101 Ways to Calculate Significance

Jim Linnemann

Michigan State University &

Los Alamos National Laboratory

July 11, 2003

Goal

What Can HEP and Astrophysics Practice Teach Each Other

Astrophysics:

- aims at simple formulae (very fast)
- calculates Sigmas directly
- hope it's a good formula

HEP:

- calculates probabilities by MC (general; slow)
- translates into Sigmas for communication
- loses track of analytic structure

Report at PHYSTAT2003 (Sept)

Executive (Jordanian) Summary

- For high and moderate Non, Noff, Li Ma Eq 17 fine
  - Anything works for Crab, but not for short GRB's
  - Li Ma Eq 9 not too bad
  - Bad formula typically overestimates significance

- Probably should use Binomial Test for small N
  - Optimal Frequentist, and Plausible Bayesian, Technique
  - want some MC confirmation
  - numerically, more work than Li Ma Eq 9
- Interesting relations exist among methods
  - Bayes with Gamma (not Gaussian) = Binomial
    - And same as Alexandreas et. al. (possibly within a constant factor)…
  - Li Ma Eq 9 = Binomial for large N
- Fraser-Reid Approximation Promising but not done

Significance

- Z value: ~ Normal(0,1) (Milagro; Li Ma)
- The art is to pick a good variable for this

More Generally:

- P(more extreme "signal" | background)
  - Assume Null Hypothesis: background only
  - Translate probability p into Z value by

$$Z = \Phi^{-1}(p); \quad \Phi = \int_{-\infty}^{Z} e^{-t^2/2} dt$$

$$Z \approx \sqrt{u - Ln\, u} \approx \sqrt{-Ln\, p}, \quad u = -2Ln\left(p\sqrt{2\pi}\right)$$

Prospective vs. Observed Significance

- This discussion: Observed Significance
  - Post-hoc: (after data)
- Prospective Observability (before data) involves more :
  - definition of Z, as for post-hoc;   but also:
  - Choice of Zmin   = max P(observed|background)
    - Very Similar to post-hoc: Zmin to make observation claim
  - Consideration of probability of meeting criterion
    - Simplest calculation:
      - Non = $\mu_{on} + \alpha\, \mu_{off}$; Noff = μ_{off} (ignores fluctuations)
      - Significance for Expected Conditions
      - Optimistic: crudely, ½ time less signal; or ½ time more background!
    - Better: Source Strength for 50% probability of observation? 90% ?
      - More related to Lazar's "upper bound" criterion
      - Similar discussions in HEP literature

Backgrounds in Astro and HEP

- Astrophysics: on-source vs. off-source
  - side observation with **α = Ton/Toff** (sensitivity ratio)
  - b = α Noff; db = α $\sqrt{\text{Noff}}$
  - $\alpha = (db)^2 / b$ (deduced from above)

- HEP: estimate background in defined signal region
  - Sometimes a sideband measurement, like Astrophysics
  - Often a MC estimate; rescaled to signal sensitivity
  - More often a sum of terms of both types
  - b ± db db: uncertainties in quadrature
  - **$\alpha = (db)^2 / b$** I'll use as a definition of effective α

Can apply astrophysics formulae

Li and Ma Equations

$$Z = S/\ \sigma(S)$$

$S = Non - b$ $\qquad$ $b = \alpha\ Noff$

N is observation; b is background estimate

Eq 5: Var(S) = Var(Non) + Var(b) = Non + α^2 Noff

Ignores key null hyp constraint: $\mu_{on} = \alpha\ \mu_{off}$ (anti-signal bias!)

Eq 9: Var(S) = α (Non + Noff)

Obeys constraint; uses Non and Noff to estimate μ_{off}

Eq 17: Log Likelihood Ratio (Wilks' Theorem)

$$Z = \sqrt{2}\sqrt{x \bullet Ln[\left(\tfrac{1+\alpha}{\alpha}\right)\left(\tfrac{x}{x+y}\right)] + y \bullet Ln[(1+\alpha)\left(\tfrac{y}{x+y}\right)]}$$

x = Noff; y = Noff

Li and Ma Variant

- Apply null by using only Noff to estimate Var(Non) and Var(b)
  - Obviously, bad if $\alpha > 1$
- Eq 5c: Var(S) = $\alpha\,(1+\alpha)$ Noff
- Get Eq 9 if use both (Max Likelihood)

Other Frequentist Methods

Ignoring uncertainty in b:

- S/√b Li Ma 10a
- Poisson(≥Non|b) (often much better)
- Feldman & Cousins? confidence limits!
  - For significance, just Poisson(≥ k|b), I believe

Using Uncertainty in b:

- b + db instead of b in above (I've seen it!)
- Near-Constant Variance (Zhang and Ramsden)

$$Z_{VS} = \frac{2}{\sqrt{1+\alpha}}\left(\sqrt{x+\frac{3}{8}} - \sqrt{y+\frac{3}{8}}\right)$$

- Fraser Reid
- Binomial Test

Fraser and Reid

- Interesting approximate method (last 15 yr)
- Significance from likelihood curve
  - Combine Z(Likelihood Ratio), Z(t/σ)
  - correct each other to $O(n^{-1.5})$
  - One version: improved Z value
- Redo algebra for each new kind of problem
  - I'm still working to apply it to Non, Noff fully
- Fast & simple numerically to apply formula

Binomial Test For Ratio of Poisson Means

(Compare means for on, off measurements)

- UMPU (Uniformly Most Powerful Unbiased)
  – **If best test, probably it's using the best variable**
- Holds k = Non+Noff fixed (nuisance parameter)
- Test is PrBinomial($\geq$Non | p,k), p = $\alpha/(1+\alpha)$
- Not in common use; probably should be

  Known in HEP and Astrophysics: not as optimal, nor standard procedure

  – Zhang and Ramsden claim too conservative for Z small? Even if true, we want Z > 4

  Experimental Astronomy 1 (1990) 145-163; I have pdf

  – Closed form in term of special functions, or sums
    - Applying for large N requires some delicacy; **slower than Eq 17**!
- Gaussian Limit of Binomial Test is Li Ma Eq 9!

Bayesian Methods

- Allow for correlations among background contributions (MC integration)
- Extension to efficiency, upper limits natural
  - In common use in HEP
    - Cousins & Highland "smeared likelihood" efficiency
- **Predictive Posterior** (after background measurement)
  - Natural avenue for connection with p-values

    But: typical Bayes analysis isn't significance, but odds ratio
  - Truncated Gaussian often used to represent db
  - A flat prior for background, gives gamma for db
  - **P value calc using gamma: (same(?) as Alexandreas)**
    - **same as Frequentist Binomial Test**

Comparing the Methods

- Some test cases from literature
  - Range of Non, Noff values
  - Different α values
- Color Code Accuracy
  - Assume Frequentist Binomial as Gold Standard
    - May change after I've run Monte Carlo

	Top 1	Top 2	Top 3	Crab X2>2.5	Crab > 5	Whipple	Hegra	Alexandr	Fake	Zhang1	Zhang2
Non	9	17	6	2,119,449	167,589	498,426	523	4	200	67	50
Noff	17.83	40.11	18.76	23,671,193	1,864,910	493,434	2327	5	10	15	55
alpha	0.2132	0.0947	0.0692	0.0891	0.0891	1.000	0.167	0.2	10.0	2.0	0.5
Nb	3.8	3.8	1.3	2,109,732	166,213	493,434	388.6	1.0	100.0	30.0	27.5
Ns = Non - Nb	5.2	13.2	4.7	9717	1376	4992	134.4	3.0	100	37	22.5
Sigma(Nb)	0.9	0.6	0.3	433.6	121.7	702.4	8.1	0.45	31.6	7.75	3.71
Sigma/Nb	0.237	0.158	0.231	0.000206	0.000732	0.00142	0.0207	0.447	0.316	0.258	0.135
Reported p	2.70E-02	2.00E-06	3.00E-03								
Reported S	1.9	4.6	2.7	6.4	3.2	5.0	5.9			3	3
Li Ma:											
Eq 5	1.66	3.17	1.90	6.397	3.22	5.01	5.54	1.46	2.89	3.28	2.82
Eq 9	2.17	5.67	3.59	6.409	3.23	5.01	6.16	2.24	2.18	2.89	3.11
Eq 17	1.99	4.57	2.81	6.405	3.23	5.01	5.93	1.95	2.38	3.04	3.02
Eq 5c	2.42	6.47	3.99	6.410	3.23	5.03	6.31	2.74	3.02	3.90	3.50
Eq 10 a = S/Sqrt B	2.67	6.77	4.12	6.69	3.38	7.11	6.82	3.00	10.00	6.76	4.29
Eq 10a + 1 sigma	2.40	6.29	3.72	6.69	3.37	7.10	6.75	2.49	6.72	6.02	4.03
Poisson	2.14	4.87	2.84	6.69	3.37	7.09	6.44	2.08	7.72	5.76	3.80
Poisson + 1 sigma	1.64	4.47	2.46	6.386	3.07	6.09	6.01	1.56	5.51	4.24	3.04
Binomial, Bayes prec	1.82	4.46	2.63	6.4048	3.23	5.01	5.93	1.66	2.20	2.89	2.93
Bayes Flat											
Bayes Poisson num	1.82	4.46						1.66	2.20		
Bayes Gauss num	1.94	4.55	2.71	6.4044	3.23	5.02	5.93	1.88	2.90	3.44	3.06
Square Root 3/8	1.98	4.22	2.66	6.4032	3.23	5.01	5.86	1.93	2.39	3.07	3.00
Fraser-Reid 1	2.14	4.87	2.84				6.44	2.07	8.95	5.76	3.80
Fraser-Reid 2	1.11	1.93	1.05		0.96	5.02	2.51	1.20			
Fraser-Reid 3	2.34	5.02	3.10				6.48	2.38			
Fraser-Reid 4											
Monte Carlo											
> .4 sigma high	correct										
> .4 sigma low	nearly correct										
correct approximation											
poor approximation											
input value											

What is a Good (Z) Variable?

Standard Method of MC Testing a Variable:

- "self-test": compare Z with distribution of statistic for MC assuming background only
  - i.e. convert back from Z to probability
    - **Good if PrMC(Z>Zo) = PrGauss(Z>Zo)**
  - Intuition: want fast convergence to Gaussian

Why not just compare with "right answer"?

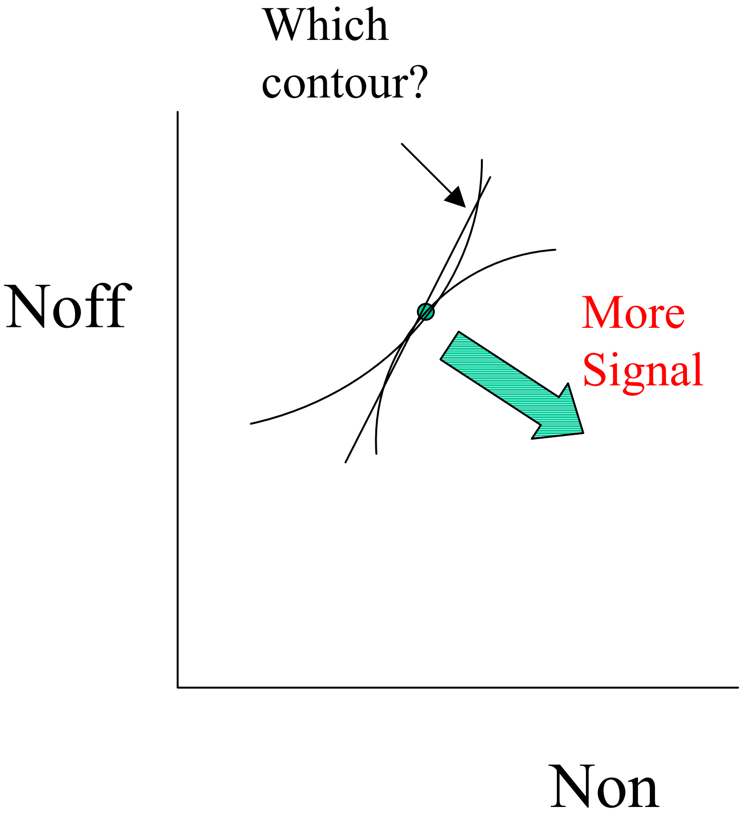
- Variables all supposed to give same Z, right?

But it's not really well-defined!

What is a Bigger Deviation? Part of Significance Definition!

- Measure Non, Noff = (x,y)
- Which values are worse?
  - Farther from line x = α y?
    - Angle? Perpendicular?
  - Larger s = x - α y?
- Trying to order 2-dim ∞ set!
  - Points on (x,y) plane
  - Nuisance parameter bites again
- Statistics give different metrics

  contours of equal deviation
- Convergence (to Gaussian)?
  - Perhaps for large N?
  - Enough peaking so overlapping regions dominate integrals?



Thank you Milagro!

Especially Gus and Jordan for making it possible

- I've Learned a Lot (Thanks for explaining!)
- Stimulating Company
- Excellent Surroundings
- Chance to work on some long-deferred things
- Interesting Experiment
- Hope I've contributed something useful!
- I also hope to find a way to continue…

A Prickly Problem
not to everyone's taste…

- What is Significance?
- Li and Ma Equations
- Frequentist Methods
- Bayesian Methods
- What is Significance, Really?
- To Do

Conclusions

- Bad formula typically overestimates significance
- For the Crab, any formula will do
  - Not true for GRB's with smaller Non, Noff
    - LR quite good, though maybe Binomial better
- Several interesting relationships among methods
- Fraser Reid remarkably good for P(n|b)
  - Haven't deciphered for interesting case (Non,Noff)
- Binomial Test should be used more

To Do

- Finish algebra for comparison of Bayes Gamma and Frequentist Binomial
- Monte Carlo Tests
- Fraser Reid Approximation to full problem
  - Simpler numerics, if it works!