Introduction to Bayesian inference
Lecture 2: Key examples

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

CASt Summer School — 5 June 2014

Lecture 2: Key examples

1. **Simple examples**
   - Normal Distribution
   - Poisson Distribution
2. **Multilevel models for measurement error**
3. **Bayesian computation**

Key examples

1. **Simple examples**
   - Normal Distribution
   - Poisson Distribution
2. **Multilevel models for measurement error**
3. **Bayesian computation**

Supplement

- Binary classification with binary data
- Bernoulli, binomial, negative binomial distributions
- Parameter estimation & model comparison
- Likelihood principle
- Relationships between probability & frequency

Inference With Normals/Gaussians

Gaussian PDF

$$p(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{over } [-\infty, \infty]$$

Common abbreviated notation: $x \sim N(\mu, \sigma^2)$

Parameters

$$\begin{aligned} \mu &= \langle x \rangle \equiv \int dx\, x\, p(x|\mu,\sigma) \\ \sigma^2 &= \langle (x-\mu)^2 \rangle \equiv \int dx\, (x-\mu)^2\, p(x|\mu,\sigma) \end{aligned}$$

Gauss's Observation: Sufficiency

Suppose our data consist of N measurements, $d_i = \mu + \epsilon_i$.
Suppose the noise contributions are independent, and
$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

$$
\begin{aligned}
p(D|\mu, \sigma, M) &= \prod_i p(d_i|\mu, \sigma, M) \\
&= \prod_i p(\epsilon_i = d_i - \mu|\mu, \sigma, M) \\
&= \prod_i \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(d_i - \mu)^2}{2\sigma^2}\right] \\
&= \frac{1}{\sigma^N (2\pi)^{N/2}} e^{-Q(\mu)/2\sigma^2}
\end{aligned}
$$

Find dependence of Q on μ by completing the square:

$$
\begin{aligned}
Q &= \sum_i (d_i - \mu)^2 \qquad\qquad [\text{Note: } Q/\sigma^2 = \chi^2(\mu)] \\
&= \sum_i d_i^2 + \sum_i \mu^2 - 2\sum_i d_i \mu \\
&= \left(\sum_i d_i^2\right) + N\mu^2 - 2N\mu\overline{d} \qquad \text{where } \overline{d} \equiv \frac{1}{N}\sum_i d_i \\
&= N(\mu - \overline{d})^2 + \left(\sum_i d_i^2\right) - N\overline{d}^2 \\
&= N(\mu - \overline{d})^2 + Nr^2 \quad \text{where } r^2 \equiv \frac{1}{N}\sum_i (d_i - \overline{d})^2
\end{aligned}
$$

Likelihood depends on $\{d_i\}$ only through $\overline{d}$ and r:

$$\mathcal{L}(\mu, \sigma) = \frac{1}{\sigma^N (2\pi)^{N/2}} \exp\left(-\frac{Nr^2}{2\sigma^2}\right) \exp\left(-\frac{N(\mu - \overline{d})^2}{2\sigma^2}\right)$$

The sample mean and variance are *sufficient statistics*.

This is a miraculous compression of information—the normal dist'n is highly *abnormal* in this respect!

Estimating a Normal Mean

Problem specification

Model: $d_i = \mu + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$, σ is known $\rightarrow I = (\sigma, M)$.

Parameter space: μ; seek $p(\mu|D, \sigma, M)$

Likelihood

$$
\begin{aligned}
p(D|\mu, \sigma, M) &= \frac{1}{\sigma^N (2\pi)^{N/2}} \exp\left(-\frac{Nr^2}{2\sigma^2}\right) \exp\left(-\frac{N(\mu - \overline{d})^2}{2\sigma^2}\right) \\
&\propto \exp\left(-\frac{N(\mu - \overline{d})^2}{2\sigma^2}\right)
\end{aligned}
$$

"Uninformative" prior

Translation invariance $\Rightarrow p(\mu) \propto C$, a constant.
This prior is *improper* unless bounded.

Prior predictive/normalization

$$\begin{aligned} p(D|\sigma, M) &= \int d\mu\ C \exp\left(-\frac{N(\mu - \overline{d})^2}{2\sigma^2}\right) \\ &= C(\sigma/\sqrt{N})\sqrt{2\pi} \end{aligned}$$

... minus a tiny bit from tails, using a proper prior.

Posterior

$$p(\mu|D, \sigma, M) = \frac{1}{(\sigma/\sqrt{N})\sqrt{2\pi}} \exp\left(-\frac{N(\mu - \overline{d})^2}{2\sigma^2}\right)$$

Posterior is $N(\overline{d}, w^2)$, with standard deviation $w = \sigma/\sqrt{N}$.

68.3% HPD credible region for μ is $\overline{d} \pm \sigma/\sqrt{N}$.

Note that C drops out $\rightarrow$ limit of infinite prior range is well behaved.

Informative Conjugate Prior

Use a normal prior, $\mu \sim N(\mu_0, w_0^2)$.

Conjugate because the posterior turns out also to be normal.

Posterior

Normal $N(\tilde{\mu}, \tilde{w}^2)$, but mean, std. deviation *"shrink"* towards prior.

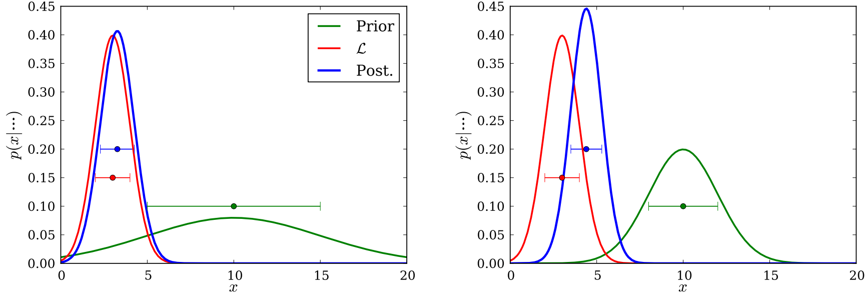
Define $B = \frac{w^2}{w^2 + w_0^2}$, so $B < 1$ and $B = 0$ when w_0 is large. Then

$$\begin{aligned} \widetilde{\mu} &= \overline{d} + B \cdot (\mu_0 - \overline{d}) \\ \widetilde{w} &= w \cdot \sqrt{1 - B} \end{aligned}$$

"Principle of stable estimation" — The prior affects estimates only when data are not informative relative to prior.

Conjugate normal examples:

- Data have $\overline{d} = 3$, $\sigma/\sqrt{N} = 1$
- Priors at $\mu_0 = 10$, with $w = \{5, 2\}$



Estimating a Normal Mean: Unknown σ

Supplement: Marginalize over σ $\rightarrow$ Student's t distribution

Gaussian Background Subtraction

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):

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

$\Rightarrow$ Background *subtraction* is a special case of background *marginalization*; i.e., marginalization "told us" to subtract a background estimate.

Recall the standard derivation of background uncertainty via "propagation of errors" based on Taylor expansion (statistician's *Delta-method*).

Marginalization provides a generalization of error propagation—without approximation!

Supplement: Handling σ uncertainty by marginalizing over σ; Student's t distribution

Bayesian Curve Fitting & Least Squares

Setup

Data $D = \{d_i\}$ are measurements of an underlying function $f(x;\theta)$ at N sample points $\{x_i\}$. Let $f_i(\theta) \equiv f(x_i;\theta)$:

$$d_i = f_i(\theta) + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma_i^2)$$

We seek learn θ, or to compare different functional forms (model choice, M).

Likelihood

$$\begin{aligned} p(D|\theta, M) &= \prod_{i=1}^{N} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{d_i - f_i(\theta)}{\sigma_i}\right)^2\right] \\ &\propto \exp\left[-\frac{1}{2}\sum_i \left(\frac{d_i - f_i(\theta)}{\sigma_i}\right)^2\right] \\ &= \exp\left[-\frac{\chi^2(\theta)}{2}\right] \end{aligned}$$

Bayesian Curve Fitting & Least Squares

Posterior

For prior density $\pi(\theta)$,

$$p(\theta|D, M) \propto \pi(\theta) \exp\left[-\frac{\chi^2(\theta)}{2}\right]$$

If you have a least-squares or χ^2 code:

- Think of $\chi^2(\theta)$ as $-2 \log \mathcal{L}(\theta)$.
- Bayesian inference amounts to exploration and numerical integration of $\pi(\theta)e^{-\chi^2(\theta)/2}$.

Important Case: Separable Nonlinear Models

A (linearly) separable model has parameters $\theta = (A, \psi)$:

- Linear amplitudes $A = \{A_\alpha\}$
- Nonlinear parameters ψ

$f(x; \theta)$ is a linear superposition of M nonlinear components $g_\alpha(x; \psi)$:

$$
\begin{aligned}
d_i &= \sum_{\alpha=1}^{M} A_\alpha g_\alpha(x_i; \psi) + \epsilon_i \\
\text{or} \qquad \vec{d} &= \sum_{\alpha} A_\alpha \vec{g}_\alpha(\psi) + \vec{\epsilon}.
\end{aligned}
$$

Why this is important: You can marginalize over A *analytically*
$\rightarrow$ *Bretthorst algorithm* ("Bayesian Spectrum Analysis & Param. Est'n" 1988)

Algorithm is closely related to linear least squares, diagonalization, SVD; for sinusoidal g_α, generalizes periodograms.

Poisson Dist'n: Infer a Rate from Counts

Problem:

Observe n counts in T; infer rate, r

Likelihood

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

Prior

Two simple standard choices (or conjugate gamma dist'n):

- 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

A 2-parameter family of distributions over nonnegative x, with shape parameter α and scale parameter s:

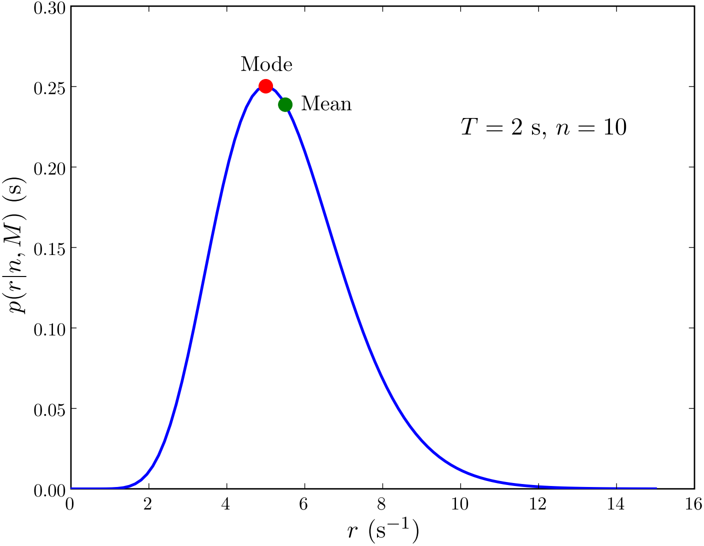
$$p_\Gamma(x|\alpha, s) \quad = \quad \frac{1}{s\Gamma(\alpha)} \left(\frac{x}{s}\right)^{\alpha-1} e^{-x/s}$$

Moments:

$$\mathrm{E}(x) = s\alpha \qquad \mathrm{Var}(x) = s^2\alpha$$

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)



Mode
Mean
$T = 2$ s, $n = 10$
$p(r|n, M)$ (s)
r (s^{-1})
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
2
4
6
8
10
12
14
16

The flat prior

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

Require (discrete) predictive distribution to be flat:

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

Useful conventions

- 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

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

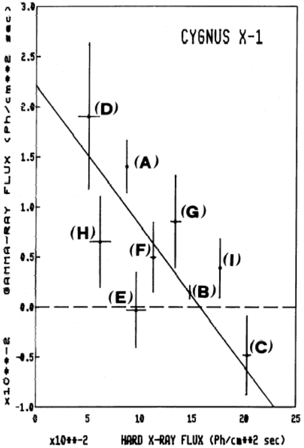
$$\begin{aligned}
\hat{b} &= N_{\rm off}/T_{\rm off}; \quad & \sigma_b &= \sqrt{N_{\rm off}}/T_{\rm off} \\
\hat{r} &= N_{\rm on}/T_{\rm on}; \quad & \sigma_r &= \sqrt{N_{\rm on}}/T_{\rm on} \\
\hat{s} &= \hat{r} - \hat{b}; \quad & \sigma_s &= \sqrt{\sigma_r^2 + \sigma_b^2}
\end{aligned}$$

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

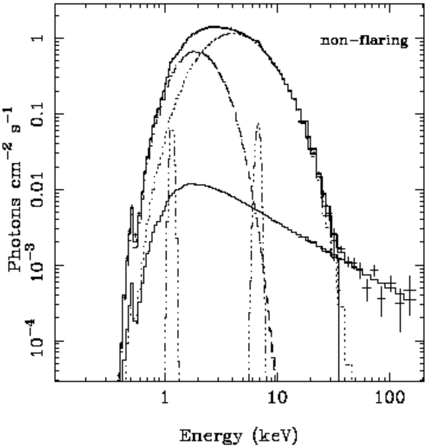
Examples

Spectra of X-Ray Sources

Bassani et al. 1989

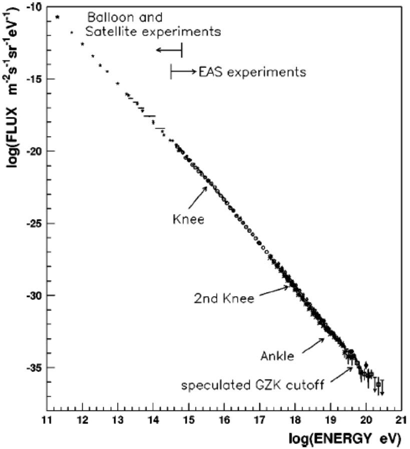


Di Salvo et al. 2001

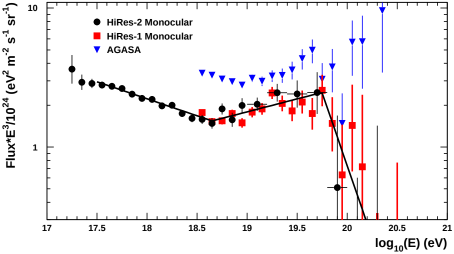


Spectrum of Ultrahigh-Energy Cosmic Rays

Nagano & Watson 2000



HiRes Team 2007



N is Never Large

Sample sizes are never large. If N is too small to get a sufficiently-precise estimate, you need to get more data (or make more assumptions). But once N is 'large enough,' you can start subdividing the data to learn more (for example, in a public opinion poll, once you have a good estimate for the entire country, you can estimate among men and women, northerners and southerners, different age groups, etc etc). N is never enough because if it were 'enough' you'd already be on to the next problem for which you need more data.

— Andrew Gelman (blog entry, 31 July 2005)

N is Never Large

Sample sizes are never large. If N is too small to get a sufficiently-precise estimate, you need to get more data (or make more assumptions). But once N is 'large enough,' you can start subdividing the data to learn more (for example, in a public opinion poll, once you have a good estimate for the entire country, you can estimate among men and women, northerners and southerners, different age groups, etc etc). N is never enough because if it were 'enough' you'd already be on to the next problem for which you need more data.

Similarly, you never have quite enough money. But that's another story.

— Andrew Gelman (blog entry, 31 July 2005)

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 $I_{\rm all} = (I_{\rm on}, N_{\rm off}, I_{\rm off})$:

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

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

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

Now marginalize over b;

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

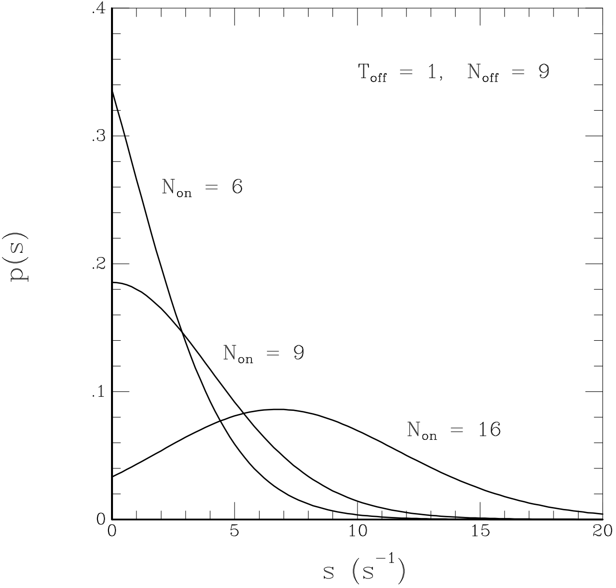
Expand $(s+b)^{N_{\rm on}}$ and do the resulting Γ integrals:

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

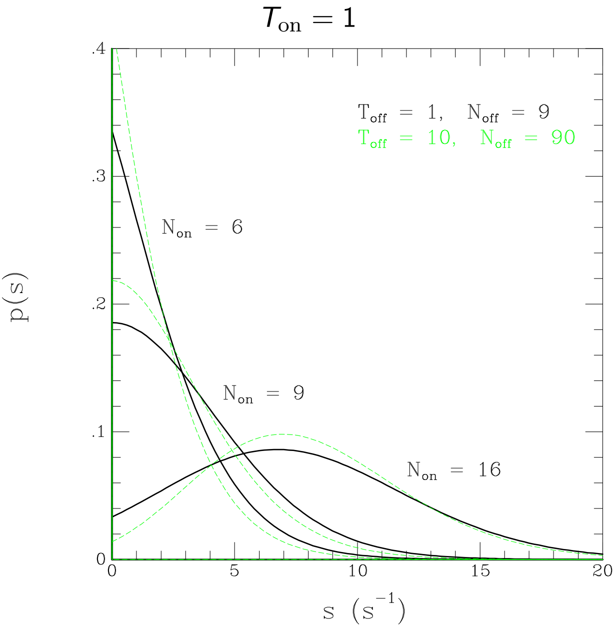
Posterior is a weighted sum of Gamma distributions, each assigning a different number of on-source counts to the source. (Evaluate via recursive algorithm or confluent hypergeometric function.)

Example On/Off Posteriors—Short Integrations

$T_{\rm on} = 1$



Example On/Off Posteriors—Long Background Integrations



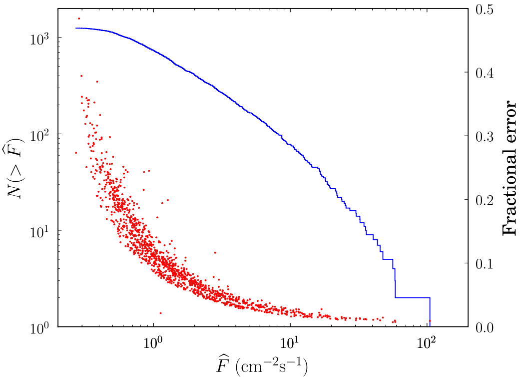
Supplement: Two more solutions of on/off problem (including data augmentation); multibin case

Recap of Key Ideas From Examples

- Sufficient statistic: Model-dependent summary of data
- Conjugate priors
- Marginalization: Generalizes background subtraction, propagation of errors
- Exact treatment of Poisson background uncertainty (don't subtract!)
- Likelihood principle
- Student's t for handling σ uncertainty

Key examples

Complications With Survey Data

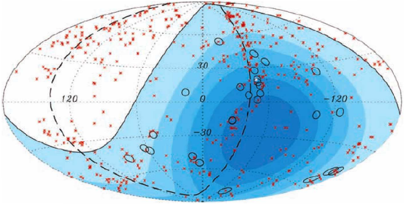


- *Selection effects* (truncation, censoring) — *obvious* (usually)
  Typically treated by "correcting" data
  Most sophisticated: product-limit estimators

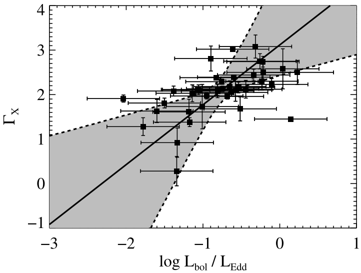
- *"Scatter" effects* (measurement error, etc.) — *insidious*
  Typically ignored (average out???)

Many Guises of Measurement Error

Auger data above GZK cutoff (PAO 2007; Soiaporn^{+} 2013)



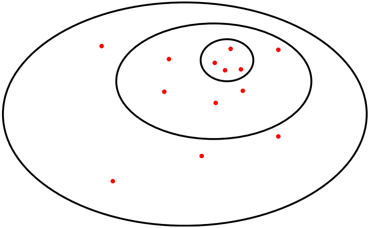
QSO hardness vs. luminosity (Kelly 2007, 2012)



Accounting For Measurement Error

Introduce latent/hidden/incidental parameters

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



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

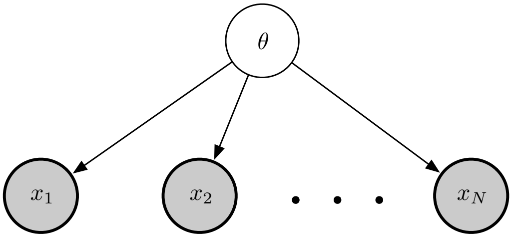
$$\mathcal{L}(\theta) \equiv p(\{x_i\}|\theta) = \prod_i f(x_i|\theta)$$

$$p(\theta|\{x_i\}) \propto p(\theta)\mathcal{L}(\theta) = p(\theta, \{x_i\})$$

(A *binomial point process*)

Graphical representation

- Nodes/vertices = uncertain quantities (gray → known)
- Edges specify conditional dependence
- Absence of an edge denotes *conditional independence*

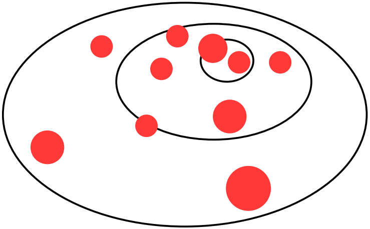


Graph specifies the form of the *joint distribution*:

$$p(\theta, \{x_i\}) \;=\; p(\theta)\, p(\{x_i\}|\theta) \;=\; p(\theta) \prod_i f(x_i|\theta)$$

Posterior from BT: $p(\theta|\{x_i\}) = p(\theta, \{x_i\})/p(\{x_i\})$

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



$\{x_i\}$ are now *uncertain (latent) parameters*
We should somehow incorporate $\ell_i(x_i) = p(D_i|x_i)$:

$$
\begin{aligned}
p(\theta, \{x_i\}, \{D_i\}) &= p(\theta)\, p(\{x_i\}|\theta)\, p(\{D_i\}|\{x_i\}) \\
&= p(\theta) \prod_i f(x_i|\theta)\, \ell_i(x_i)
\end{aligned}
$$

Marginalize over $\{x_i\}$ to summarize inferences for θ.
Marginalize over θ to summarize inferences for $\{x_i\}$.

Key point: *Maximizing over x_i and integrating over x_i can give very different results!*

To estimate x_1:

$$\begin{aligned} p(x_1|\{x_2,\ldots\}) &= \int d\theta\; p(\theta)\, f(x_1|\theta)\,\ell_1(x_1) \times \prod_{i=2}^{N} \int dx_i\; f(x_i|\theta)\,\ell_i(x_i) \\ &= \ell_1(x_1) \int d\theta\; p(\theta)\, f(x_1|\theta) \mathcal{L}_{m,\check{1}}(\theta) \\ &\approx \ell_1(x_1) f(x_1|\hat{\theta}) \end{aligned}$$

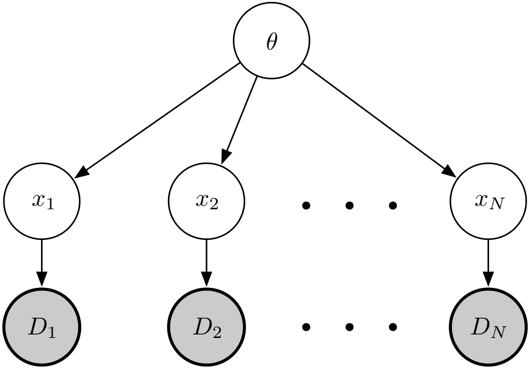
with $\hat{\theta}$ determined by the remaining data.

$f(x_1|\hat{\theta})$ behaves like a prior that shifts the x_1 estimate away from the peak of $\ell_1(x_i)$.

This generalizes the corrections derived by Eddington, Malmquist and Lutz-Kelker.

Landy & Szalay (1992) proposed adaptive Malmquist corrections that can be viewed as an approximation to this.

Graphical representation



$$
\begin{aligned}
p(\theta, \{x_i\}, \{D_i\}) &= p(\theta)\, p(\{x_i\}|\theta)\, p(\{D_i\}|\{x_i\}) \\
&= p(\theta) \prod_i f(x_i|\theta)\, p(D_i|x_i) \;=\; p(\theta) \prod_i f(x_i|\theta)\, \ell_i(x_i)
\end{aligned}
$$

A two-level *multi-level model* (MLM).

Bayesian MLMs in Astronomy

Surveys (number counts/"$\log N$–$\log S$"/Malmquist):

- GRB peak flux dist'n (Loredo & Wasserman 1998)
- TNO/KBO magnitude distribution (Gladman+ 1998; Petit+ 2008)
- MLM tutorial; Malmquist-type biases in cosmology (*Loredo & Hendry 2009* in *BMIC* book)
- "Extreme deconvolution" for proper motion surveys (Bovy, Hogg, & Roweis 2011)

Directional & spatio-temporal coincidences:

- GRB repetition (Luo+ 1996; Graziani+ 1996)
- GRB host ID (Band 1998; Graziani+ 1999)
- VO cross-matching (Budavári & Szalay 2008)

Time series:

- SN 1987A neutrinos, uncertain energy vs. time (Loredo & Lamb 2002)
- Multivariate "Bayesian Blocks" (Dobigeon, Tourneret & Scargle 2007)
- SN Ia multicolor light curve modeling (Mandel^{+} 2009, 2011)

Linear & nonlinear regression with measurement error:

- QSO hardness vs. luminosity (Kelly 2007, 2012)
- Dust SEDs (Kelly^{+} 2012)

More information:
`http://astrostatistics.psu.edu/su10/surveys.html`
Overview of MLMs in astronomy: arXiv:1208.3036
In progress: GPU software (Budavari, Kelly, TJL)

Key examples

1. Simple examples
   - Normal Distribution
   - Poisson Distribution
2. Multilevel models for measurement error
3. Bayesian computation

Statistical Integrals

Inference with independent data

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

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

Frequentist integrals

Find long-run properties of procedures via sample space integrals:

$$\mathcal{I}(\theta) = \int dx_1\, p(x_1|\theta) \int dx_2\, p(x_2|\theta) \cdots \int dx_N\, p(x_N|\theta) f(D, \theta)$$

Rigorous analysis must explore the θ dependence; rarely done in practice.

"Plug-in" approximation: Report properties of procedure for $\theta = \hat{\theta}$. *Asymptotically* accurate (for large N, expect $\hat{\theta} \to \theta$).

"Plug-in" results are easy via Monte Carlo (due to independence).

Bayesian integrals

$$\int d^m\theta \, g(\theta) \, p(\theta|M) \, \mathcal{L}(\theta) \quad = \quad \int d^m\theta \, g(\theta) \, \overbrace{q(\theta)}^{p(\theta|M)\,\mathcal{L}(\theta)}$$

- $g(\theta) = 1 \rightarrow p(D|M)$ (norm. const., model likelihood)
- $g(\theta) =$ 'box' $\rightarrow$ credible region
- $g(\theta) = \theta \rightarrow$ posterior mean for θ

Such integrals are sometimes easy if analytic (especially in low dimensions), often easier than frequentist counterparts (e.g., normal credible regions, Student's t).

Asymptotic approximations: Require ingredients familiar from frequentist calculations. Bayesian calculation is *not significantly harder* than frequentist calculation in this limit.

Numerical calculation: For "large" m (> 4 is often enough!) the integrals are often very challenging because of structure (e.g., correlations) in parameter space. This is usually pursued *without making any procedural approximations*.

Bayesian Computation

Large sample size: Laplace approximation

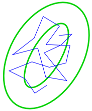
- Approximate posterior as multivariate normal $\rightarrow$ det(covar) factors
- Uses ingredients available in χ^2/ML fitting software (MLE, Hessian)
- Often accurate to $O(1/N)$

Modest-dimensional models ($d \lesssim 10$ to 20)

- Adaptive cubature
- Monte Carlo integration (importance & stratified sampling, adaptive importance sampling, quasirandom MC)

High-dimensional models ($d \gtrsim 5$)

- Posterior sampling — create RNG that samples posterior
- MCMC is most general framework — *Murali Haran's lab*



See SCMA 5 Bayesian Computation tutorial notes,
and notes from next week's sessions,
for more on MLMs & computation!

See online resource list for an annotated list
of Bayesian books and software