Ordering of Trials

Jim Linnemann
MSU
Banff Birs Statistics
July 14, 2010

Look Elsewhere

= Trials Effects

= Multiple Comparisons

use this to talk to statisticians

Multiple discoveries possible?

False Discovery Rate

Look Elsewhere Problems

Worst: *unknown* number of trials

- blind analysis
- tuning sample, then freeze cuts

Next worst:

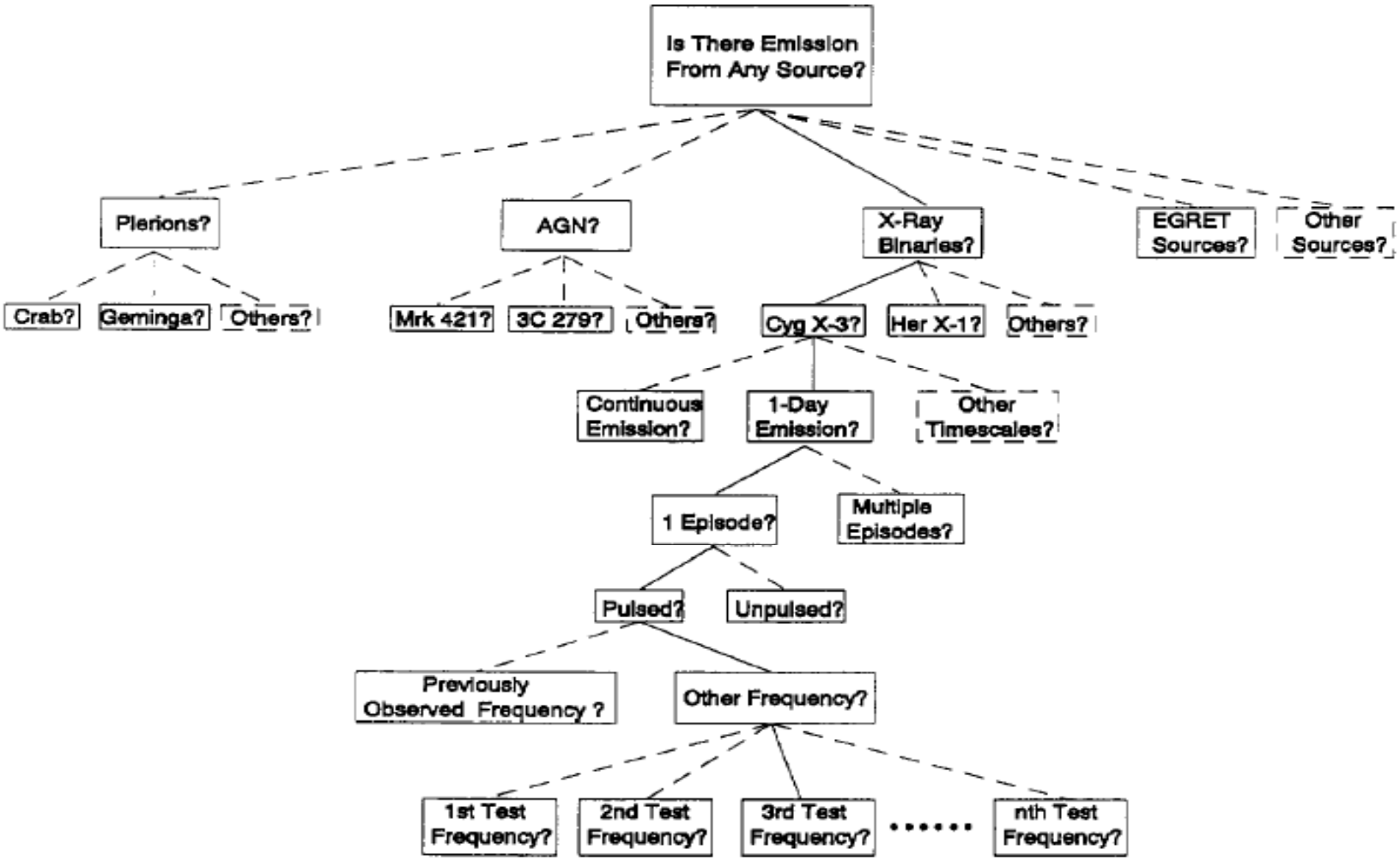
loss of power due to large Ntrials

eg, lots of places to look on sky

How to "spend" trials

Importance ordering: write out a protocol

S.D. Biller / Astroparticle Physics 4 (1996) 285–291



How to order?

Your (collaboration's) choice:

Physics interest

Prior probability

MC: expected sensitivity

Result of Ordering

First hypothesis:	1 trial (best sensitivity)
2nd	2
..	
Nth	N (full Bonferroni penalty)

On average, ½ the trials

Only “last” searches pay the full price

Trials Degrade Apparent Significance

Nominal significance must be de-rated by trials

$p_{corrected}$ ~ N $p_{nominal}$ if you looked at N plots

$$\int_{\sigma}^{\infty} Gau(0,1) \;=\; N \int_{\sigma_{nom}}^{\infty} Gau(0,1)$$

To achieve significance σ after trials correction

σ =	2	**3**	4	5	(corrected σ desired)

requires larger σ_{nom}(N) before trials correction for N trials

N	equivalent observed significance σ_{nom}=			
10	2.9	**3.7**	4.6	5.5
50	3.4	**4.1**	4.9	5.7
400	3.9	**4.5**	5.3	6.1
16K	4.7	**5.3**	5.9	6.7

bigger change for smaller corrected σ

Details: Bonferroni Correction Math:

Derive: $p_{corrected} \sim q = N \times p_{nominal}$

Exact Binomial probability for ≥1 of N found above p_n :

$$p_c = 1-(1-p_n)^N = (1-e^{-q}) + O(\frac{q^2}{N}) \approx q\,(1-\frac{q}{2})$$

$p_c = q$ for N=1, and $p_c \sim q$ for $q << 1$

Np=q sufficient for p_c of 2 σ or more

p_c=q=Np in terms of σ:

$$p_c = \int_{\sigma}^{\infty} Gau(0,1) \quad = q = N \int_{\sigma_{nom}}^{\infty} Gau(0,1)$$