Statistical Cross-Identification: Commentary

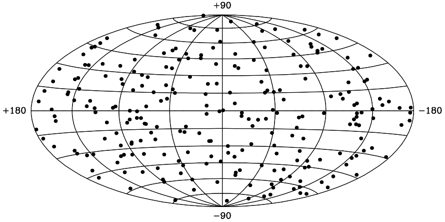
Tom Loredo
Dept. of Astronomy, Cornell University
http://www.astro.cornell.edu/staff/loredo/bayes/
Work with Kunlaya Soiaporn, David Ruppert,
David Chernoff, Ira Wasserman

SCMA V — 14 June 2011

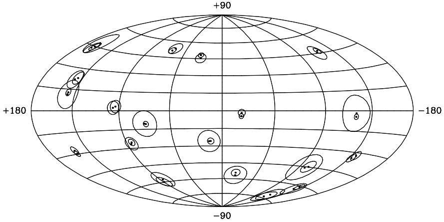
Just a Coincidence?

Do gamma-ray burst sources repeat?

250 GRB directions



Subset with neighbor within 3° (39)



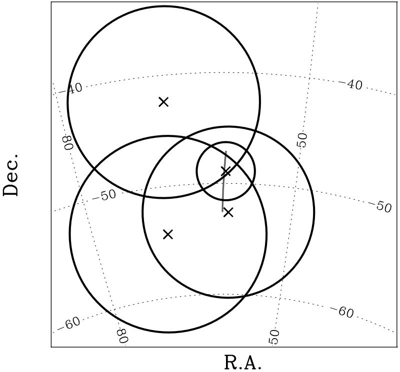
Later catalogs:

- 485 out of 1000 are close
- 2280 out of 2702 are close

BATSE GRB directions have 5–25° uncertainties

Error circles for 4 (3?) GRBs from 4B catalog

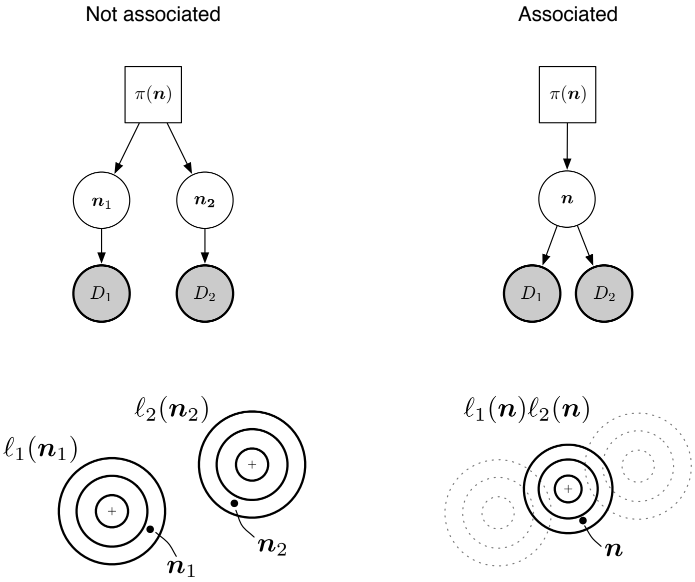
Seen in a 1.8 d period



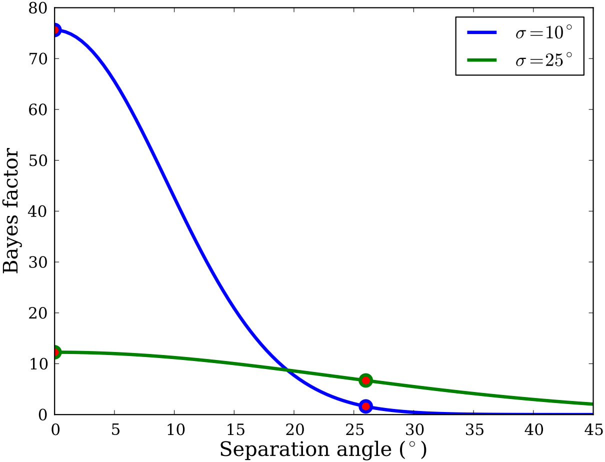
Graziani & Lamb 1998

Are these particular bursts from the same source?

Bayesian coincidence assessment, ca. 1995



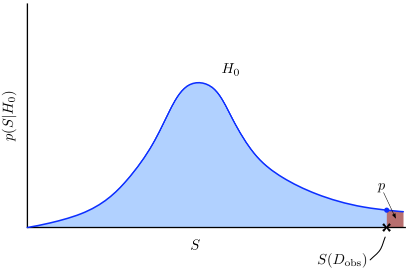
Bayes factor for two directions



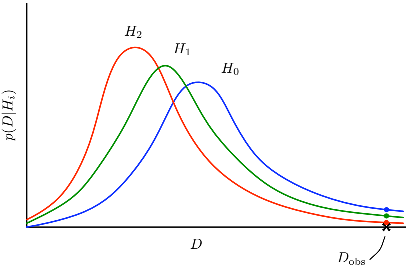
Nearest neighbor test: $p(< 26^\circ) = 0.05$; $p(< 0^\circ) = 0$

Bayes factor will never be compelling if $\sigma = 25^\circ$

Hypothesis testing with *p*-values



Model comparison with Bayes factors



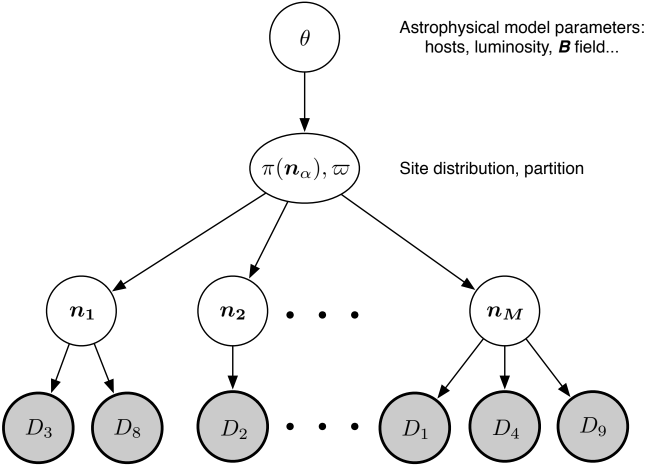
Challenge: Large hypothesis spaces

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; we must assign sensible priors to them, and sum over them (or at least all important ones).



Cornell group: Patch-based approximate counting; directional and spatio-temporal coincidences
Chicago group: Exhaustive enumeration; apply to small datasets

Why it's worth it

- No ambiguity in choice of proximity statistic
- Uncertain parameters (source extent, multiplicity, durations) handled by marginalization rather than optimization+adjustment for test multiplicity
- Bayes factors enable building multilevel models relating coincidences to astrophysically interesting quantities (e.g., source event rates; multiplicities)
- Bayes factors usefully quantify strength of an experiment

Bayesian Coincidence Assessment References

- Cornell group (1996) — GRB repetition; directional & time
  `http://adsabs.harvard.edu/abs/1996AIPC..384..477L`
- Graziani & Lamb (1996$^+$) — GRB repetition; SN assoc'n
  `http://adsabs.harvard.edu/abs/1996AIPC..366..196G`
  `http://adsabs.harvard.edu/abs/1998AIPC..428..161G`
  `http://adsabs.harvard.edu/abs/1999astro.ph..9025G`
  `http://adsabs.harvard.edu/abs/1999A%26AS..138..469G`
- Band (1998) — GRB no-host problem
  `http://adsabs.harvard.edu/abs/1998ApJ...493..555B`
- Budavári's team (2008$^+$) — General-purpose matching; VO
  `http://adsabs.harvard.edu/abs/2008ApJ...679..301B`
  astro-ph/1006.2096
- Cornell group (2011) — UHE cosmic ray source ID
  Kunlaya Soiaporn's poster

Why the gap?

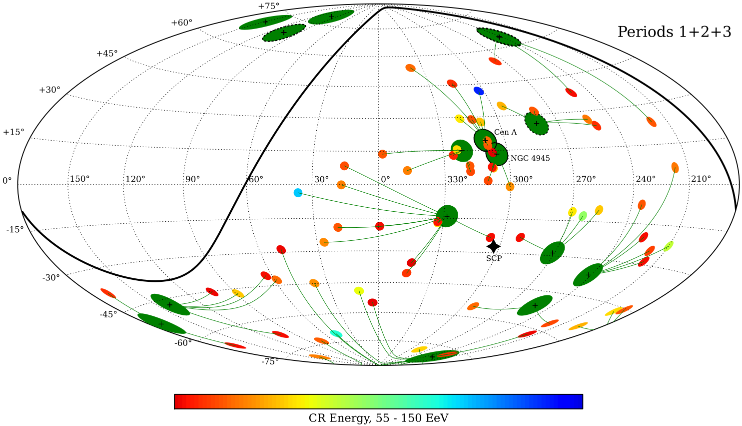
NVO 2005 proposal review:

“Arguments over the superiority of Baysean [sic] statistical techniques are nothing new: the committee doubted any real, practical advantages to the statistical approaches described. It seems like such a capability would not be much more than a ‘few-liner’ addition to XMatch. . . ”

Pierre Auger Observatory UHE Cosmic Rays

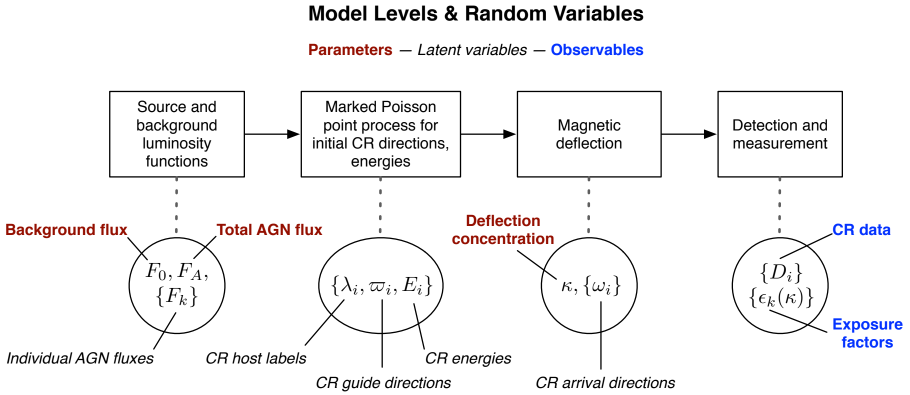
69 UHE CRs from PAO

17 AGN from a volume-complete survey to 15 Mpc



Arcs connect each CR to its nearest AGN

Associating UHE CRs and AGN



Many important uncertainties accounted for via marginalization

Unavoidable subjectivity: Choice of candidate source population

Kunlaya's algorithm:
Gibbs sampling + Chibb's marginal likelihood estimator