Significance In the On-Off Problem

Jim Linnemann

Michigan State University

Banff/Birs

July 12, 2010

What Can HEP and Astrophysics Practice Teach Each Other?

Astrophysics **(especially γ ray)**

aims at simple formulae (very fast)

calculates σ's directly (Asymptotic Normal)

hope it's a good formula

HEP **(especially Fermilab practice)**

calculates probabilities by MC (general; slow)

translates into σ's for communication

loses track of analytic structure

Cousins, Linnemann, Tucker, NIM A 595 (2008) 480-501

Observed vs. Prospective Significance

- This discussion: Observed Significance (of my data)
  - Post-hoc: (after data)
  - Need definition of Z
  - **Choice of Zmin for observational claim**

    = max P(observed|background)
- Prospective Observability (before data, to optimize expt.)
  - Should consider **<u>Pr</u> (Z > Zmin) (making observational claim)**

Backgrounds in Astro and HEP

- Astrophysics: **On**-source vs. **Off**-source
  - side observation with **τ = Toff/Ton (sensitivity ratio)**

$$\hat{\mu}_b = n_{off} / \tau;$$

- HEP: estimate background in defined sideband region

  τ is ratio of signal and sideband region

Li and Ma (Gamma Ray)

$$Z_{PL} = \sqrt{2}\sqrt{x \bullet Ln[(1+\tau)\left(\tfrac{x}{x+y}\right)] + y \bullet Ln[(\frac{1+\tau}{\tau})\left(\tfrac{y}{x+y}\right)]}$$

x = Non; y = Noff

Generic test for composite hypothesis

+ Wilks' Theorem (conditions not satisfied)

Binomial Proportion Test: Ratio of Poisson Means

P-value = Pr Binomial(≥Non | p, k) where p = α/(1+α)

$$p - value = \sum_{j=x}^{k} \binom{k}{j} p^{j} (1-p)^{k-j}$$

Holds k = Non + Noff fixed (k a nuisance parameter)

UMPU (Uniformly Most Powerful Unbiased)

for **Composite Hypothesis test $\mu_{on} / \alpha\, \mu_{off} > 1$**

Optimal? Not continuous—issues for small n

Not in common use; probably should be

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

- Zhang and Ramsden claim too conservative for Z small
  Even if true, we want Z > 4
- Closed form in term of special functions, or sums
  - Applying for large N requires some delicacy; **ZPL**

Bayesian Methods

- In common use in HEP
  - Cousins & Highland "smeared likelihood" efficiency

- **Predictive Posterior** (after background measurement)

  P(Non | Noff) (integrate posterior μ_b)

  A flat prior for background, gives Gamma dist. for p(μ_b | Noff)

  P value calc using Gamma: **(also Alexandreas--Astro)**

  IDENTICAL to Frequentist Binomial Test

Predictive Posterior Bayes P-value (HEP)

In words: tail sum averaged over Bayes posterior for mean

or: integrate before sum

$$P-value\,(x,y) = \sum_{j=x}^{\infty} p(j\,|y)$$

$$p(j\,|\,y) = \int p(j\,|\,\mu)\,p(\mu\,|\,y)\,d\mu$$

$$p(j\,|\,\mu) = \frac{\mu^{j} e^{-\mu}}{j!}$$

$$p_{\Gamma}(\mu\,|\,y) = \frac{\beta^{y} e^{-\beta}}{y!}, \qquad \beta = \mu/\alpha$$

$$p_{N}(\mu\,|\,y) = Normal\,[(\mu - b)/\delta b]$$

Comparing the Methods

Some test cases from published literature

And a few artificial cases

Range of Non, Noff values

Different τ values (mostly > 1)

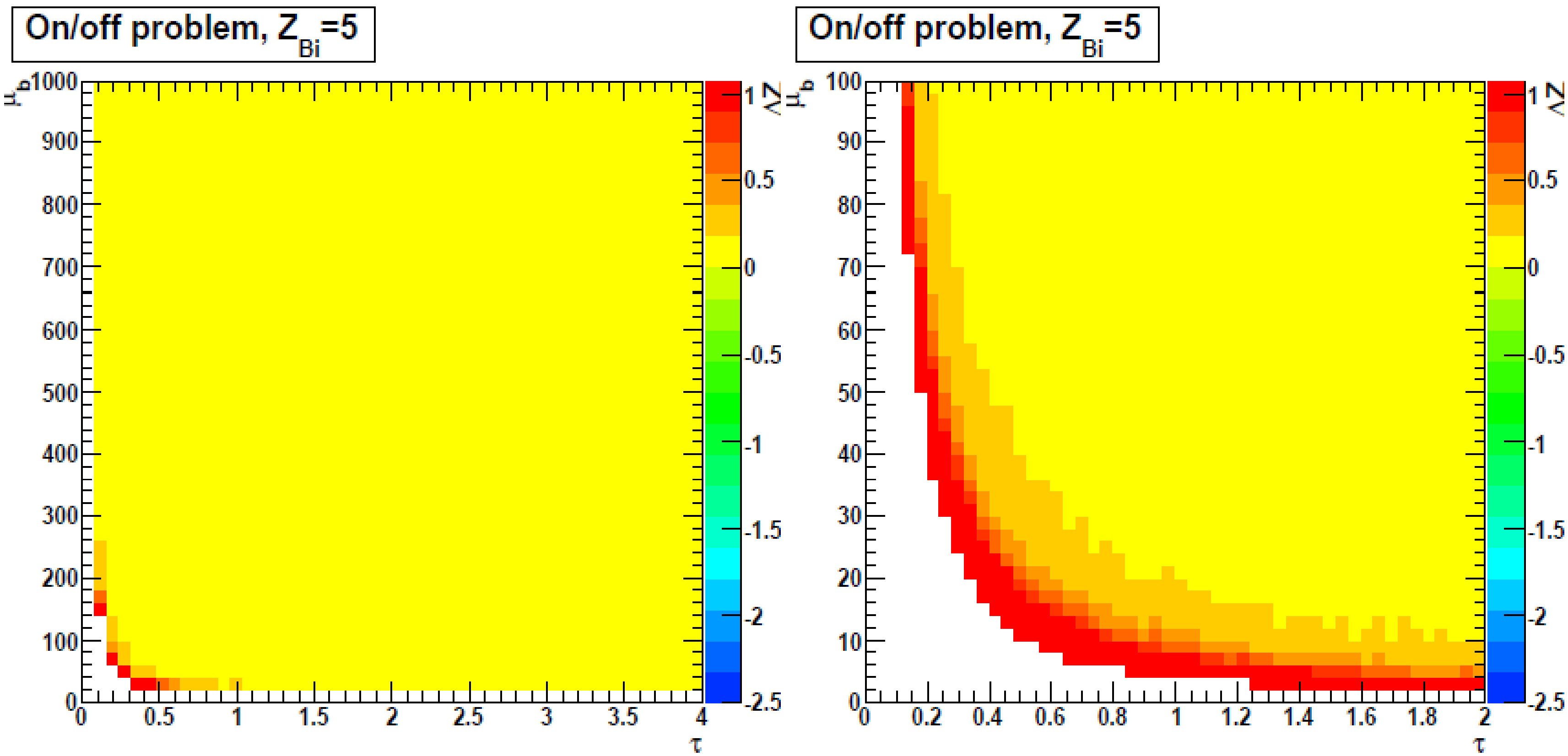
Can show some approximate Z's strictly > others

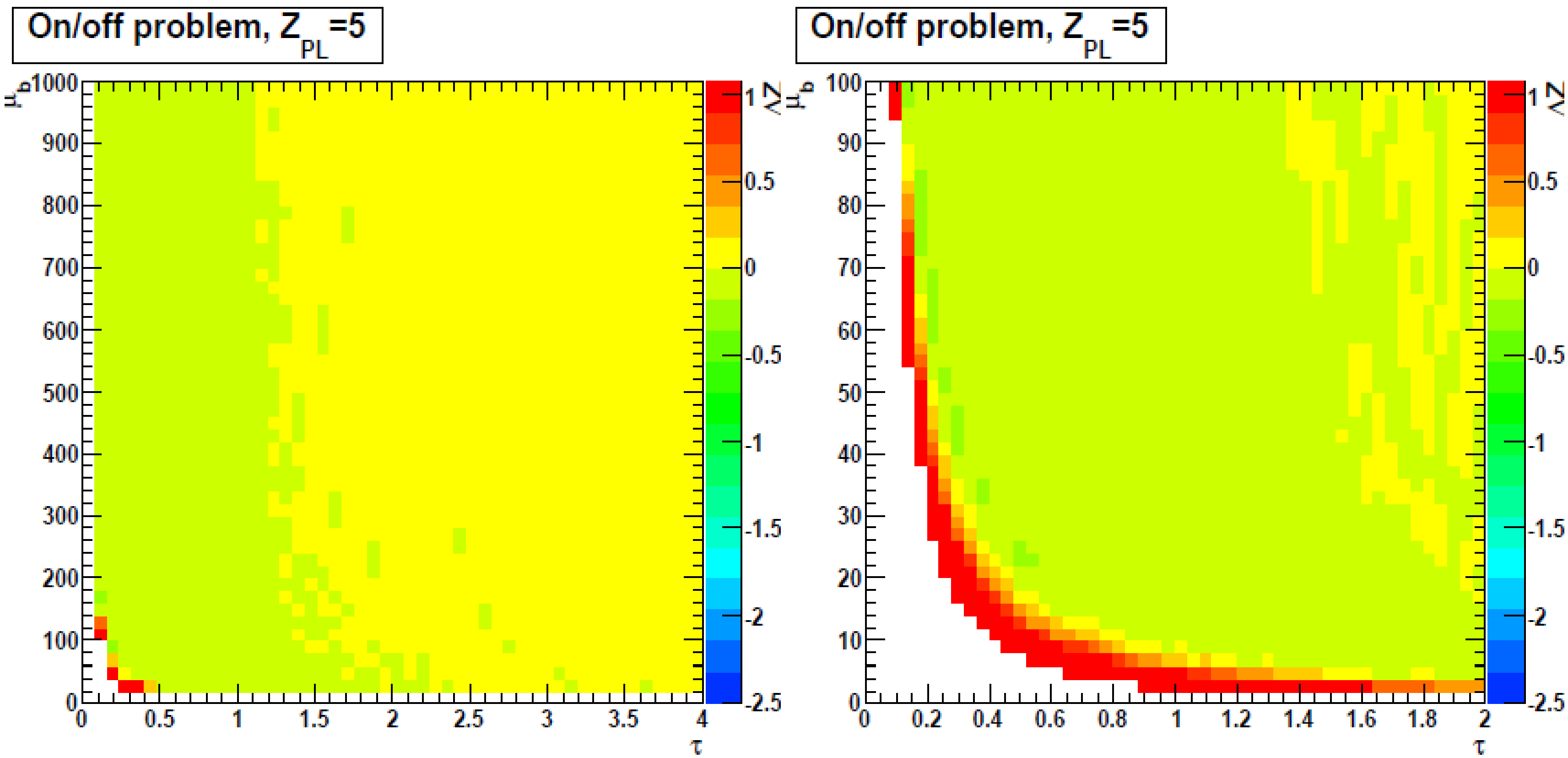
including popular shortcut formulas

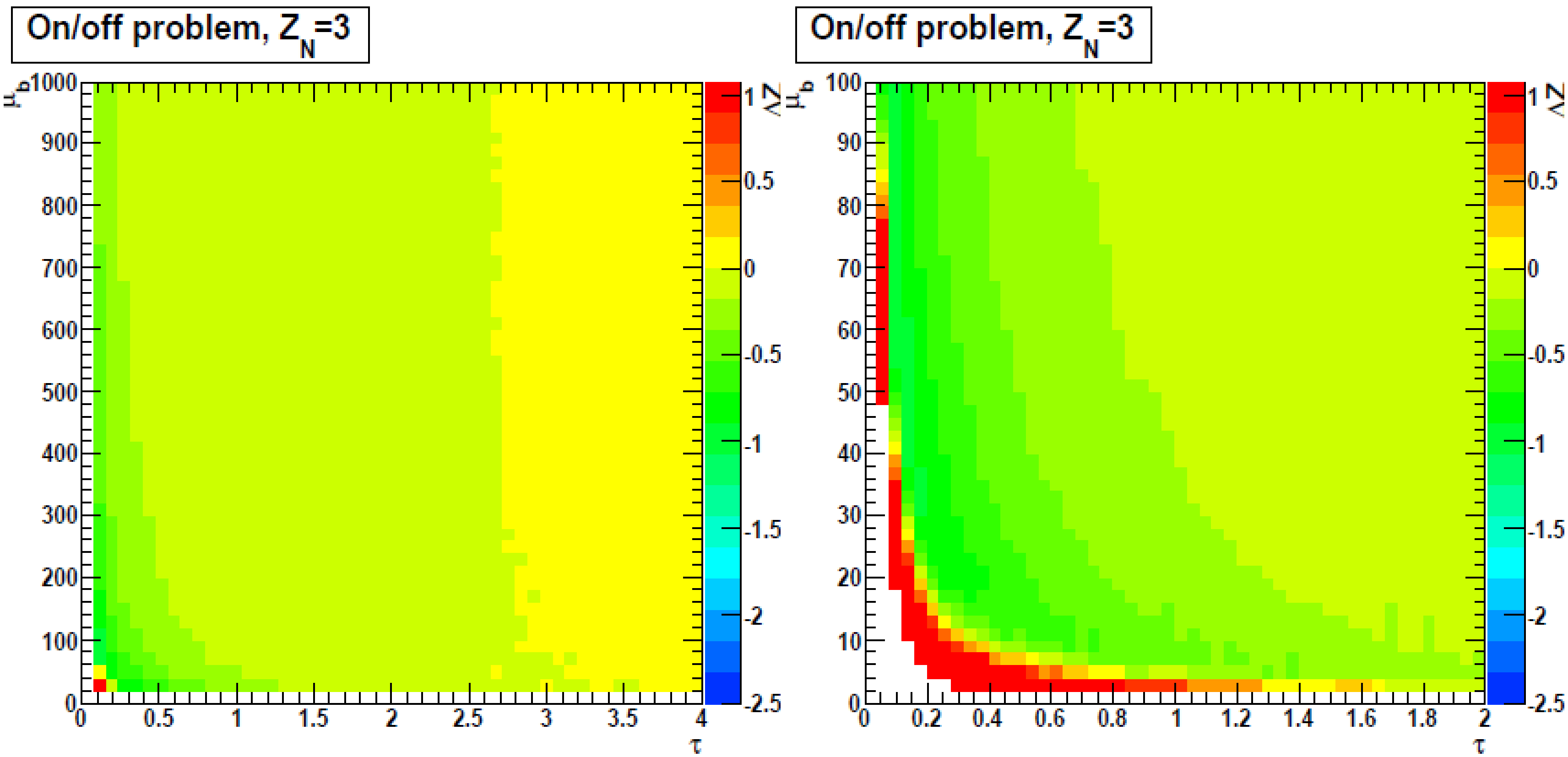
Others cross over as τ varies

Coverage Calculations (Tucker, Cousins)

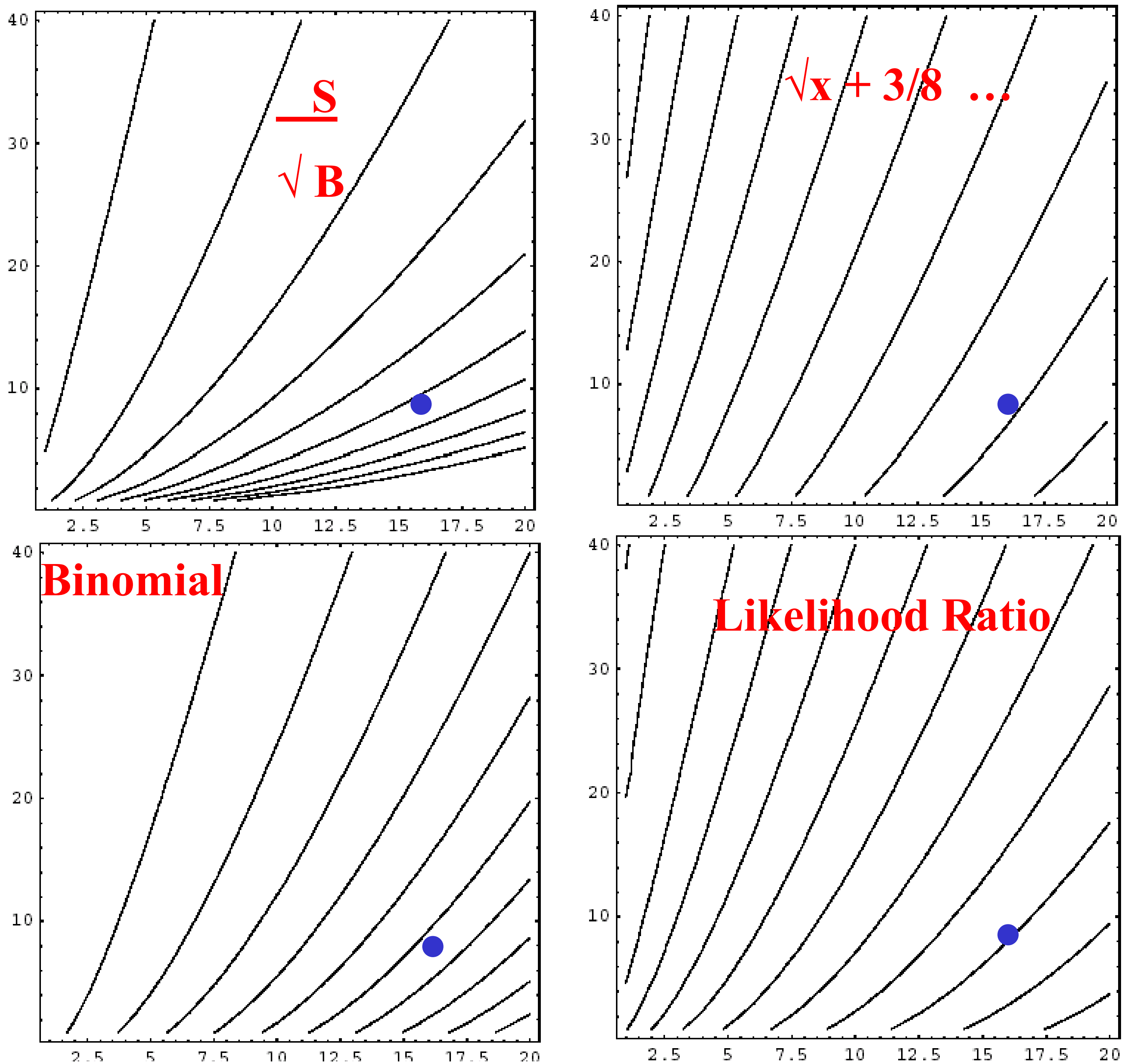
Reference	[40]	[41]	[42]	[43]	[44]	[44]	[45]	[46]
n_{on}	**4**	**6**	**9**	**17**	**50**	**67**	**200**	**523**
n_{off}	**5**	18.78	17.83	40.11	**55**	**15**	**10**	**2327**
τ	**5.0**	14.44	4.69	10.56	**2.0**	**0.5**	**0.1**	**5.99**
$\hat{\mu}_{\mathrm{b}}$	1.0	**1.3**	**3.8**	**3.8**	27.5	30.0	100.0	388.6
$s = n_{\mathrm{on}} - \hat{\mu}_{\mathrm{b}}$	3.0	4.7	5.2	13.2	22.5	37	100	**134**
σ_{b}	0.447	**0.3**	**0.9**	**0.6**	3.71	7.75	31.6	8.1
$f = \sigma_{\mathrm{b}}/\hat{\mu}_{\mathrm{b}}$	0.447	0.231	0.237	0.158	0.135	0.258	0.316	0.0207
Reported p		**0.003**	**0.027**	**2E−06**				
Reported Z		2.7	1.9	**4.6**				**5.9**
See conclusion								
$Z_{\mathrm{Bi}} = Z_{\Gamma}$ binomial	**1.66**	**2.63**	**1.82**	**4.46**	**2.93**	**2.89**	**2.20**	**5.93**
Z_{N} Bayes Gaussian	1.88	2.71	1.94	4.55	3.08	*3.44*	*2.90*	**5.93**
Z_{PL} profile likelihood	1.95	2.81	1.99	4.57	3.02	3.04	2.38	**5.93**
Z_{ZR} variance stabilization	1.93	2.66	1.98	4.22	3.00	3.07	2.39	5.86
Not recommended								
$Z_{\mathrm{BiN}} = s/\sqrt{n_{\mathrm{tot}}/\tau}$	*2.24*	*3.59*	2.17	*5.67*	3.11	**2.89**	**2.18**	6.16
$Z_{\mathrm{nn}} = s/\sqrt{n_{\mathrm{on}} + n_{\mathrm{off}}/\tau^2}$	1.46	*1.90*	1.66	*3.17*	2.82	3.28	*2.89*	*5.54*
$Z_{\mathrm{ssb}} = s/\sqrt{\hat{\mu}_{\mathrm{b}} + s}$	1.50	*1.92*	1.73	*3.20*	3.18	*4.52*	*7.07*	5.88
$Z_{\mathrm{bo}} = s/\sqrt{n_{\mathrm{off}}(1+\tau)/\tau^2}$	*2.74*	*3.99*	*2.42*	*6.47*	*3.50*	*3.90*	*3.02*	6.31
Ignore σ_{b}								
Z_{P} Poisson: ignore σ_{b}	2.08	2.84	2.14	4.87	*3.80*	*5.76*	*8.76*	*6.44*
$Z_{\mathrm{sb}} = s/\sqrt{\hat{\mu}_{\mathrm{b}}}$	*3.00*	*4.12*	*2.67*	*6.77*	*4.29*	*6.76*	*10.00*	*6.82*
Unsuccessful ad hockery								
Poisson: $\mu_{\mathrm{b}} \to \hat{\mu}_{\mathrm{b}} + \sigma_{\mathrm{b}}$	1.56	2.51	1.64	**4.47**	3.04	*4.24*	*5.51*	6.01
$s/\sqrt{\hat{\mu}_{\mathrm{b}} + \sigma_{\mathrm{b}}}$	*2.49*	*3.72*	*2.40*	*6.29*	*4.03*	*6.02*	*8.72*	*6.75*


On/off problem, Z_{Bi}=5
μ_b
τ
ΔZ
On/off problem, Z_{Bi}=5
μ_b
τ
ΔZ


On/off problem, Z_{PL} =5
μ_b
τ
ΔZ
On/off problem, Z_{PL} =5
μ_b
τ
ΔZ



On/off problem, Z_N=3
μ_b
τ
ΔZ
On/off problem, Z_N=3
μ_b
τ
ΔZ



Compare line shape near point, not spacing

What did we learn?

Shapes of tails matter at 3-5 sigma

Z_{Bi}: no undercoverage; can overcover for small N

Recommended

Z_{PL} quite reasonable behavior (despite Wilks);

pretty fast to calculate

Z_N undercovers, worse at high Z (Cranmer)

$S/\sqrt{B}$ and $S/\sqrt{(S+B)}$: just don’t: tails are wrong

Small τ is hard to get right

Summary

Should use Binomial Test for small N, large Zmin

Good Frequentist Properties

smallest N, overcovers a bit

numerically, more work than Z_{PL}

Binomial Test and L. Ratio have roots in Hyp Testing

For high and moderate N, Z_{PL} Likelihood Ratio Good

Not so much for low N or negative

Most wrong formulae overestimate significance

$S/\sqrt{B}$ is way too optimistic—ignores uncertainty in B

You MUST check properties

Z_N has coverage problems at large Zmin

References

Li & Ma   Astroph. Journ. 272 (1983) 314-324

Zhang & Ramsden   Experimental Astronomy 1 (1990) 145-163

Fraser   Journ. Am. Stat. Soc. 86 (1990) 258-265

Alexandreas et. al.   Nuc. Inst. & Meth. A328 (1993) 570-577

Gelman et. al., Bayesian Data Analysis, Chapman & Hall (1998)
(predictive p-value terminology)