directions



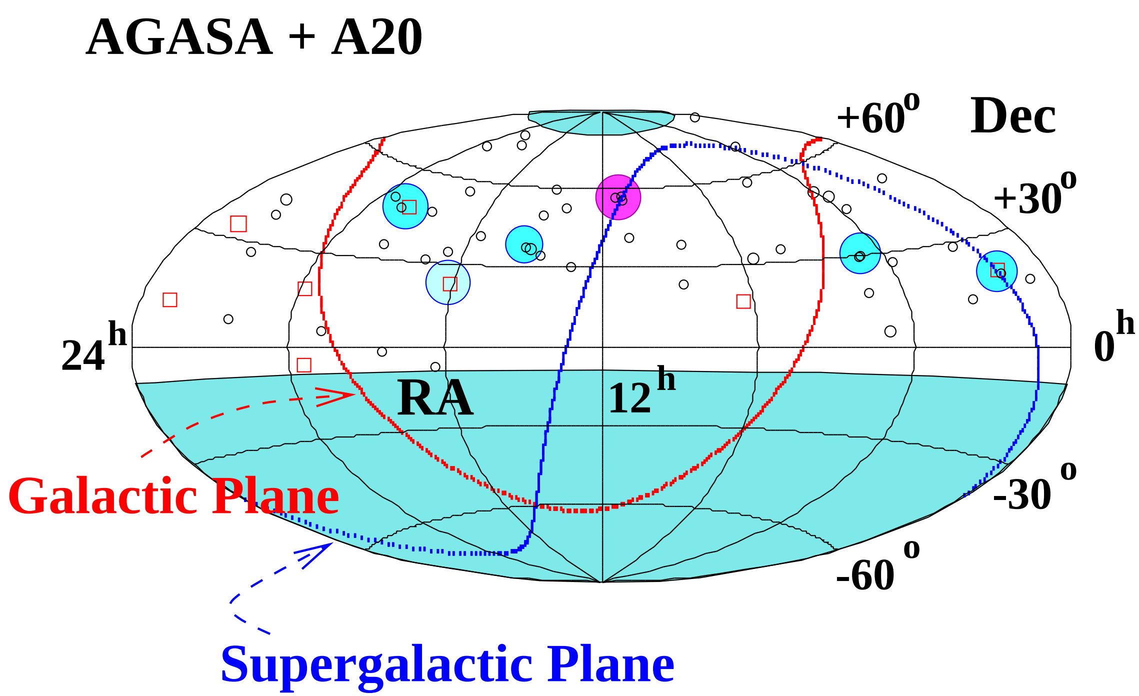
Subset with neighbor within 3° (39)



If GRBs repeat, many existing models are ruled out!

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

Frequentist nearest neighbor analysis—two objects:

Null hypothesis H_0: no repetition, isotropic source dist'n

Statistic: Angle to nearest neighbor, θ_{12}

Sampling Dist'n:

$$
\begin{aligned}
p(\cos\theta_{12}, \phi_{12}) &= \frac{1}{4\pi}, \quad \textit{independent of uncertainty} \\
\rightarrow p(\theta_{12}) &= \frac{\sin\theta_{12}}{2} \\
p(<\theta_{12}) &= \frac{1-\cos\theta_{12}}{2}
\end{aligned}
$$

Reject H_0 if this probability is small; e.g.:

- $\theta_{12} = 26^\circ \quad \rightarrow \quad p(< 26^\circ) = 0.05$
- $\theta_{12} = 0^\circ \quad \rightarrow \quad p(< 0^\circ) = 0$

Bayesian coincidence assessment—two objects:

Direction uncertainties accounted for via likelihoods for object directions:

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

H_0: No repetition

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

H_1: Repeating (same direction!)

$$p(d_1, d_2|H_0) = \int d\mathbf{n}\, p(\mathbf{n}|H_0)\, \mathcal{L}_1(\mathbf{n})\, \mathcal{L}_2(\mathbf{n})$$

Odds favoring repetition:

$$O = 4\pi \int d\mathbf{n}\, \mathcal{L}_1(\mathbf{n})\, \mathcal{L}_2(\mathbf{n})$$

$$\approx \frac{2}{\sigma_{12}^2} \exp\left[-\frac{\theta_{12}^2}{2\sigma_{12}^2}\right]; \quad \sigma_{12}^2 = \sigma_1^2 + \sigma_2^2$$

E.g.: $\sigma_1 = \sigma_2 = 10^\circ$ $\quad O \approx 6$ for $\theta_{12} = 26^\circ$

$O \approx 33$ for $\theta_{12} = 0^\circ$

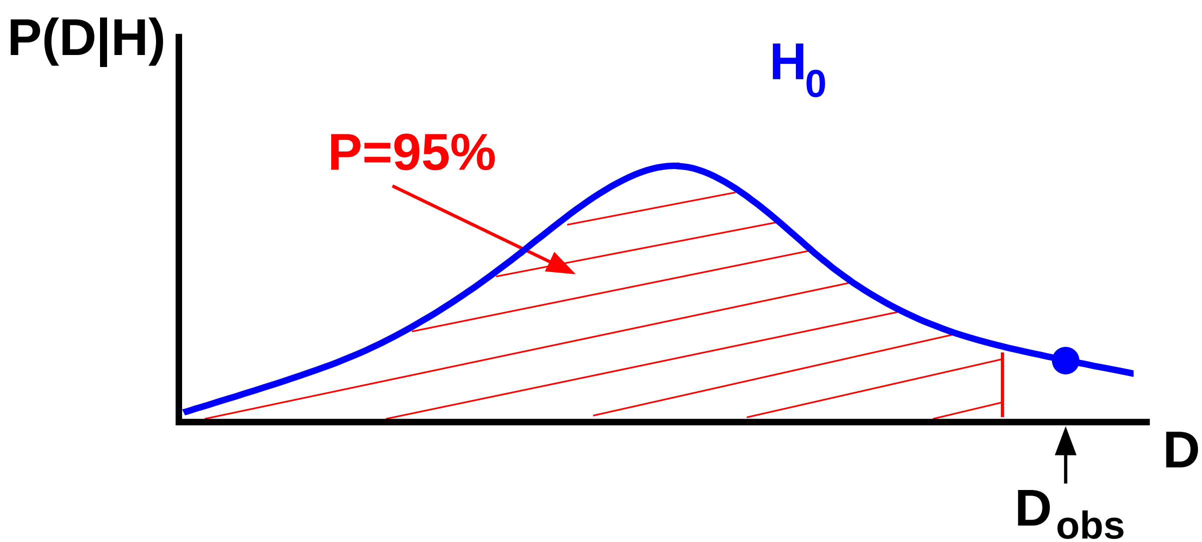
$\sigma_1 = \sigma_2 = 20^\circ$ $\quad O \approx 5$ for $\theta_{12} = 26^\circ$

$O \approx 8$ for $\theta_{12} = 0^\circ$

Compare or Reject Hypotheses?

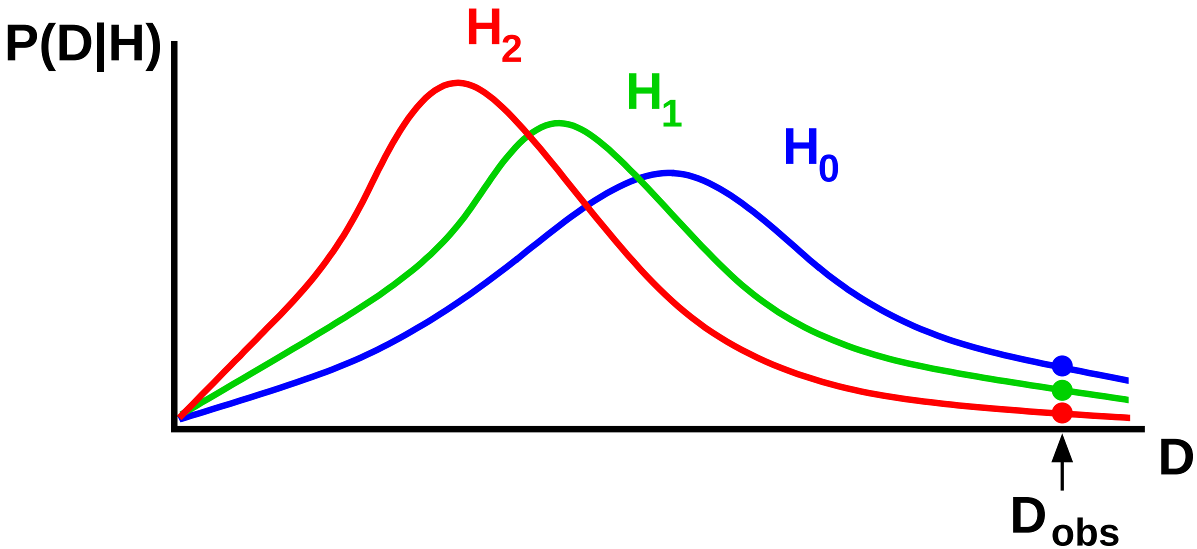
Frequentist Significance Testing (G.O.F. tests):

- Specify simple null hypothesis H_0 such that rejecting it implies an interesting effect is present
- Divide sample space into probable and improbable parts (for H_0)
- If $D_{\rm obs}$ lies in improbable region, reject H_0; otherwise accept it



Bayesian Model Comparison:

- 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. GOF tests reject the latter possibility at the outset.

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

For N events, # of hypotheses with n_k k-tuplets (n_2 doublets, n_3 triplets...)

$$\mathcal{N} = \frac{N!}{\prod_{k=1}^{K} (k!)^{n_k} n_k!}$$

E.g. for $n_2 = 2$, $\mathcal{N} \approx N^4/8$.

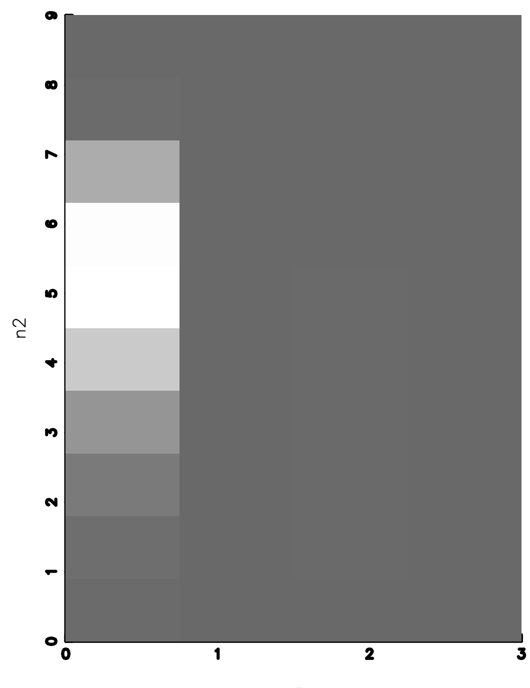
Bayesian Analysis of AGASA Cosmic Rays

M_0: $N = 58$ different directions

M_1: Unknown number of pairs (n_2) and triplets (n_3)

$\rightarrow O_{10} = 1.4$ favoring clusters (i.e., no significant evidence)

If indeed clusters are present, we can constrain the number by calculating $p(n_2, n_3|D, M_1)$:



Key Ideas

Poisson processes handled without approximation

- Counting processes:
  - Can treat rigorously for any n
  - Backgrounds handled straightforwardly
- Point processes: No binning necessary!
- Point processes with error: Uncertainties easily handled