Data Analysis Using Bayesian Inference With Applications in Astrophysics

A Survey

Tom Loredo

Dept. of Astronomy, Cornell University

Outline

- Overview of Bayesian inference
  - What to do
  - How to do it
  - Why do it this way
- Astrophysical examples
  - The “on/off” problem
  - Supernova Neutrinos

What To Do: The Bayesian Recipe

Assess hypotheses by calculating their probabilities $p(H_i|\ldots)$ conditional on known and/or presumed information using the rules of probability theory.

But . . . what does $p(H_i|\ldots)$ *mean*?

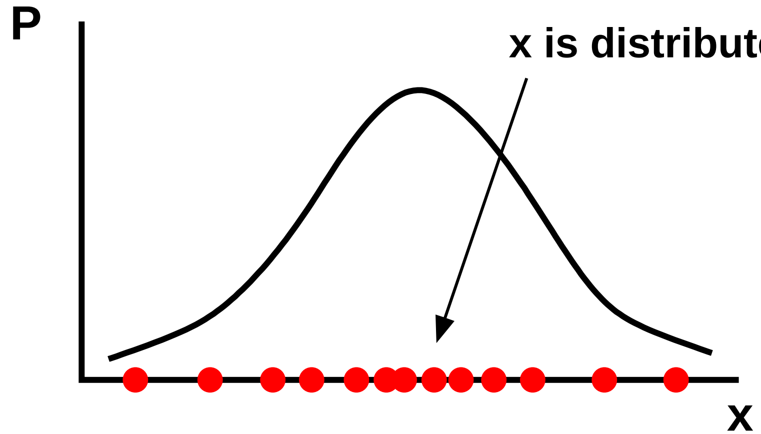
What is distributed in $p(x)$?

Frequentist: Probability describes "randomness"

Venn, Boole, Fisher, Neymann, Pearson. . .

x is a *random variable* if it takes different values throughout an infinite (imaginary?) ensemble of "identical" sytems/experiments.

$p(x)$ describes how x is distributed throughout the ensemble.



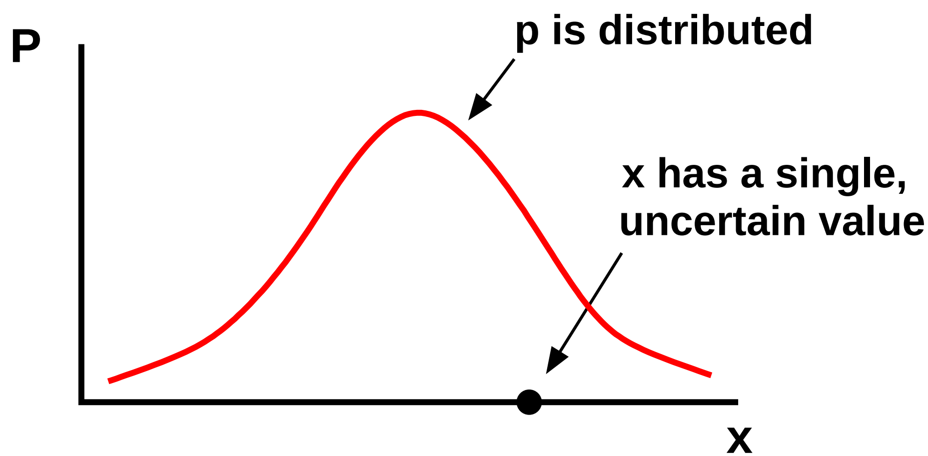
Probability $\equiv$ frequency (pdf $\equiv$ histogram).

Bayesian: Probability describes uncertainty

Bernoulli, Laplace, Bayes, Gauss...

$p(x)$ describes how probability (plausibility) is distributed among the possible choices for x in the case at hand.

Analog: a mass density, $\rho(x)$



Relationships between probability and frequency were demonstrated mathematically (large number theorems, Bayes's theorem).

Interpreting Abstract Probabilities

Symmetry/Invariance/Counting

- Resolve possibilities into equally plausible "microstates" using symmetries
- Count microstates in each possibility

Frequency from probability

Bernoulli's laws of large numbers: In repeated trials, given $P(\text{success})$, predict

$$\frac{N_{\text{success}}}{N_{\text{total}}} \to P \quad \text{as} \quad N \to \infty$$

Probability from frequency

Bayes's "An Essay Towards Solving a Problem in the Doctrine of Chances" $\longrightarrow$ Bayes's theorem

Probability $\neq$ *Frequency!*

Bayesian Probability: A Thermal Analogy

Intuitive notion	*Quantification*	*Calibration*
Hot, cold	Temperature, T	Cold as ice $= 273$K Boiling hot $= 373$K
uncertainty	Probability, P	Certainty $= 0$, 1 $p = 1/36$: plausible as "snake's eyes" $p = 1/1024$: plausible as 10 heads

The Bayesian Recipe

Assess hypotheses by calculating their probabilities $p(H_i|\ldots)$ conditional on known and/or presumed information using the rules of probability theory.

Probability Theory Axioms ("grammar"):

'OR' (sum rule)
$$P(H_1 + H_2|I) = P(H_1|I) + P(H_2|I) - P(H_1, H_2|I)$$

'AND' (product rule)
$$\begin{aligned} P(H_1, D|I) &= P(H_1|I)\, P(D|H_1, I) \\ &= P(D|I)\, P(H_1|D, I) \end{aligned}$$

Direct Probabilities ("vocabulary"):

- Certainty: If A is certainly true given B, $P(A|B) = 1$
- Falsity: If A is certainly false given B, $P(A|B) = 0$
- Other rules exist for more complicated types of information; for example, invariance arguments, maximum (information) entropy, limit theorems (CLT; tying probabilities to frequencies), bold (or desperate!) presumption...

Important Theorems

Normalization:

For *exclusive, exhaustive* H_i

$$\sum_i P(H_i|\cdots) = 1$$

Bayes's Theorem:

$$P(H_i|D,I) = P(H_i|I)\,\frac{P(D|H_i,I)}{P(D|I)}$$

posterior $\propto$ prior $\times$ likelihood

Marginalization:

Note that for exclusive, exhaustive $\{B_i\}$,

$$\begin{aligned}\sum_i P(A, B_i|I) &= \sum_i P(B_i|A, I)P(A|I) = P(A|I) \\ &= \sum_i P(B_i|I)P(A|B_i, I)\end{aligned}$$

$\rightarrow$ We can use $\{B_i\}$ as a "basis" to get $P(A|I)$.

Example: Take $A = D$, $B_i = H_i$; then

$$\begin{aligned}P(D|I) &= \sum_i P(D, H_i|I) \\ &= \sum_i P(H_i|I)P(D|H_i, I)\end{aligned}$$

prior predictive for D = Average likelihood for H_i

Inference With Parametric Models

Parameter Estimation

$I =$ Model M with parameters θ ($+$ any add'l info)

$H_i =$ statements about θ; e.g. "$\theta \in [2.5, 3.5]$," or "$\theta > 0$"

Probability for any such statement can be found using a *probability density function* (pdf) for θ:

$$\begin{aligned} P(\theta \in [\theta, \theta + d\theta] | \cdots) &= f(\theta) d\theta \\ &= p(\theta | \cdots) d\theta \end{aligned}$$

Posterior probability density:

$$p(\theta|D, M) = \frac{p(\theta|M)\ \mathcal{L}(\theta)}{\int d\theta\ p(\theta|M)\ \mathcal{L}(\theta)}$$

Summaries of posterior:

- “Best fit” values: mode, posterior mean
- Uncertainties: Credible regions (e.g., HPD regions)
- Marginal distributions:
  - Interesting parameters ψ, nuisance parameters ϕ
  - Marginal dist’n for ψ:

$$p(\psi|D, M) = \int d\phi\, p(\psi, \phi|D, M)$$

Generalizes “propagation of errors”

Model Uncertainty: Model Comparison

$I = (M_1 + M_2 + \ldots)$ — Specify a set of models.
$H_i = M_i$ — Hypothesis chooses a model.

Posterior probability for a model:

$$\begin{aligned} p(M_i|D, I) &= p(M_i|I)\frac{p(D|M_i, I)}{p(D|I)} \\ &\propto p(M_i)\mathcal{L}(M_i) \end{aligned}$$

But $\mathcal{L}(M_i) = p(D|M_i) = \int d\theta_i \, p(\theta_i|M_i)p(D|\theta_i, M_i)$.

Likelihood for model $=$ Average likelihood for its parameters

$$\mathcal{L}(M_i) = \langle \mathcal{L}(\theta_i) \rangle$$

Model Uncertainty: Model Averaging

Models have a common subset of interesting parameters, ψ.

Each has different set of nuisance parameters ϕ_i (or different prior info about them).

$H_i =$ statements about ψ.

Calculate posterior PDF for ψ:

$$
\begin{aligned}
p(\psi|D, I) &= \sum_i p(\psi|D, M_i)p(M_i|D, I) \\
&\propto \sum_i \mathcal{L}(M_i) \int d\theta_i \, p(\psi, \phi_i|D, M_i)
\end{aligned}
$$

The model choice is itself a (discrete) nuisance parameter here.

What's the Difference?

Bayesian Inference (BI):

- Specify at least two competing hypotheses and priors
- Calculate their probabilities using probability theory
  - Parameter estimation:

$$p(\theta|D, M) = \frac{p(\theta|M)\mathcal{L}(\theta)}{\int d\theta\, p(\theta|M)\mathcal{L}(\theta)}$$

  - Model Comparison:

$$O \propto \frac{\int d\theta_1\, p(\theta_1|M_1)\, \mathcal{L}(\theta_1)}{\int d\theta_2\, p(\theta_2|M_2)\, \mathcal{L}(\theta_2)}$$

Frequentist Statistics (FS):

- Specify null hypothesis H_0 such that rejecting it implies an interesting effect is present
- Specify statistic $S(D)$ that measures departure of the data from null expectations
- Calculate $p(S|H_0) = \int dD\, p(D|H_0)\delta[S - S(D)]$ (e.g. by Monte Carlo simulation of data)
- Evaluate $S(D_{\rm obs})$; decide whether to reject H_0 based on, e.g., $\int_{>S_{\rm obs}} dS\, p(S|H_0)$

Crucial Distinctions

The role of subjectivity:

BI exchanges (implicit) subjectivity in the choice of null & statistic for (explicit) subjectivity in the specification of alternatives.

- Makes assumptions explicit
- Guides specification of further alternatives that generalize the analysis
- Automates identification of statistics:
  - BI is a problem-solving approach
  - FS is a solution-characterization approach

The types of mathematical calculations:

- BI requires integrals over hypothesis/parameter space
- FS requires integrals over sample/data space

An Example Confidence/Credible Region

Infer μ : $\quad x_i = \mu + \epsilon_i; \qquad p(x_i|\mu, M) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$

$$\rightarrow \mathcal{L}(\mu) \quad \propto \quad \exp\left[-\frac{(\bar{x}-\mu)^2}{2(\sigma/\sqrt{N})^2}\right]$$

68% confidence region: $\bar{x} \pm \sigma/\sqrt{N}$

$$\int d^N x_i \cdots = \int d(\text{angles}) \int_{\bar{x}-\sigma/\sqrt{N}}^{\bar{x}+\sigma/\sqrt{N}} d\bar{x} \cdots = 0.683$$

68% credible region: $\bar{x} \pm \sigma/\sqrt{N}$

$$\frac{\int_{\bar{x}-\sigma/\sqrt{N}}^{\bar{x}+\sigma/\sqrt{N}} d\mu \exp\left[-\frac{(\bar{x}-\mu)^2}{2(\sigma/\sqrt{N})^2}\right]}{\int_{-\infty}^{\infty} d\mu \exp\left[-\frac{(\bar{x}-\mu)^2}{2(\sigma/\sqrt{N})^2}\right]} \approx 0.683$$

Difficulty of Parameter Space Integrals

Inference with independent data:

Consider N data, $D = \{x_i\}$; and model M with m parameters $(m \ll N)$.

Suppose $\mathcal{L}(\theta) = p(x_1|\theta)\, p(x_2|\theta) \cdots p(x_N|\theta)$.

Frequentist integrals:

$$\int dx_1\, p(x_1|\theta) \int dx_2\, p(x_2|\theta) \cdots \int dx_N\, p(x_N|\theta) f(D)$$

Seek integrals with properties independent of θ. Such rigorous frequentist integrals usually can't be found.

Approximate (e.g., asymptotic) results are easy via Monte Carlo (due to independence).

Bayesian integrals:

$$\int d^m\theta \; g(\theta)\, p(\theta|M)\, \mathcal{L}(\theta)$$

Such integrals are sometimes easy if analytic (especially in low dimensions).

Asymptotic approximations require ingredients familiar from frequentist calculations.

For large m (> 4 is often enough!) the integrals are often very challenging because of correlations (lack of independence) in parameter space.

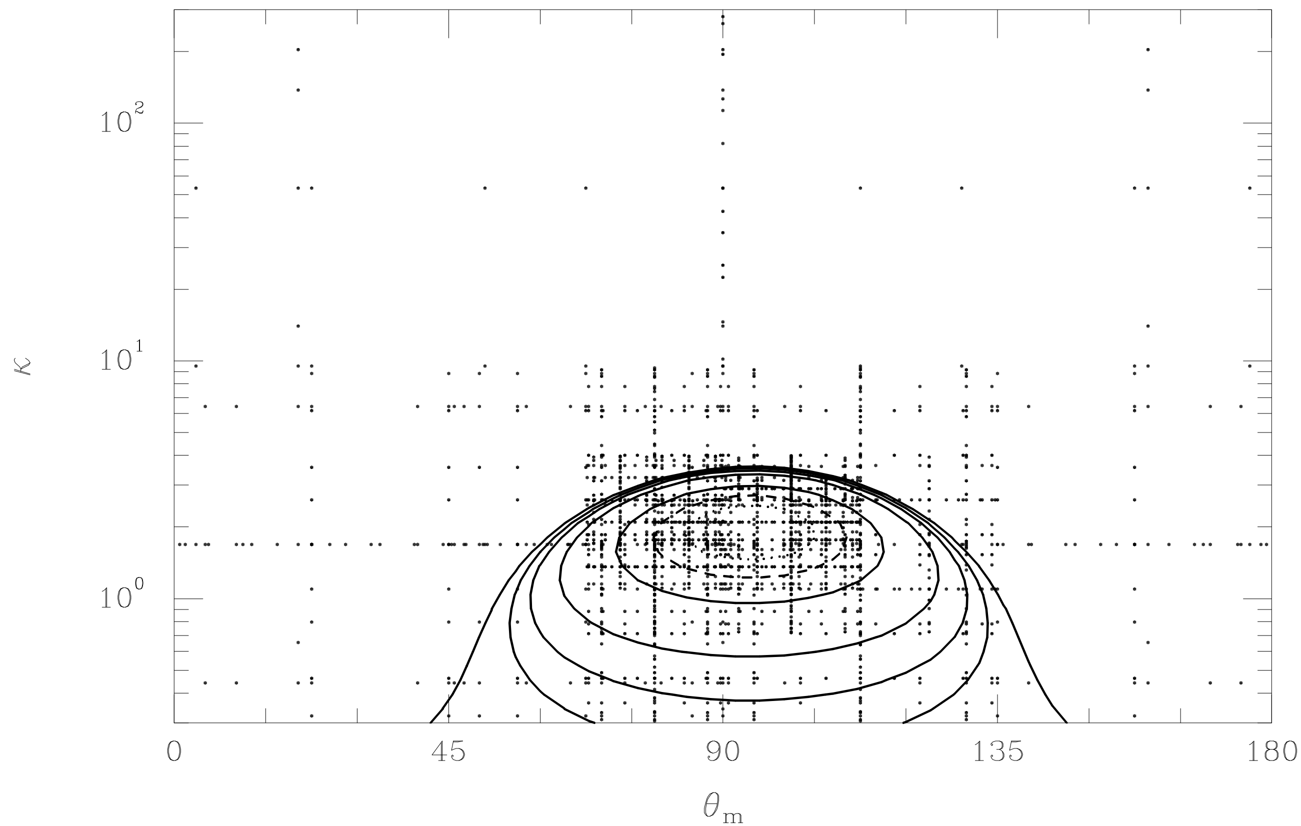
How To Do It

Tools for Bayesian Calculation

- Asymptotic (large N) approximation: Laplace approximation
- Low-D Models $(m \lesssim 10)$:
  - Randomized Quadrature: Quadrature $+$ dithering
  - Subregion-Adaptive Quadrature: `ADAPT`, `DCUHRE`, `BAYESPACK`
  - Adaptive Monte Carlo: `VEGAS`, `miser`
- High-D Models ($m \sim 5$–10^6): Posterior Sampling
  - Rejection method
  - Markov Chain Monte Carlo (MCMC)

Subregion-Adaptive Quadrature

Concentrate points where most of the probability lies via recursion. Use a pair of lattice rules (for error estim'n), subdivide regions w/ large error.



ADAPT in action (galaxy polarizations)

Tools for Bayesian Calculation

- Asymptotic (large N) approximation: Laplace approximation
- Low-D Models $(m \lesssim 10)$:
  - Randomized Quadrature: Quadrature $+$ dithering
  - Subregion-Adaptive Quadrature: `ADAPT`, `DCUHRE`, `BAYESPACK`
  - Adaptive Monte Carlo: `VEGAS`, `miser`
- High-D Models ($m \sim 5$–10^6): Posterior Sampling
  - Rejection method
  - Markov Chain Monte Carlo (MCMC)

Posterior Sampling

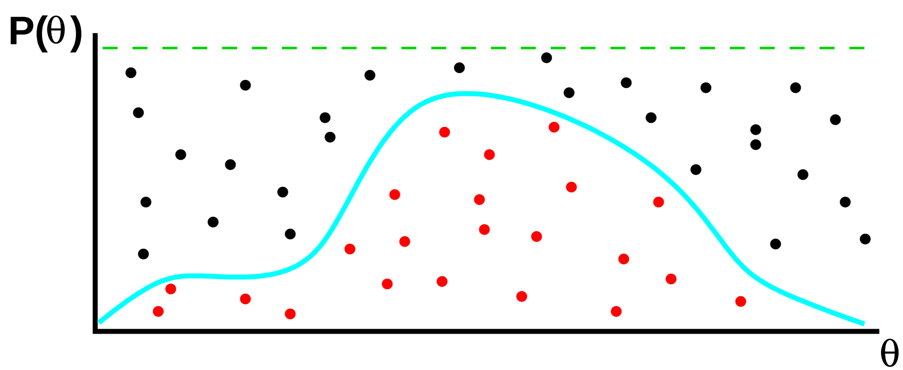
General Approach:

Draw samples of θ, ϕ from $p(\theta, \phi|D, M)$; then:

- Integrals, moments easily found via $\sum_i f(\theta_i, \phi_i)$
- $\{\theta_i\}$ are samples from $p(\theta|D, M)$

But how can we obtain $\{\theta_i, \phi_i\}$?

Rejection Method:



Hard to find efficient comparison function if $m \gtrsim 6$.

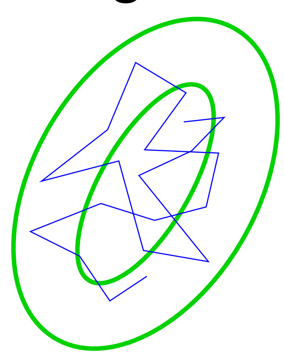
Markov Chain Monte Carlo (MCMC)

Let $$-\Lambda(\theta) = \ln\left[p(\theta|M)\, p(D|\theta, M)\right]$$

Then $$p(\theta|D, M) = \frac{e^{-\Lambda(\theta)}}{Z} \qquad Z \equiv \int d\theta\; e^{-\Lambda(\theta)}$$

Bayesian integration looks like problems addressed in computational statmech and Euclidean QFT.

Markov chain methods are standard: Metropolis; Metropolis-Hastings; molecular dynamics; hybrid Monte Carlo; simulated annealing



The MCMC Recipe:

Create a “time series” of samples θ_i from $p(\theta)$:

- Draw a candidate θ_{i+1} from a kernel $T(\theta_{i+1}|\theta_i)$
- Enforce “detailed balance” by accepting with $p = \alpha$

$$\alpha(\theta_{i+1}|\theta_i) = \min\left[1, \frac{T(\theta_i|\theta_{i+1})p(\theta_{i+1})}{T(\theta_{i+1}|\theta_i)p(\theta_i)}\right]$$

Choosing T to minimize “burn-in” and corr’ns is an art.

Coupled, parallel chains eliminate this for select problems (“exact sampling”).

Why Do It

- What you get
- What you avoid
- Foundations

What you get

- Probabilities *for hypotheses*
  - Straightforward interpretation
  - Identify weak experiments
  - Crucial for global (hierarchical) analyses (e.g., pop'n studies)
  - Forces analyst to be explicit about assumptions
- Handle Nuisance parameters
- Valid for all sample sizes
- Handles multimodality
- Quantitative Occam's razor
- Model comparison for > 2 alternatives; needn't be nested

And there’s more . . .

- Use prior info/combine experiments
- Systematic error treatable
- Straightforward experimental design
- Good frequentist properties:
  - Consistent
  - Calibrated—E.g., if you choose a model only if odds > 100, you will be right ≈ 99% of the time
  - Coverage as good or better than common methods
- Unity/simplicity

What you avoid

- Hidden subjectivity/arbitrariness
- Dependence on “stopping rules”
- Recognizable subsets
- Defining number of “independent” trials in searches
- Inconsistency & incoherence (e.g., inadmissable estimators)
- Inconsistency with prior information
- Complexity of interpretation (e.g., significance vs. sample size)

Foundations

"Many Ways To Bayes"

- Consistency with logic $+$ internal consistency $\rightarrow$ BI
  (Cox; Jaynes; Garrett)
- "Coherence"/Optimal betting $\rightarrow$ BI (Ramsey; DeFinetti; Wald)
- Avoiding recognizable subsets $\rightarrow$ BI (Cornfield)
- Avoiding stopping rule problems $\rightarrow$ $\mathcal{L}$-principle
  (Birnbaum; Berger & Wolpert)
- Algorithmic information theory $\rightarrow$ BI
  (Rissanen; Wallace & Freeman)
- Optimal information processing $\rightarrow$ BI (Good; Zellner)

There is probably something to all of this!

What the theorems mean

When reporting numbers ordering hypotheses, values must be consistent with calculus of probabilities for hypotheses.

Many frequentist methods satisfy this requirement.

Role of priors

Priors are not fundamental!

Priors are analogous to initial conditions for ODEs.

- Sometimes crucial
- Sometimes a nuisance

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

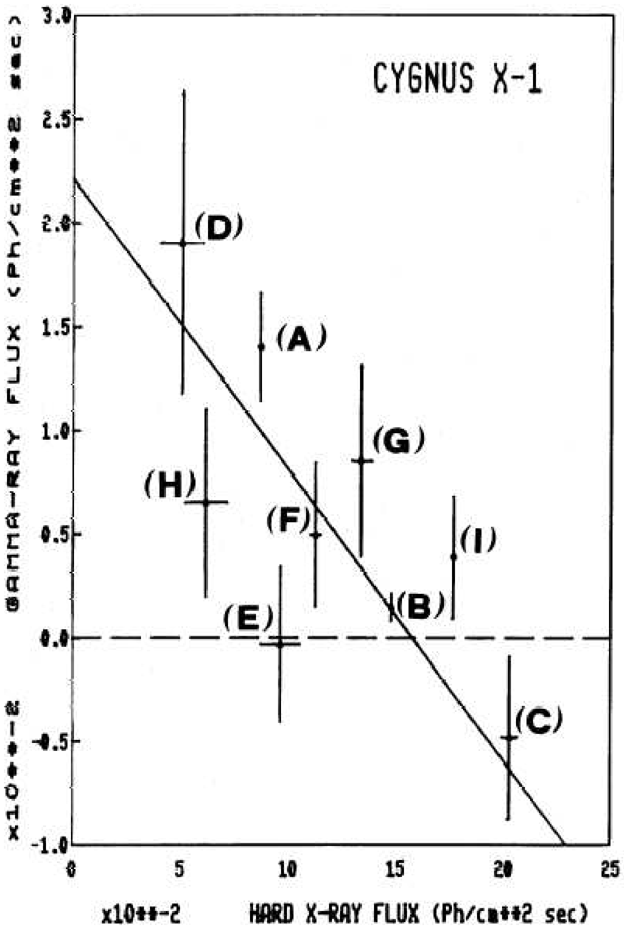
$$\begin{aligned}
\hat{b} &= N_{\rm off}/T_{\rm off}; & \sigma_b &= \sqrt{N_{\rm off}}/T_{\rm off} \\
\hat{r} &= N_{\rm on}/T_{\rm on} - \hat{b}; & \sigma_r &= \sqrt{N_{\rm on}}/T_{\rm on} \\
\hat{s} &= \hat{r} - \hat{b}; & \sigma_s &= \sqrt{\sigma_r^2 + \sigma_b^2}
\end{aligned}$$

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

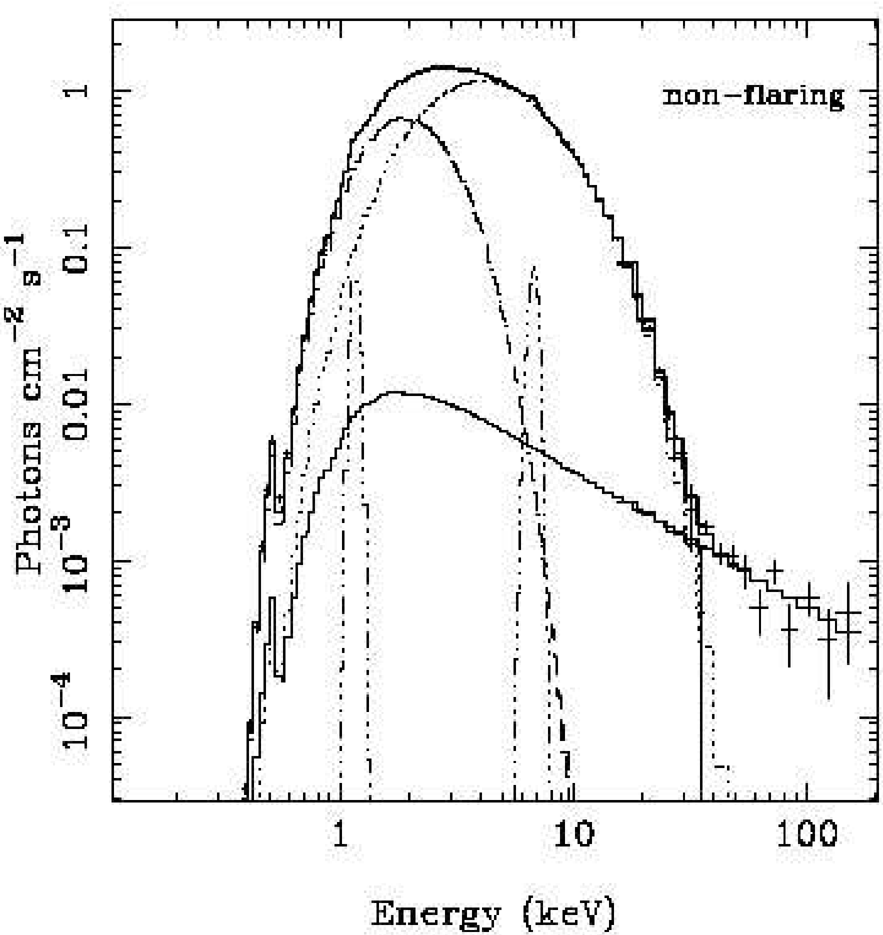
Examples

Spectra of X-Ray Sources

Bassani et al. 1989

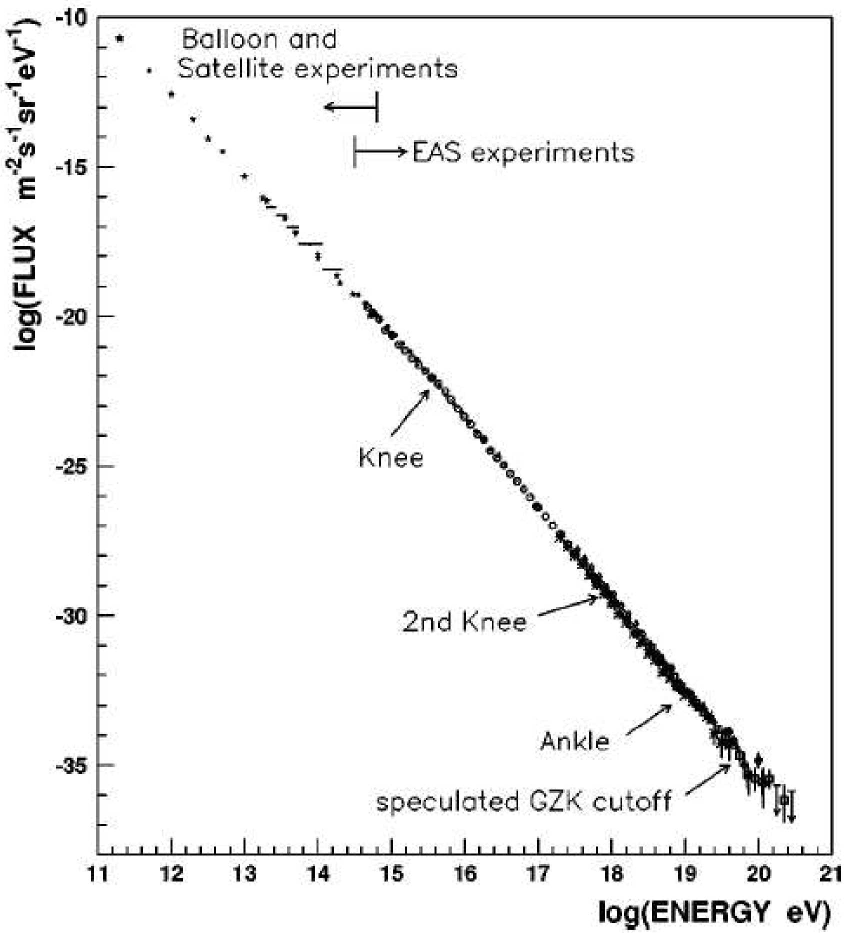


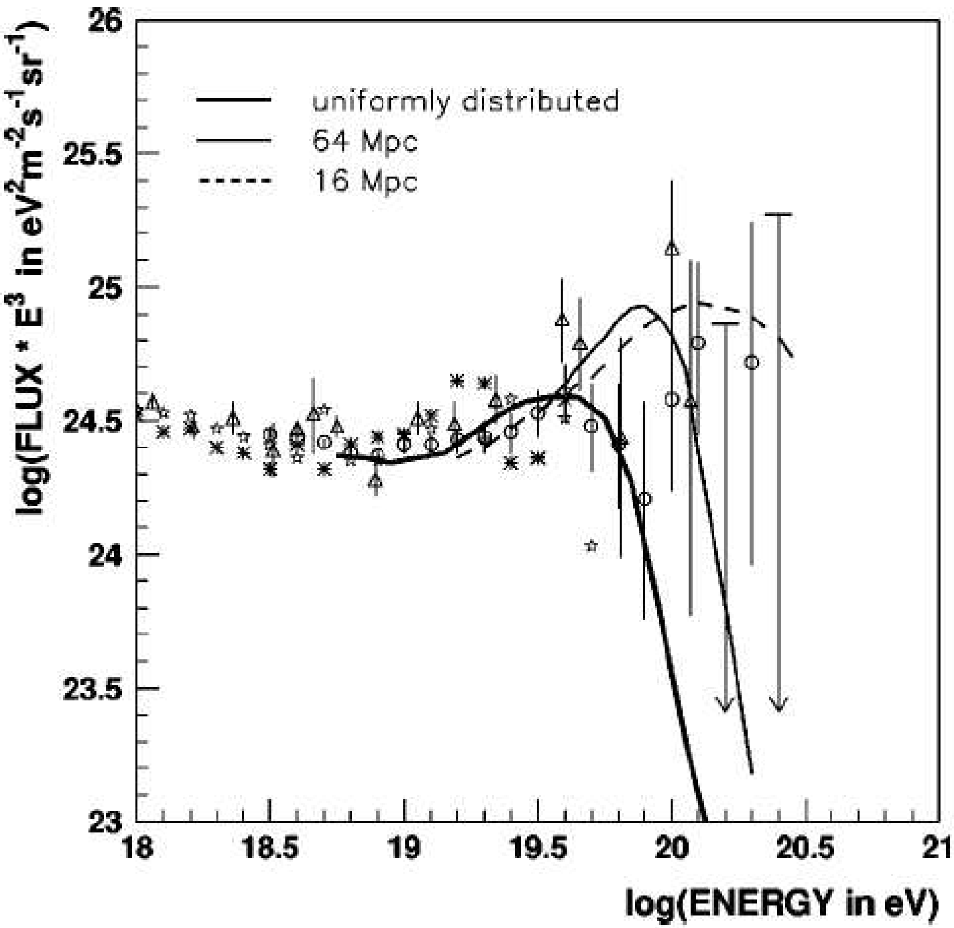
Di Salvo et al. 2001



Spectrum of Ultrahigh-Energy Cosmic Rays

Nagano & Watson 2000





Bayesian Solution

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.

About that flat prior . . .

Bayes's justification for a flat prior

Not that ignorance of $r \to p(r|I) = C$

Require (discrete) predictive distribution to be flat:

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

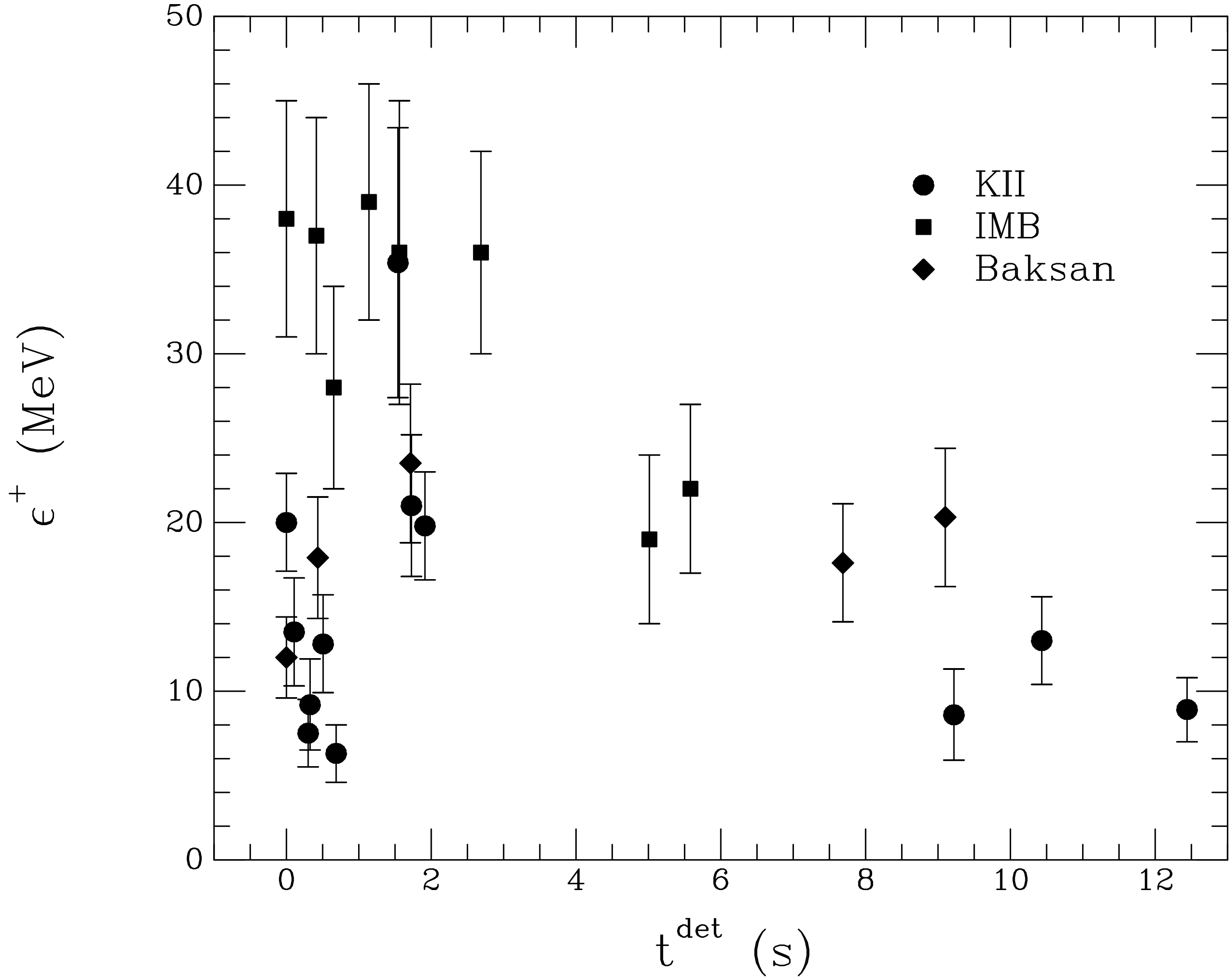
A convention

- 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

Supernova Neutrinos

Tarantula Nebula in the LMC, ca. Feb 1987

Neutrinos from Supernova SN 1987A



Why Reconsider the SN Neutrinos?

Advances in astrophysics

Two scenarios for Type II SN: prompt and delayed

’87: Delayed scenario new, poorly understood

Prompt scenario problematic, but favored

$\longrightarrow$ Most analyses presumed prompt scenario

’90s: Consensus that prompt shock fails
Better understanding of delayed scenario

Advances in statistics

’89: First applications of Bayesian methods to modern astrophysical problems

’90s: Diverse Bayesian analyses of Poisson processes
Better computational methods

Likelihood for SN Neutrino Data

Models for neutrino rate spectrum

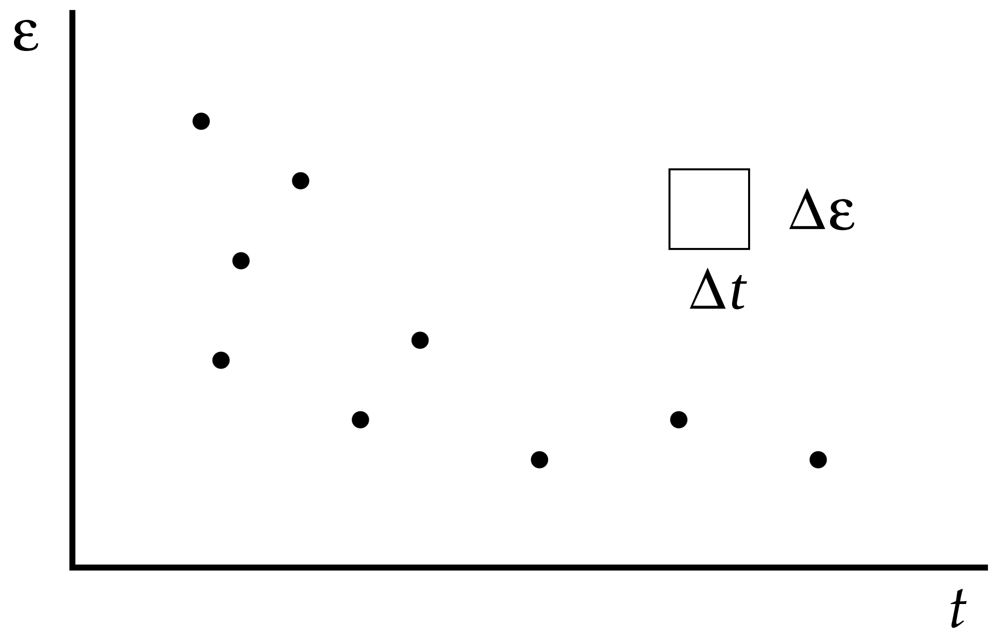
$$R(\epsilon, t) = \left[\begin{array}{c} \text{Emitted} \\ \bar{\nu}_e \text{ signal} \end{array} \right] \times \left[\begin{array}{c} \text{Propagation} \\ \text{to earth} \end{array} \right] \times \left[\begin{array}{c} \text{Interaction} \\ \text{w/ detector} \end{array} \right]$$

$$= \text{Astrophysics} \times \begin{array}{c} \text{Particle} \\ \text{physics} \end{array} \times \begin{array}{c} \text{Instrument} \\ \text{properties} \end{array}$$

Models have ≥ 6 parameters; $3+$ are nuisance parameters.

Ideal Observations

Detect all captured $\bar{\nu}_e$ with precise (ϵ, t)



$$\mathcal{L}(\theta) = \left[\prod p(\text{non-dtxns})\right] \times \left[\prod p(\text{dtxns})\right]$$

$$= \exp\left[-\int dt \int d\epsilon R(\epsilon, t)\right] \prod_i R(\epsilon_i, t_i)$$

Real Observations

- Detection efficiency $\eta(\epsilon) < 1$
- ϵ_i measured with significant uncertainty

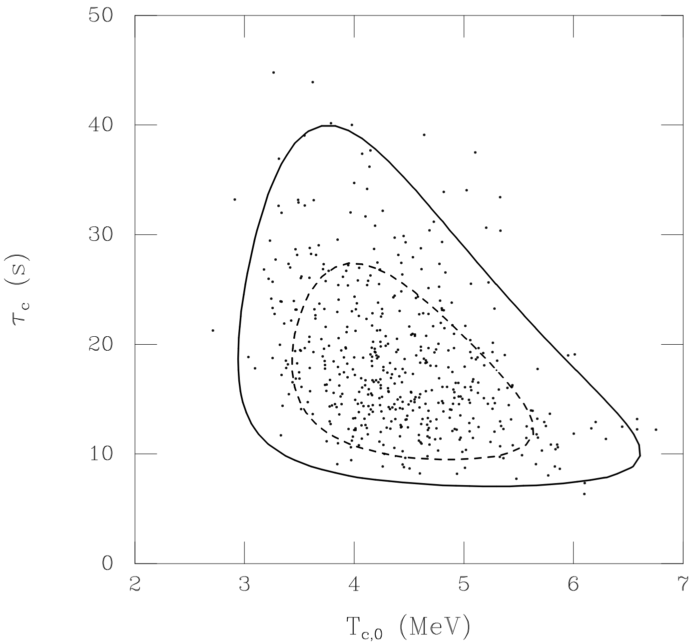
Let $\ell_i(\epsilon) = p(d_i|\epsilon, I)$; "individual event likelihood"

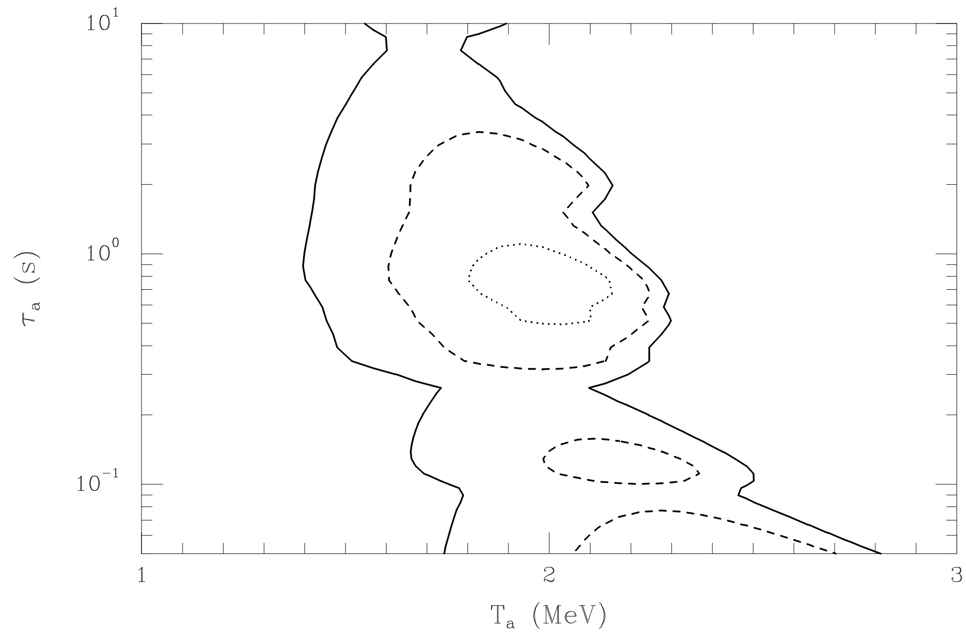
$$\mathcal{L}(\theta) = \exp\left[-\int dt \int d\epsilon\, \eta(\epsilon) R(\epsilon, t)\right] \prod_i \int d\epsilon_i\, \ell_i(\epsilon) R(\epsilon, t_i)$$

Instrument background rates and dead time further complicate $\mathcal{L}$.

Inferences for Signal Models

Two-component Model (Delayed Scenario)



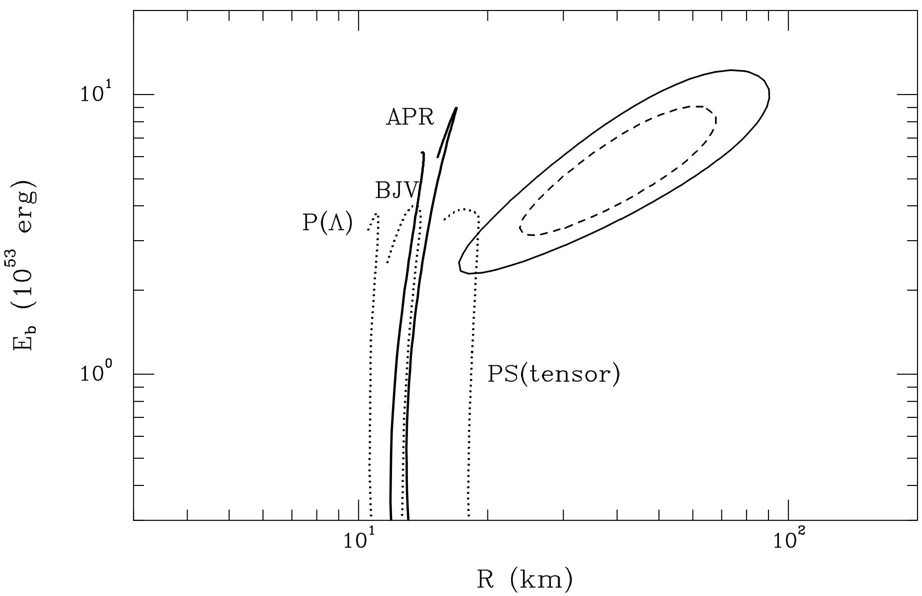


Odds favors delayed scenario by $\sim 10^2$ with conservative priors; by $\sim 10^3$ with informative priors.

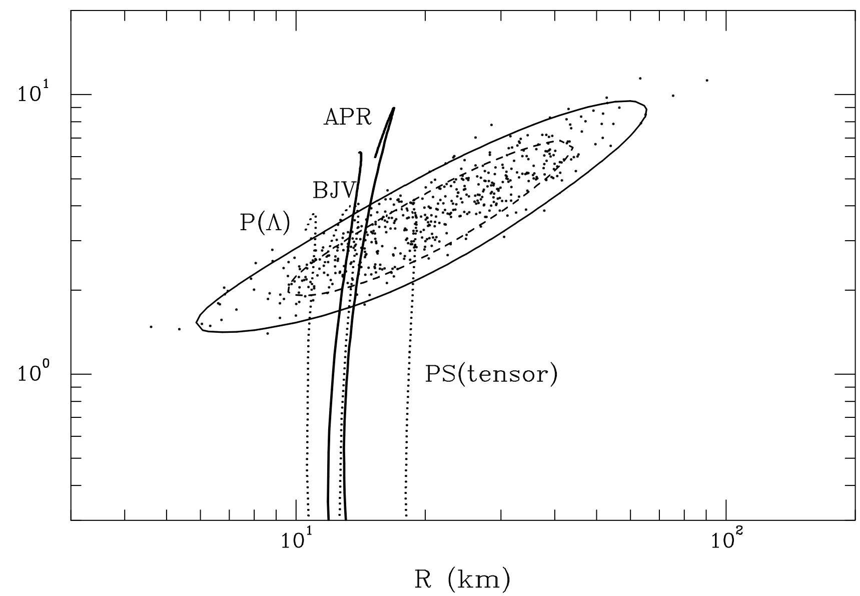
Prompt vs. Delayed SN Models

Nascent Neutron Star Properties

Prompt shock scenario



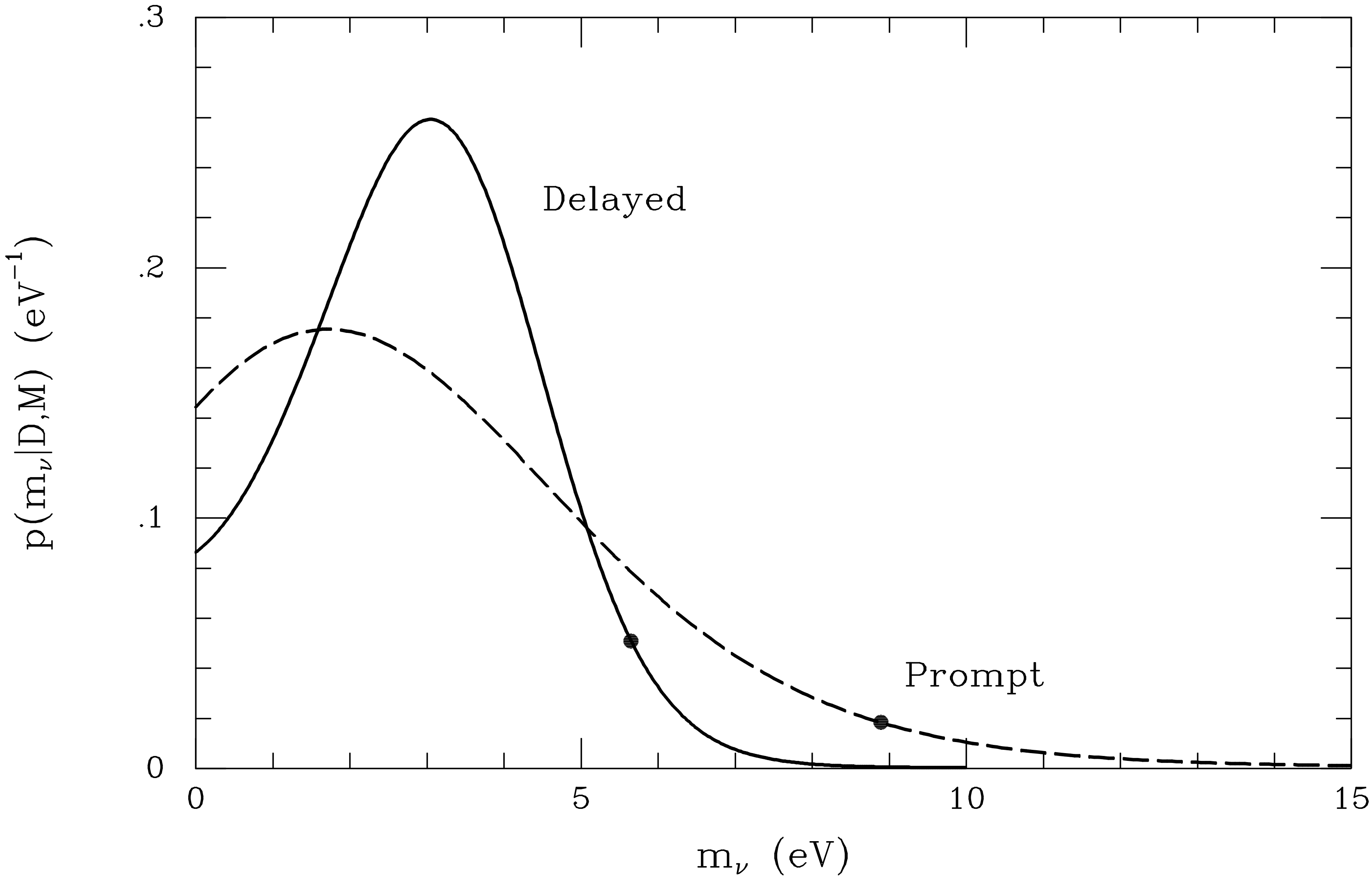
Delayed shock scenario



First direct evidence favoring delayed scenario.

Electron Antineutrino Rest Mass

Marginal Posterior for $m_{\bar{\nu}_e}$



Summary

Overview of Bayesian inference

- What to do
  - Calculate probabilities for hypotheses
  - Integrate over parameter space
- How to do it—many (unfamiliar?) tools
- Why do it this way—pragmatic & principled reasons

Astrophysical examples

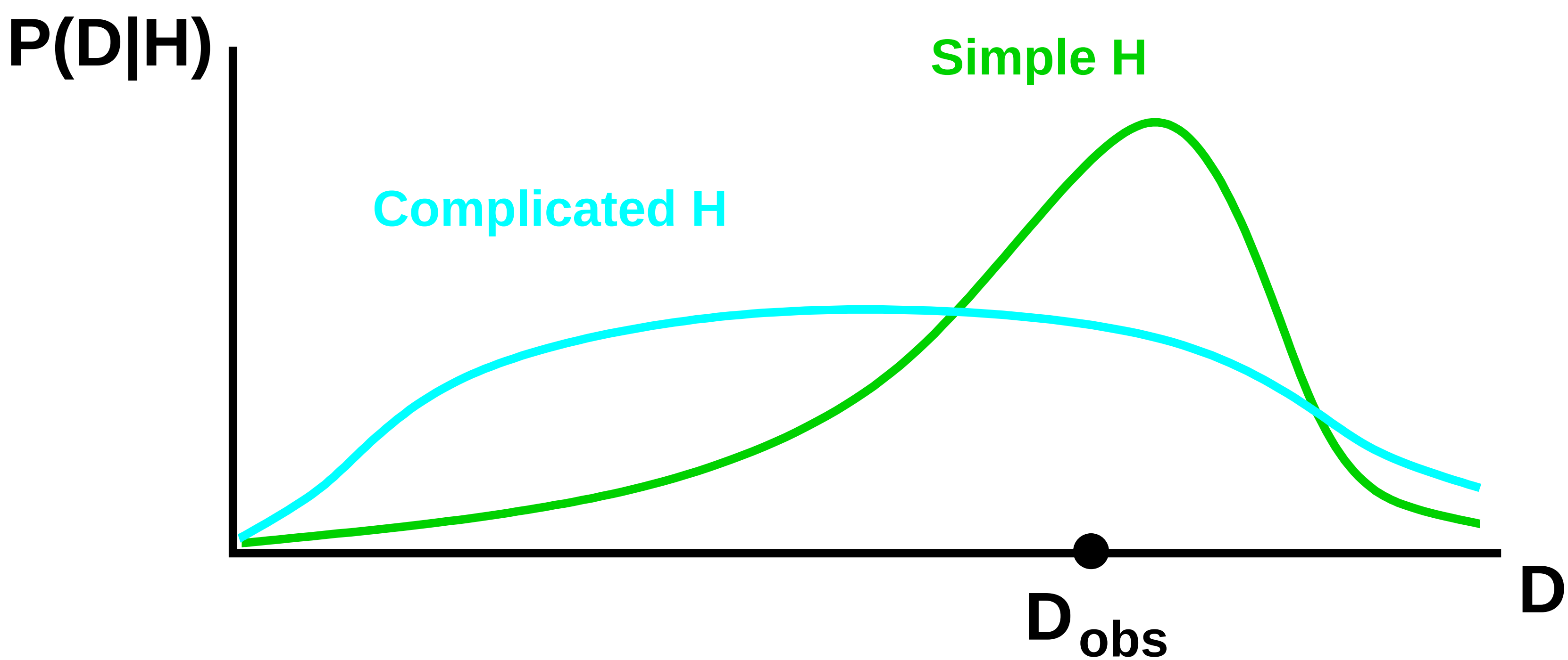
- The “on/off” problem—simple problem, new solution
- Supernova Neutrinos—A lot of info from few data!
  - Strongly favor delayed SN scenario
  - Constrain neutrino mass $\lesssim 6$ eV

That’s all, folks!

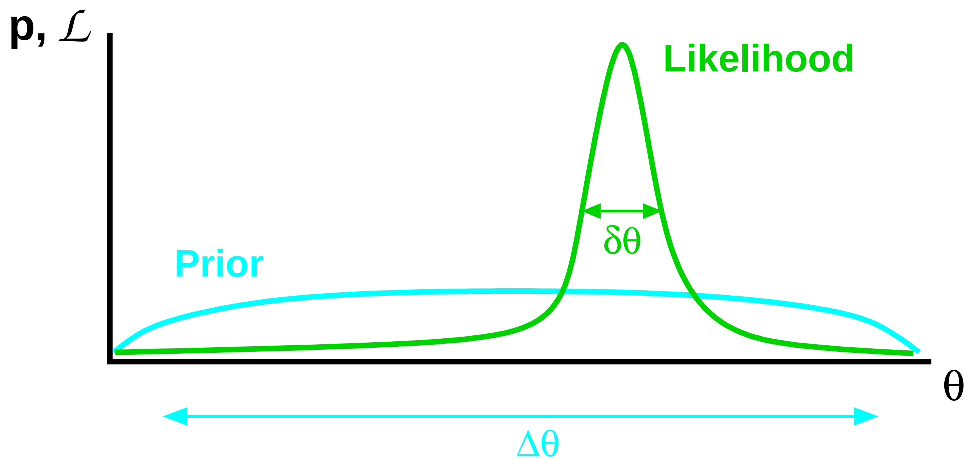
An Automatic Occam's Razor

Predictive probabilities can favor simpler models:

$$p(D|M_i) = \int d\theta_i \; p(\theta_i|M) \; \mathcal{L}(\theta_i)$$



The Occam Factor:



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

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

Occam factor penalizes models for “wasted” volume of parameter space.

Bayesian Calibration

Credible region $\Delta(D)$ with probability P:

$$P = \int_{\Delta(D)} d\theta \; p(\theta|I) \frac{p(D|\theta, I)}{p(D|I)}$$

What fraction of the time, Q, will the true θ be in $\Delta(D)$?

1. Draw θ from $p(\theta|I)$
2. Simulate data from $p(D|\theta, I)$
3. Calculate $\Delta(D)$ and see if $\theta \in \Delta(D)$

$$Q = \int d\theta \; p(\theta|I) \int dD \; p(D|\theta, I) \, [\theta \in \Delta(D)]$$

$$Q = \int d\theta \; p(\theta|I) \int dD \; p(D|\theta, I) \, [\theta \in \Delta(D)]$$

Note appearance of $p(\theta, D|I) = p(\theta|D, I)p(D|I)$:

$$\begin{aligned} Q &= \int dD \int d\theta \; p(\theta|D, I) \, p(D|I) \, [\theta \in \Delta(D)] \\ &= \int dD \; p(D|I) \; \int_{\Delta(D)} d\theta \; p(\theta|D, I) \\ &= P \int dD \; p(D|I) \\ &= P \end{aligned}$$

Bayesian inferences are "calibrated." *Always.*
Calibration is with respect to choice of prior & $\mathcal{L}$.

Real-Life Confidence Regions

Theoretical confidence regions

A rule $\delta(D)$ gives a region with covering probability:

$$C_\delta(\theta) \;=\; \int dD\, p(D|\theta, I)\, [\theta \in \delta(D)]$$

It's a *confidence region* iff $C(\theta) = P$, a *constant*.

Such rules almost never exist in practice!

Average coverage

Intuition suggests reporting some kind of average performance: $\int d\theta\ f(\theta) C_\delta(\theta)$

Recall the Bayesian calibration condition:

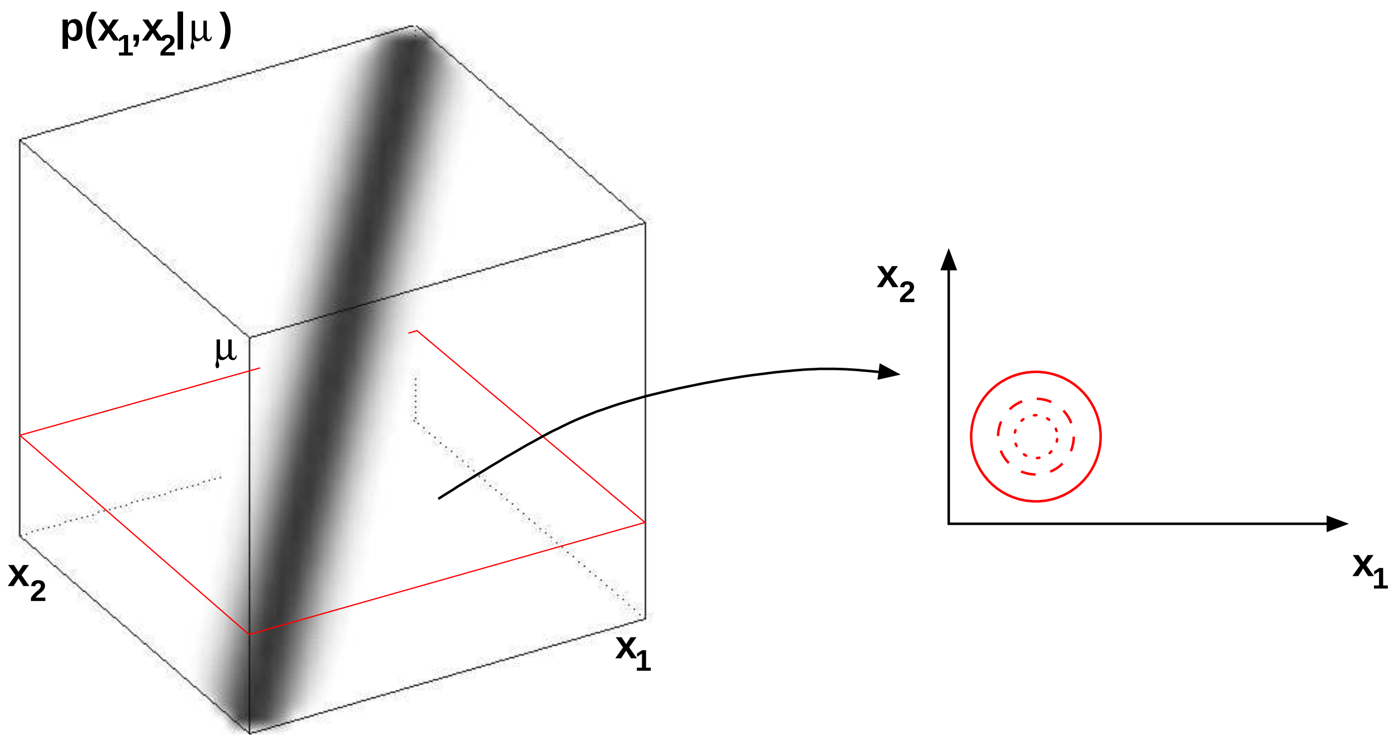
$$
\begin{aligned}
P &= \int d\theta\, p(\theta|I) \int dD\, p(D|\theta, I)\, [\theta \in \Delta(D)] \\
&= \int d\theta\, p(\theta|I)\, C_\delta(\theta)
\end{aligned}
$$

provided we take $\delta(D) = \Delta(D)$.

- If $C_\Delta(\theta) = P$, the credible region *is* a confidence region.
- Otherwise, the credible region accounts for a priori uncertainty in θ—we *need* priors for this.

A Frequentist Confidence Region

Infer μ : $x_i = \mu + \epsilon_i;$ $p(x_i|\mu, M) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$



68% confidence region: $\bar{x} \pm \sigma/\sqrt{N}$

Monte Carlo Algorithm:

1. Pick a null hypothesis, $\mu = \mu_0$
2. Draw $x_i \sim N(\mu_0, \sigma^2)$ for $i = 1$ to N
3. Find $\bar{x}$; check if $\mu_0 \in \bar{x} \pm \sigma/\sqrt{N}$
4. Repeat $M >> 1$ times; report fraction (≈ 0.683)
5. *Hope result is independent of μ_0!*

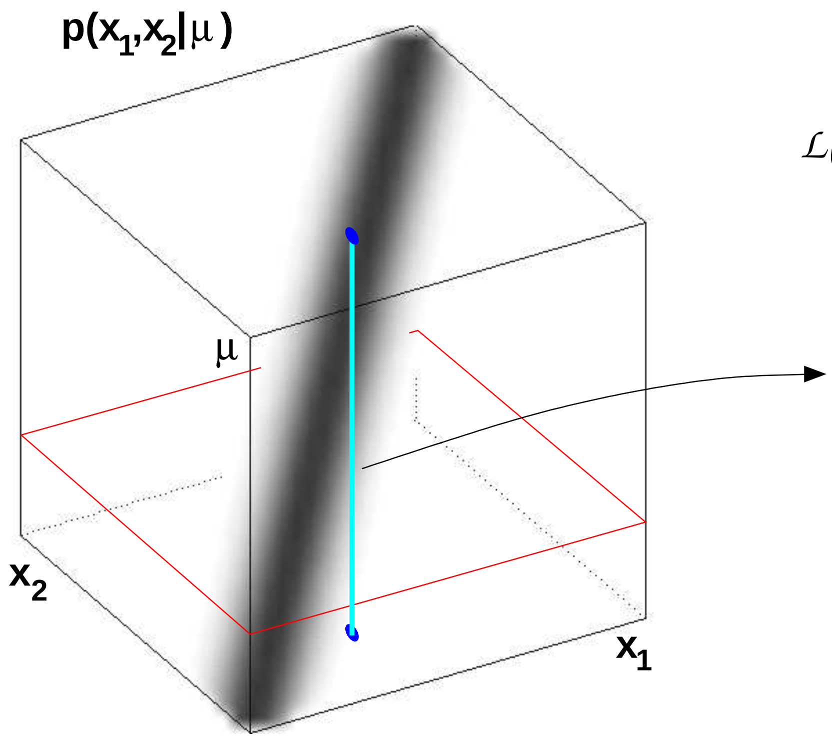
A Monte Carlo calculation of the N-dimensional integral:

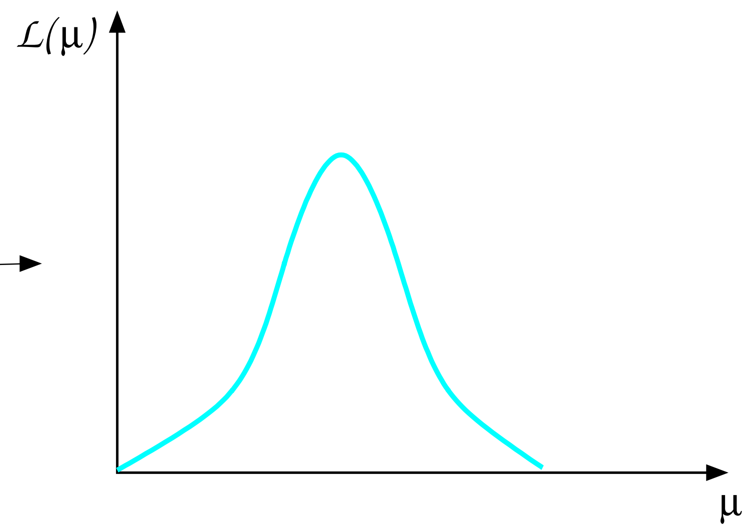
$$\int dx_1 \frac{e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \cdots \int dx_N \frac{e^{-\frac{(x_N-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \times [\mu_0 \in \bar{x} \pm \sigma/\sqrt{N}]$$

$$= \int d(\text{angles}) \int_{\bar{x}-\sigma/\sqrt{N}}^{\bar{x}+\sigma/\sqrt{N}} d\bar{x} \cdots \approx 0.683$$

A Bayesian Credible Region

Infer μ : Flat prior; $\mathcal{L}(\mu) \propto \exp\left[-\frac{(\bar{x}-\mu)^2}{2(\sigma/\sqrt{N})^2}\right]$





68% credible region: $\bar{x} \pm \sigma/\sqrt{N}$

68% credible region: $\bar{x} \pm \sigma/\sqrt{N}$

$$\frac{\int_{\bar{x}-\sigma/\sqrt{N}}^{\bar{x}+\sigma/\sqrt{N}} d\mu \exp\left[-\frac{(\bar{x}-\mu)^2}{2(\sigma/\sqrt{N})^2}\right]}{\int_{-\infty}^{\infty} d\mu \exp\left[-\frac{(\bar{x}-\mu)^2}{2(\sigma/\sqrt{N})^2}\right]} \approx 0.683$$

Equivalent to a Monte Carlo calculation of a 1-d integral:

1. Draw μ from $N(\bar{x}, \sigma^2/N)$ (i.e., prior $\times \mathcal{L}$)
2. Repeat $M >> 1$ times; histogram
3. Report most probable 68.3% region

This simulation uses hypothetical *hypotheses* rather than hypothetical *data*.