Learning How To Count:
Inference With Poisson Process Models

(Lecture 2)

Tom Loredo

Dept. of Astronomy, Cornell University

Motivation

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

Terminology

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

If our measurements report the *locations* of every individual event, the appropriate model is the *Poisson point process*. If the individual event locations are measured with uncertainty, it is a *point process with measurement error*.

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

Lecture 2

- Poisson Process Fundamentals
- Poisson counting processes
- Poisson point processes

Lecture 2

- Poisson Process Fundamentals
- Poisson counting processes
- Poisson point processes

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, separate intervals. E.g.,

- Q_1: 3 events occured in $[t_1, t_2]$
- Q_2: No events occured in $[t_3, t_4]$
- $Q_3 = (Q_1, Q_2)$

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:

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

  “Small” means $dt:\ r(t)dt \ll 1$.

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

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

Homog. 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:

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

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

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

$$\Rightarrow h(t) = e^{-rt}$$

One event:

$B =$ "One event is seen in $[0, T]$, within $[t_1, t_1 + dt_1]$"

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

Probability the event is seen *somewhere* in $[0, T]$:

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

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:

$$\mathsf{E}(n) = \langle n \rangle \equiv \sum_{n=0}^{\infty} n\, p(n|r, M) = rT \equiv \mu$$

$$\mathsf{Var}(n) = \langle (n - \bar{n})^2 \rangle = \mu$$

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

$\mu = \mathsf{E}(n)$ specifies both the mean and standard deviation.

Poisson Approximation of Binomial

Recall the binomial distribution for n successes in N trials:

$$p(n|a, M) = \frac{N!}{n!(N-n)!} a^n (1-a)^{N-n}$$

Its moments are:

$$\mathsf{E}(n) = Na \equiv \mu$$
$$\mathsf{Var}(n) = Na(1-a)$$

Note for $a \ll 1$, Var$(n) \approx Na =$ E(n). Can show that when $a \ll 1$ and $N \gg 1$,

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

Lecture 2

- Poisson Process Fundamentals
- Poisson counting processes
- Poisson point processes

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:

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

Prior predictive:

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

Posterior: A gamma distribution:

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

Gamma Distributions

Definition

A 2-parameter distribution over nonnegative x, with shape parameter α and scale parameter s:

$$p_\Gamma(x|\alpha, s) \;=\; \frac{1}{s\Gamma(\alpha)}\left(\frac{x}{s}\right)^{\alpha-1} e^{-x/s}$$

Moments:

$$\text{E}(x) = s\nu \qquad \text{Var}(x) = s^2\nu$$

Our posterior corresponds to $\alpha = n + 1$, $s = 1/T$.

- 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 $r \to p(r|I) = C$

Require (discrete) predictive distribution to be flat:

$$\begin{aligned} p(n|I) &= \int dr\; p(r|I)p(n|r,I) = C \\ &\to p(r|I) = C \end{aligned}$$

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 evaluated using the incomplete gamma function. (Helene 1983)

Alternate Approach: “Missing Data”

General idea: Pretend you know something more about the data that would simplify analysis—“missing data.” Then account for uncertainty of missing data via LTP.

Recall $\mathcal{L}(s) = p(n|s, b, M)$ which we assigned via Poisson with $r = s + b$. An alternative way is to note that $n = n_s + n_b$, and use n_b to extend the conversation:

$$p(n|s,b) = \sum_{n_b=0}^{n} p(n|n_b, s, b)\, p(n_b|s, b) \qquad || \; M$$

1st factor is Poisson with $\mu = sT$ and $n_s = n - n_b$ counts; 2nd factor is Poisson with $\mu = bT$ and n_b counts.

Thus:

$$
\begin{aligned}
p(n|s,b,M) &= e^{-(s+b)T} \sum_{n_b=0}^{n} \frac{1}{(n-n_b)!\,n_b!} (sT)^{n-n_b} (bT)^{n_b} \\
&= \frac{1}{n!} [(s+b)T]^n e^{-(s+b)T}
\end{aligned}
$$

$\longrightarrow$ same result as previously

In more complicated problems, “data augmentation” may be the simplest way to get the desired result. It can be straightforwardly implemented via Monte Carlo (“imputation”), or via a famous iterative algorithm (EM algorithm).

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:

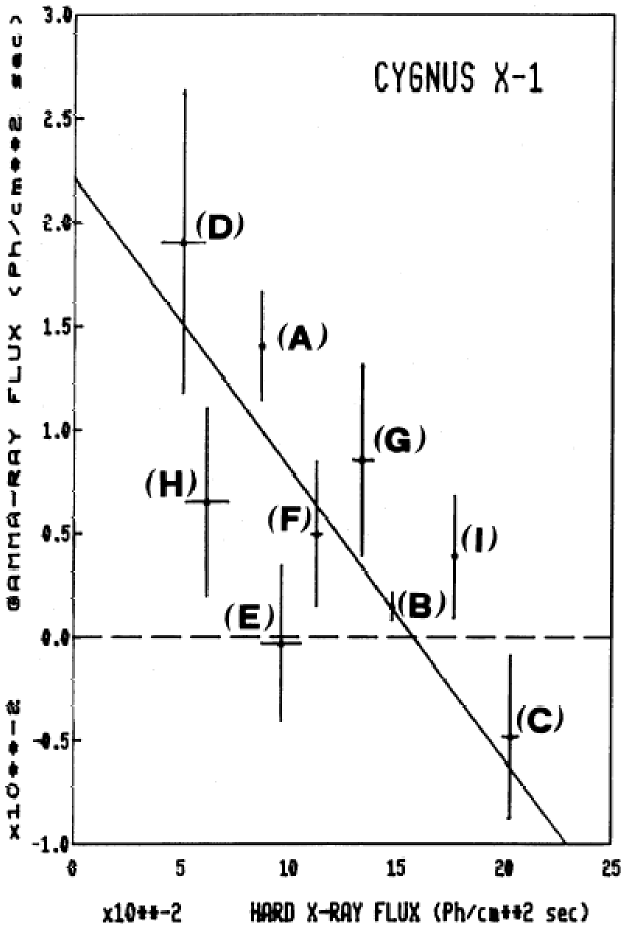
$$\hat{b} = N_{\rm off}/T_{\rm off}; \quad \sigma_b = \sqrt{N_{\rm off}}/T_{\rm off}$$
$$\hat{r} = N_{\rm on}/T_{\rm on}; \quad \sigma_r = \sqrt{N_{\rm on}}/T_{\rm on}$$
$$\hat{s} = \hat{r} - \hat{b}; \qquad \sigma_s = \sqrt{\sigma_r^2 + \sigma_b^2}$$

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

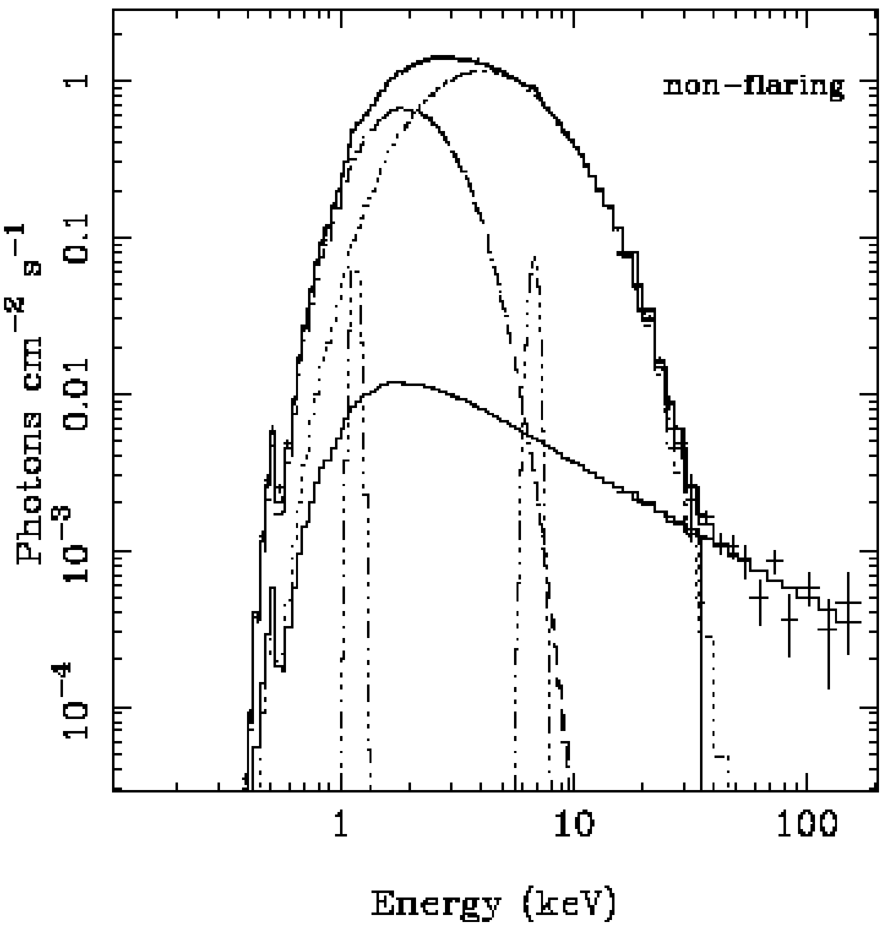
Examples

Spectra of X-Ray Sources

Bassani et al. 1989

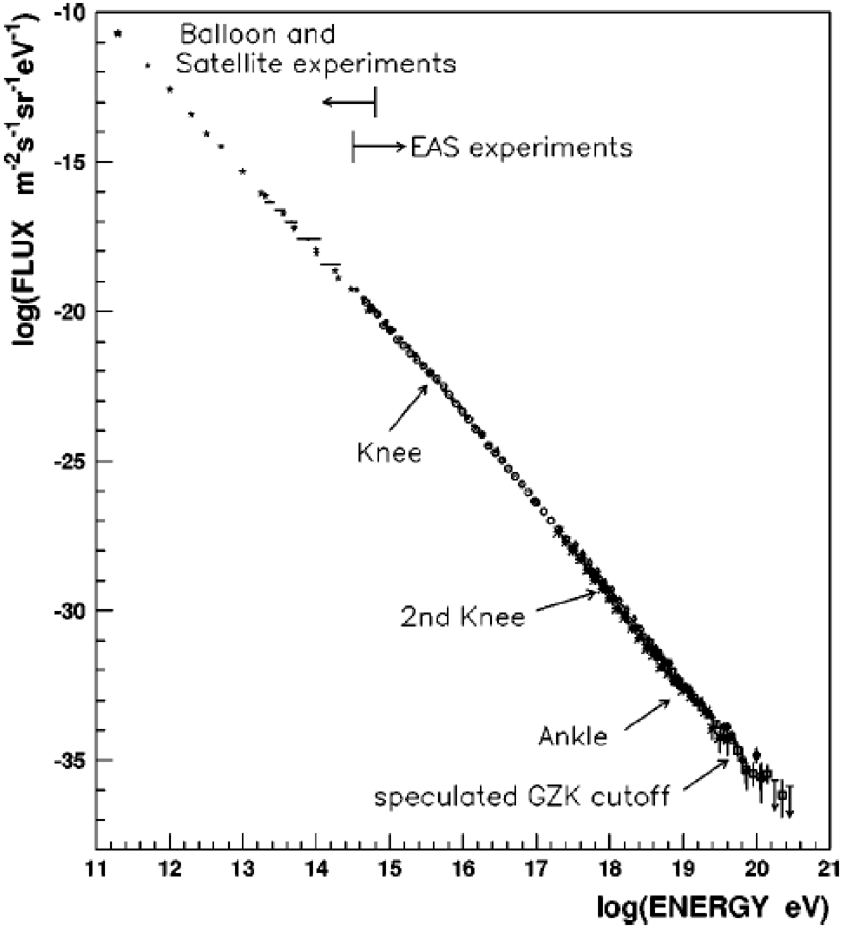


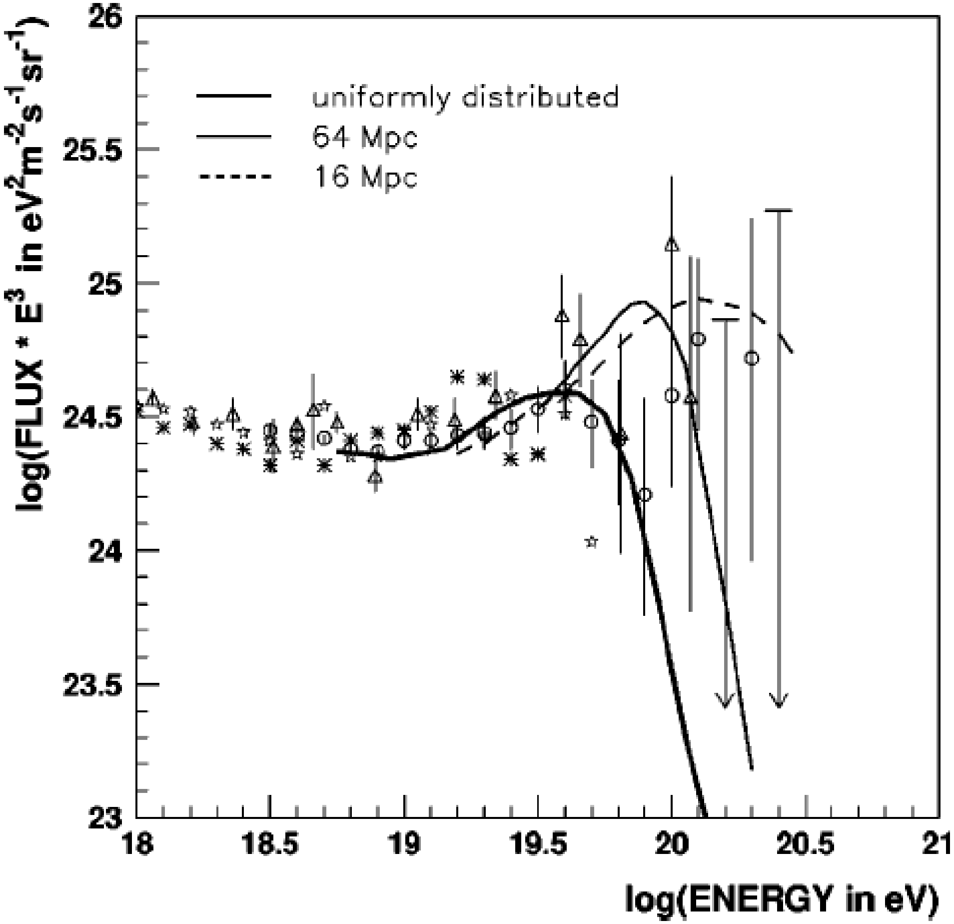
Di Salvo et al. 2001



Spectrum of Ultrahigh-Energy Cosmic Rays

Nagano & Watson 2000





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):

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

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

A Bayesian Solution to On/Off Problem

First consider off-source data; use it to estimate b:

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

Use this as a prior for b to analyze on-source data. For on-source analysis $I_{\rm all} = (I_{\rm on}, N_{\rm off}, I_{\rm off})$:

$$p(s, b|N_{\rm on}) \;\; \propto \;\; p(s)p(b)[(s+b)T_{\rm on}]^{N_{\rm on}} e^{-(s+b)T_{\rm on}} \qquad || \; I_{\rm all}$$

$p(s|I_{\rm all})$ is flat, but $p(b|I_{\rm all}) = p(b|N_{\rm off}, I_{\rm off})$, so

$$p(s, b|N_{\rm on}, I_{\rm all}) \;\; \propto \;\; (s+b)^{N_{\rm on}} b^{N_{\rm off}} e^{-sT_{\rm on}} e^{-b(T_{\rm on}+T_{\rm off})}$$

Now marginalize over b;

$$\begin{aligned} p(s|N_{\rm on}, I_{\rm all}) &= \int db\, p(s, b \mid N_{\rm on}, I_{\rm on}) \\ &\propto \int db\, (s+b)^{N_{\rm on}} b^{N_{\rm off}} e^{-sT_{\rm on}} e^{-b(T_{\rm on}+T_{\rm off})} \end{aligned}$$

Expand $(s+b)^{N_{\rm on}}$ and do the resulting Γ integrals:

$$\begin{aligned} p(s|N_{\rm on}, I_{\rm all}) &= \sum_{i=0}^{N_{\rm on}} C_i \frac{T_{\rm on}(sT_{\rm on})^i e^{-sT_{\rm on}}}{i!} \\ C_i &\propto \left(1 + \frac{T_{\rm off}}{T_{\rm on}}\right)^i \frac{(N_{\rm on} + N_{\rm off} - i)!}{(N_{\rm on} - i)!} \end{aligned}$$

Posterior is a weighted sum of Gamma distributions, each assigning a different number of on-source counts to the source. (Evaluate via recursive algorithm or confluent hypergeometric function.)

Second Solution to On/Off Problem

Consider all the data at once; the likelihood is a product of Poisson distributions for the on- and off-source counts:

$$\begin{aligned}\mathcal{L}(s,b) &\equiv p(N_{\rm on}, N_{\rm off}|s,b,I) \\ &\propto [(s+b)T_{\rm on}]^{N_{\rm on}} e^{-(s+b)T_{\rm on}} \times (bT_{\rm off})^{N_{\rm off}} e^{-bT_{\rm off}}\end{aligned}$$

Take joint prior to be flat; find the joint posterior and marginalize over b;

$$\begin{aligned}p(s|N_{\rm on}, I_{\rm on}) &= \int db\, p(s,b|I)\, \mathcal{L}(s,b) \\ &\propto \int db\, (s+b)^{N_{\rm on}} b^{N_{\rm off}} e^{-sT_{\rm on}} e^{-b(T_{\rm on}+T_{\rm off})}\end{aligned}$$

$\rightarrow$ same result as before.

Solution Via "Missing Data"

First consider the off-source data as in 1st approach $\rightarrow$ $p(b|N_{\rm off}, I_{\rm off})$.

Now consider on-source data. The likelihood for s is

$$\mathcal{L}(s) \equiv p(N_{\rm on}|s, I_{\rm all})$$

Note that $N_{\rm on} = n_s + n_b$; if only we knew n_s, we could just use $p(n_s|s, I_{\rm all})$ (Poisson!). Augment & marginalize:

$$p(N_{\rm on}|s) = \sum_{n_s=0}^{N_{\rm on}} p(N_{\rm on}|n_s, s)\, p(n_s|s) \qquad || \; I_{\rm all}$$

Last factor is just $\propto (sT)^{n_s} e^{-sT_{\rm on}}$.

First factor is probability that $n_b = N_{\rm on} - n_s$ counts are due to the background, given what we know about the background (s is irrelevant if we're given n_s).

Thus

$$
\begin{aligned}
p(N_{\rm on}|n_s, s) &= p(n_b = N_{\rm on} - n_s) \\
&= \int db\, p(n_b = N_{\rm on} - n_s|b)\, p(b) \qquad || \; I_{\rm all}
\end{aligned}
$$

where $p(b|I_{\rm all}) = p(b|N_{\rm off}, I_{\rm off})$. The first factor is

$$
p(n_b = N_{\rm on} - n_s|b, I_{\rm all}) \propto (bT_{\rm on})^{N_{\rm on}-n_s} e^{-bT_{\rm on}}
$$

Do the integral and you discover,

$$
p(N_{\rm on}|n_s, s, I_{\rm all}) = C_{n_s}
$$

i.e., C_i from the first approach, with $i = n_s$. Thus C_i is the probability that i on-source counts are due to the source (as we might have guessed).

A Profound Consistency

We solved both the known-b and on/off problems in multiple ways, always finding the same final results.

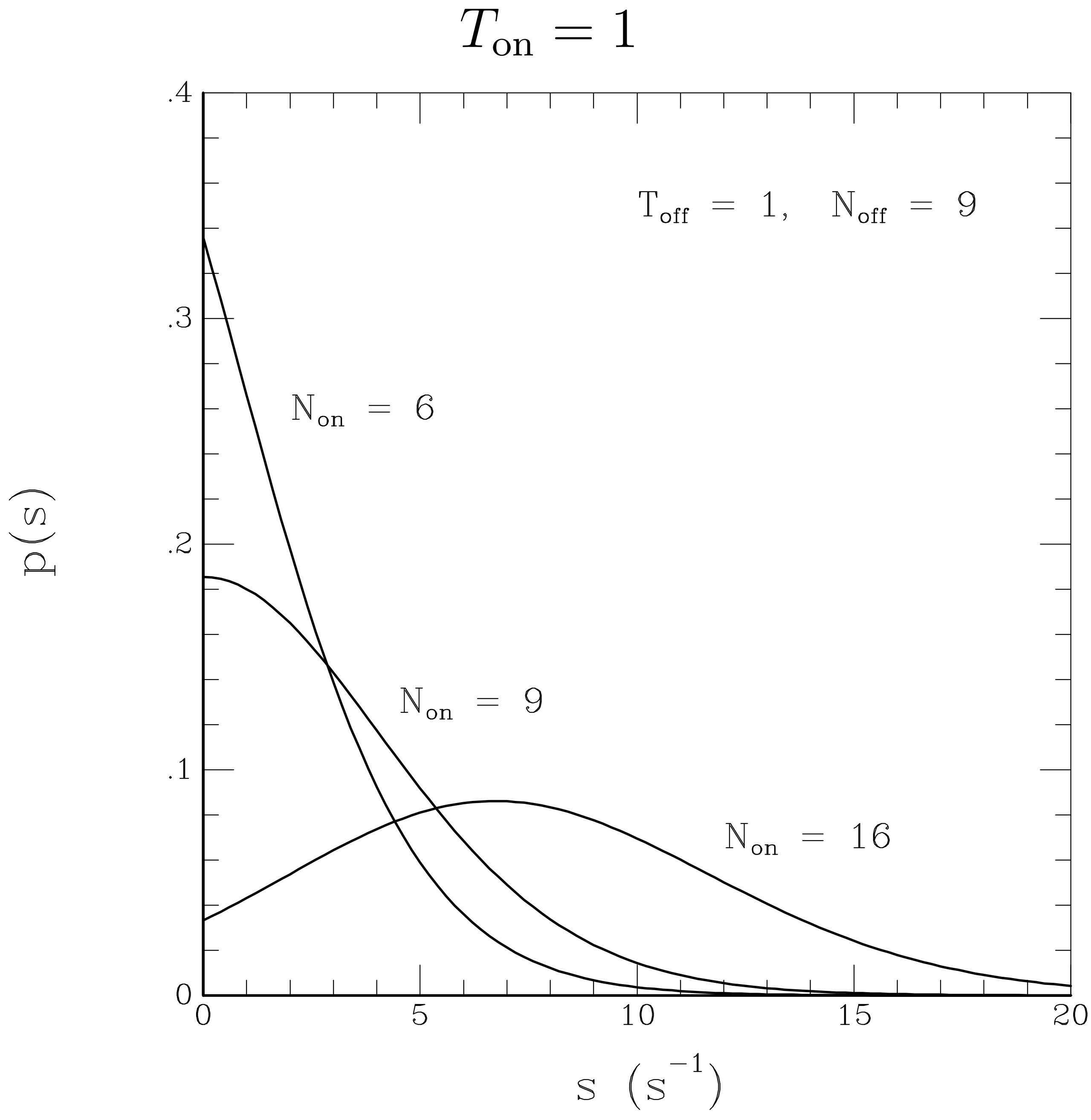
This reflects something fundamental about Bayesian inference.

R. T. Cox proposed two necessary conditions for a quantification of uncertainty:

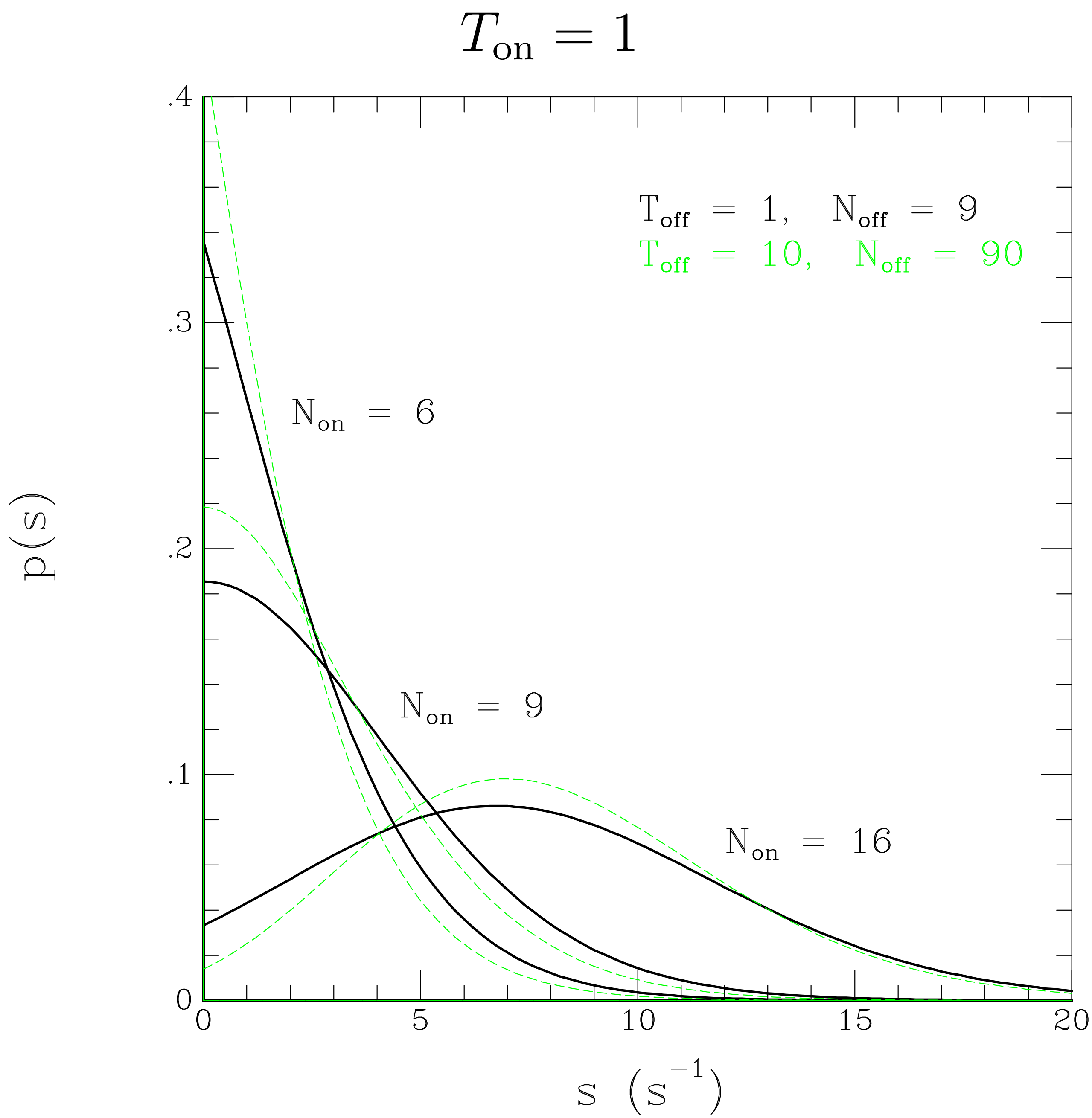
- It should duplicate deductive logic when there is no uncertainty
- Different decompositions of arguments should produce the same final quantifications (internal consistency)

Great surprise: These conditions are *sufficient*; they lead to the probability axioms. Jaynes and others refined and simplified Cox's analysis.

Example On/Off Posteriors—Short Integrations



Example On/Off Posteriors—Long Background Integrations



Multibin On/Off

The more typical on/off scenario:

Data $=$ spectrum or image with counts in many bins

Model M gives signal rate $s_k(\theta)$ in bin k, parameters θ

To infer θ, we need the likelihood:

$$\mathcal{L}(\theta) = \prod_k p(N_{\mathrm{on}k}, N_{\mathrm{off}k} | s_k(\theta), M)$$

For each k, we have an on/off problem as before, only we just need the marginal likelihood for s_k (not the posterior). The same C_i coefficients arise.

XSPEC and CIAO/Sherpa provide this as an option.

CHASC approach does the same thing via data augmentation.

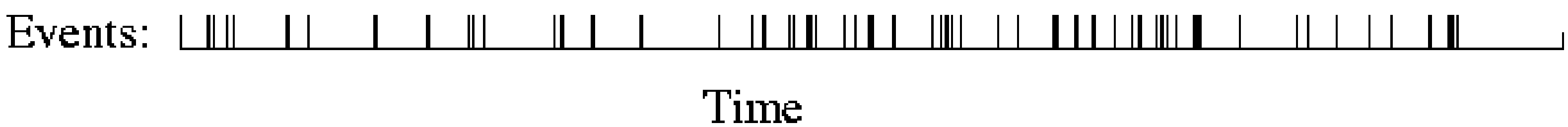
Lecture 2

- Poisson Process Fundamentals
- Poisson counting processes
- Poisson point processes

Inhomogeneous Point Processes

Example: Arrival Time Series

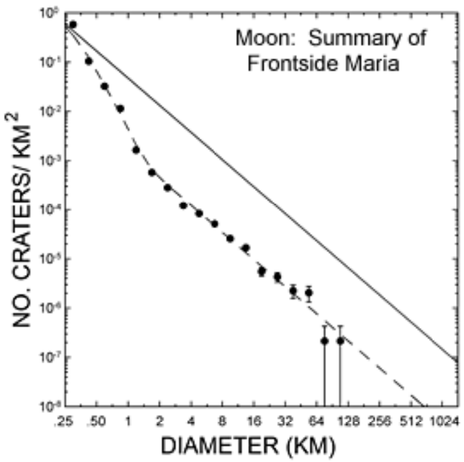
Data: Set of N arrival times $\{t_i\}$, known with small, finite resolution Δt; $N =$ dozens to millions



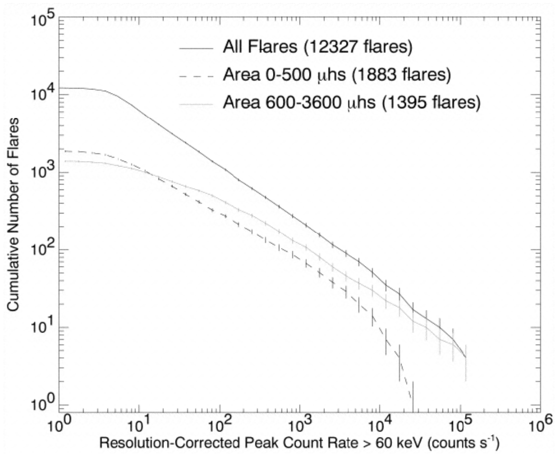
Goals: Detect & measure periodicity, bursts, structure. . .

Example: Size-Frequency/Number-Size Dist'ns

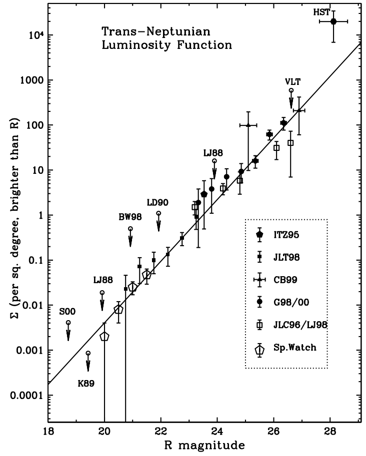
Lunar Craters



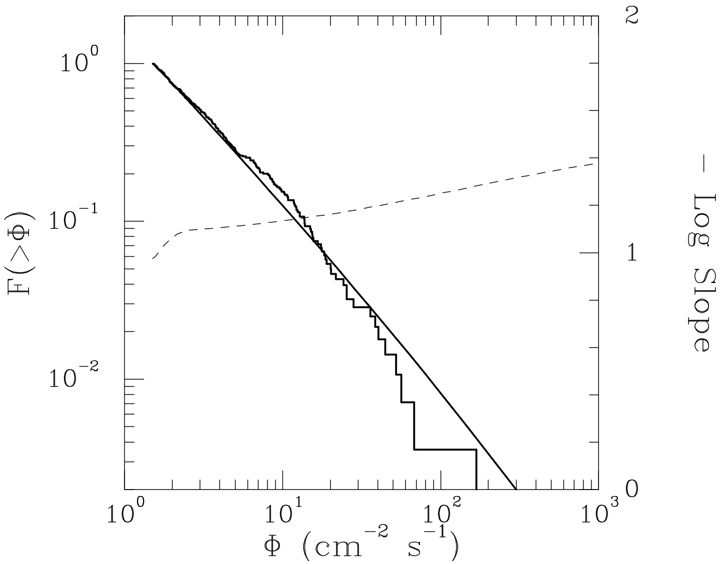
Solar Flares



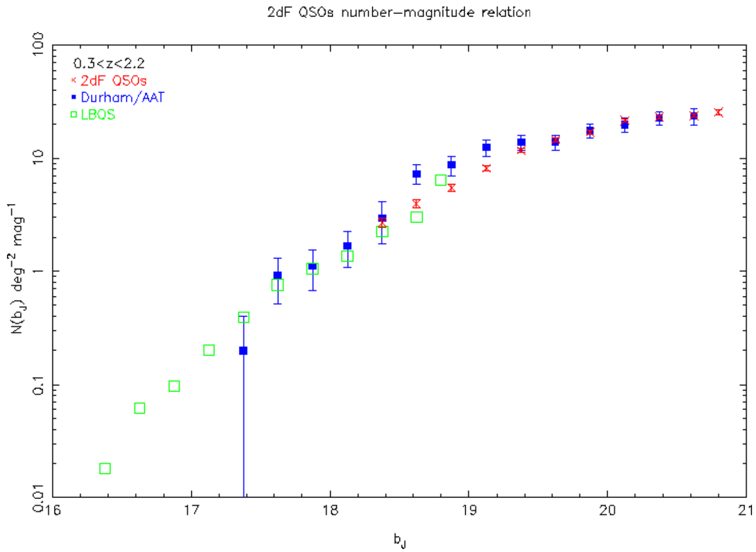
TNOs



GRBs



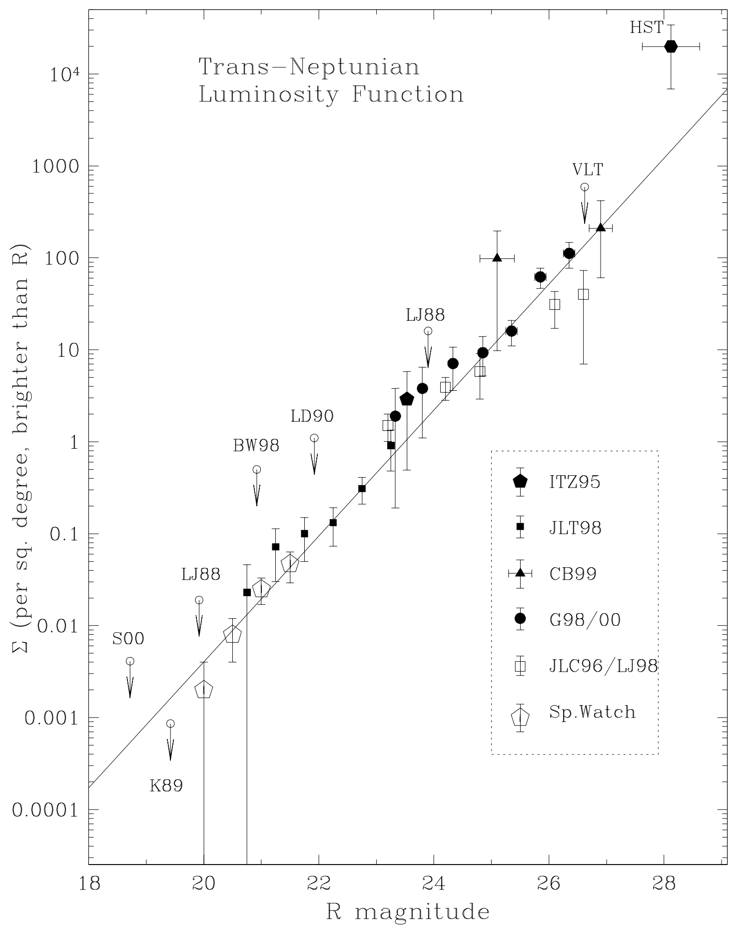
Quasars



Trans-Neptunian Objects (TNOs)
(Kuper belt objects, Plutinos, scattered KBOs . . .)

Data: Sets of magnitudes (m_i) or fluxes (F_i) from surveys with different coverage. Measurements may have significant uncertainty. $N =$ dozens to millions.

Goals: Infer TNO size and spatial distributions



Likelihood for an *Ideal* Magnitude Survey

Survey description

- Survey solid angle Ω
- Detect all TNOs brighter than m_{th}
- Report precise magnitudes, m_i

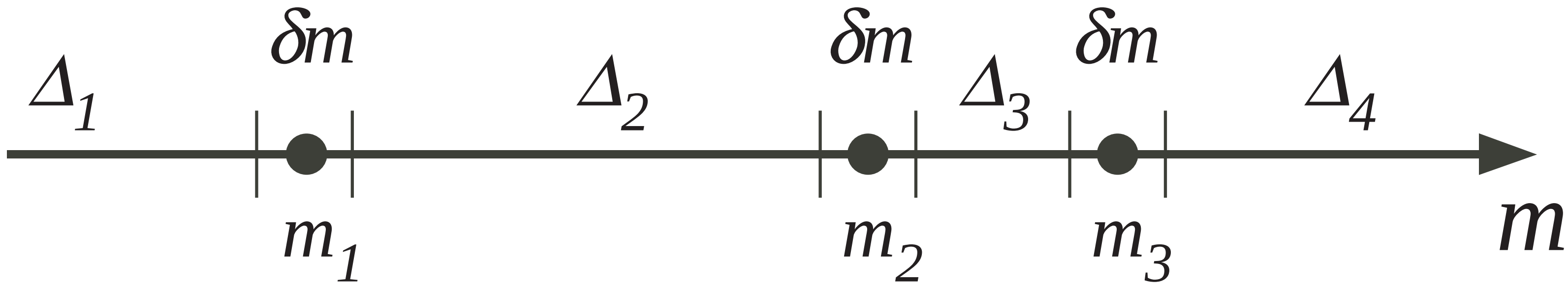
Poisson point process model

Diff'l distribution (objects/solid angle/unit m)
$\sigma(m) = \sigma(m;\theta)$ from model M with params θ

objects expected in $dm = \Omega\sigma(m)dm$

Counts in non-overlapping intervals have independent Poisson probabilities (given σ)

Point process likelihood construction:



Empty intervals: $p_0 = \exp\left[-\Omega \int_\Delta dm\ \sigma(m)\right]$

Detections: $p_1 = \Omega\,\sigma(m_i)\ \delta m \exp\left[-\Omega\,\sigma(m_i)\,\delta m\right]$

$$\rightarrow \mathcal{L} = \exp\left[-\Omega \int dm\,\Theta(m_{\mathrm{th}} - m)\sigma(m)\right] \prod_i \sigma(m_i)$$

Description of *Real* Magnitude survey

- Survey solid angle Ω
- Selection function/detection efficiency, $\eta(m)$
- Source measurements: $\ell_i(m) \equiv p(d_i|m)$, $i = 1$–N

Likelihood for a *Real* Magnitude Survey

Let $D = (\{d_i\}, \mathcal{N})$, with $\mathcal{N} \equiv$ "No other TNOs detected"'

$$\begin{aligned}\mathcal{L}(\theta) &= p(D|\theta, M) \\ &= \int \{dm_i\}\, p(\{m_i\}, \mathcal{N}|\theta, M)\, p(\{d_i\}|\{m_i\}, \mathcal{N}, \theta, M).\end{aligned}$$

First factor gives ideal likelihood terms with $\Theta \rightarrow \eta$.

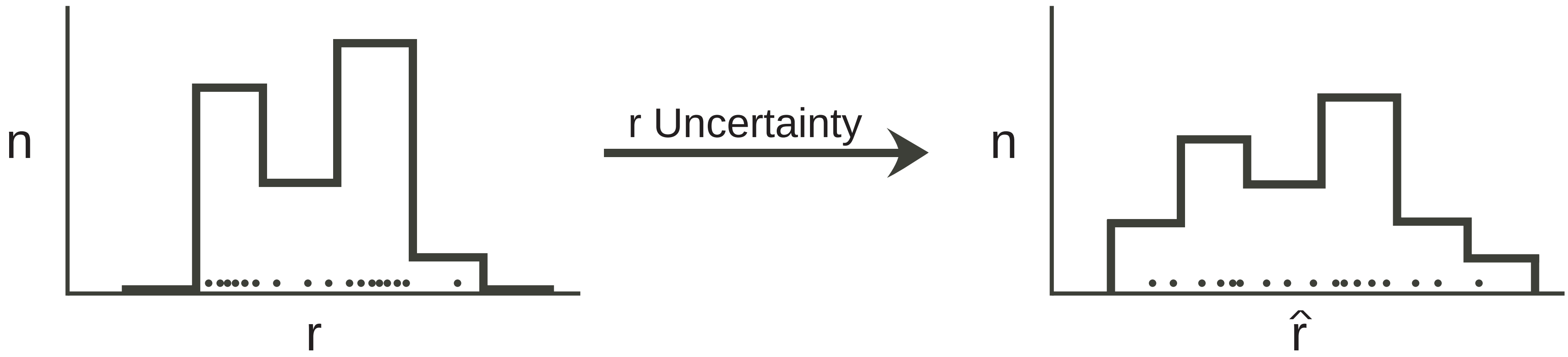
Second factor is product of $\ell_i(m_i) \quad \rightarrow$

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

Note source factors are *not* $\int dm\, \eta(m)\ell_i(m)\sigma(m)$

Importance of Source Uncertainties

Source uncertainty distorts distributions



First discussed by Eddington, Jeffreys, Malmquist

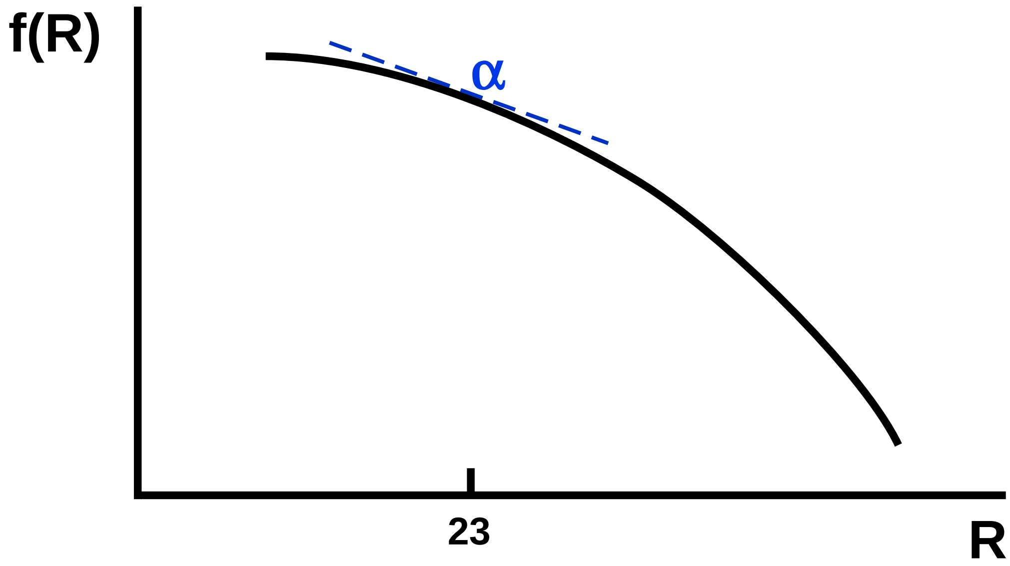
Well known in cosmology/star counts: Malmquist & Lutz-Kelker bias

To analyze SFDs, must estimate all source parameters, accounting for uncertainty—explicitly or implicitly

Example—Distribution of Source Fluxes

Measure $R = -2.5\log(\text{flux})$ from sources following a “rolling power law” distribution

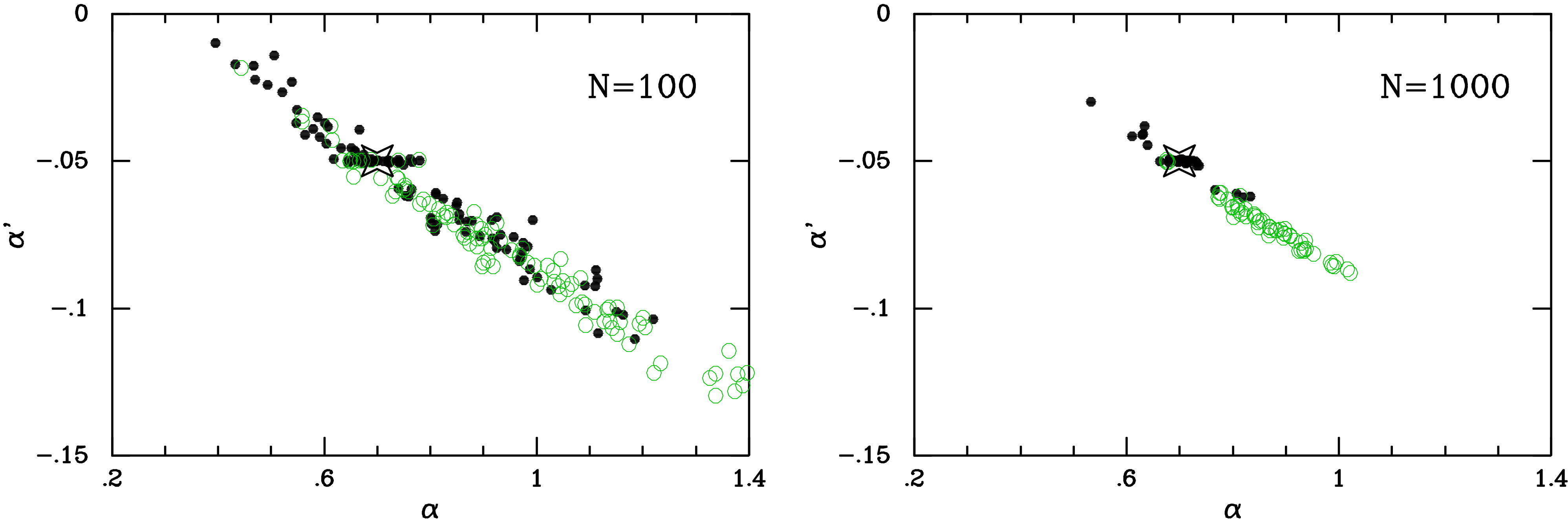
$$f(R) \propto 10^{[\alpha(R-23)+\alpha'(R-23)^2]}$$



Measurements have uncertainties 1% (bright) to ≈ 30% (dim)

Analyze simulated data with maximum likelihood and Bayes

Parameter estimates from correct likelihood (black dots) and likelihood ignoring uncertainties (green circles):



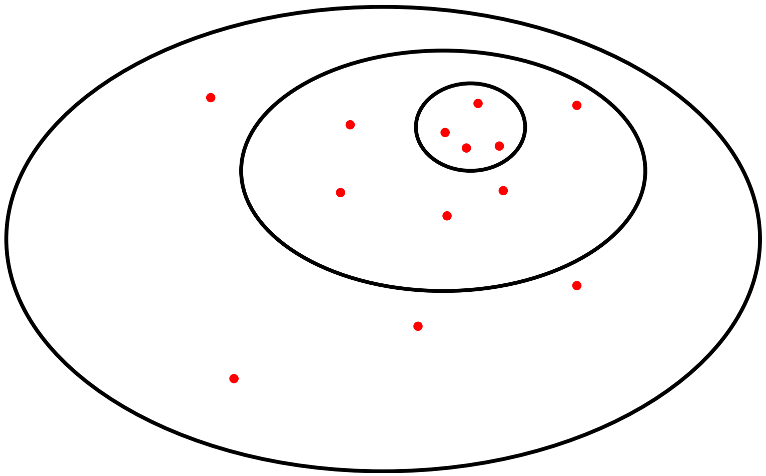
Uncertainties don't "average out"!

Bias grows *worse* with N ("inconsistent")

Measurement Error Problems

Incidental or latent parameters

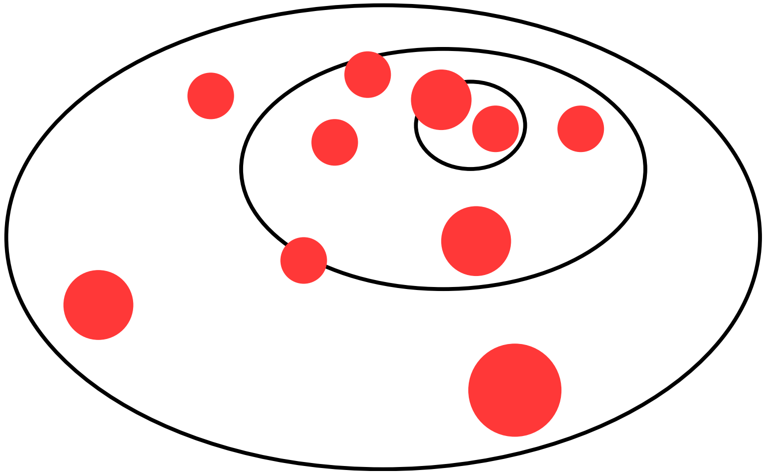
Suppose $f(x|\theta)$ is a distribution for an observable, x.



From N samples, $\{x_i\}$, we can infer θ from

$$\mathcal{L}(\theta) \equiv p(\{x_i\}|\theta) = \prod_i f(x_i|\theta)$$

But what if the x data are *noisy*, $d_i = x_i + \epsilon_i$?



We should somehow incorporate $\ell_i(x_i) = p(d_i|x_i)$

$$\begin{aligned}\mathcal{L}(\theta, \{x_i\}) &\equiv p(\{d_i\}|\theta, \{x_i\}) \\ &= \prod_i \ell_i(x_i) f(x_i|\theta)\end{aligned}$$

Maximizing over x_i and integrating over x_i give very different results!

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: Measurement error important & easily handled