Nuclear Instruments and Methods in Physics Research A 595 (2008) 480–501

Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process

Robert D. Cousins [a,*], James T. Linnemann [b], Jordan Tucker [a]

[a] Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA
[b] Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48840, USA

arXiv:physics/0702156v3 [physics.data-an]

http://arxiv.org/abs/physics/0702156

Some slides taken from Tucker's talk at PhyStat-LHC

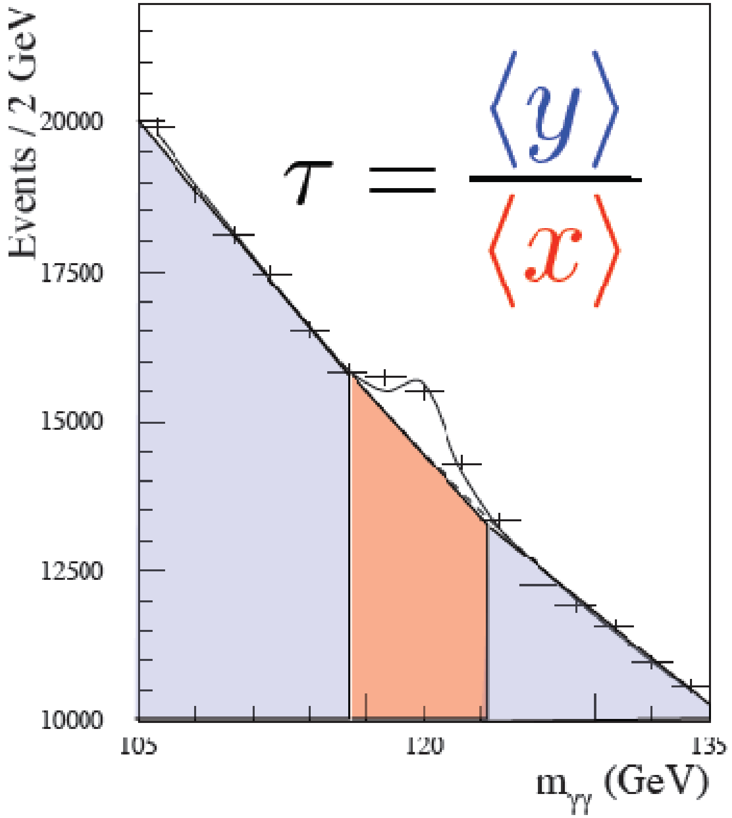
Two Problems in HEP, Gamma-Ray Astro, etc.

In both, *n_{on}* events observed from Poisson process with mean $\mu_s + \mu_b$: signal mean μ_s is of interest, background mean μ_b is estimated in subsidiary measurement.

1) On/off (sideband) problem:

GRA: n_{on} photons detected with telescope on-source;
n_{off} photons detected with telescope off source;
ratio of observing time (off/on) is τ (precisely known).
Is the data consistent with no source?

HEP: n_{on} events detected in signal region;
n_{off} events detected in sideband region;
ratio of expected events if there is only background (sideband/signal) is τ (precisely known).
Is the data consistent with no signal?



K. Cranmer, PhyStat-LHC

Two Problems in HEP, Gamma-Ray Astro, etc.

In both, n_{on} events observed from Poisson process with mean $\mu_s + \mu_b$: signal mean μ_s is of interest, background mean μ_b is estimated in subsidiary measurement.

2) Gaussian-mean background problem: subsidiary measurement of μ_b has normal (Gaussian) uncertainty with rms σ_b (precisely known, either absolutely or relatively).

Correspondence between the two problems

As detailed by Linnemann at PhyStat 2003, correspondence between on/off problems and Gaussian mean problems:

For on/off, estimate of mean background in signal region is

$$\hat{\mu}_{\mathrm{b}} = n_{\mathrm{off}}/\tau$$

(Rough) uncertainty on this estimate is $\sigma_{\mathrm{b}} = \sqrt{n_{\mathrm{off}}}/\tau$

Combining to eliminate n:

$$\boxed{\tau = \hat{\mu}_{\mathrm{b}}/\sigma_{\mathrm{b}}^2}$$

This correspondence, while rough, suggests that a recipe designed for one problem can be applied to the other problem, and performance studied: Given (n_{off}, τ), use above to get corresponding (μ_b, σ_b) and vice versa.

Key talks at past PhyStats

PhyStat 2003 at SLAC: List of methods and key points re binomial sol'n of on/off

Measures of Significance in HEP and Astrophysics

James T. Linnemann
Michigan State University, E. Lansing, MI 48840, USA and
Los Alamos National Laboratory, Los Alamos, NM 87545, USA

I compare and discuss critically several measures of statistical significance in common use in astrophysics and in high energy physics. I also exhibit some relationships among them.

PhyStat 2005 at Oxford: Critical look at integrating out uncertainty in background

STATISTICAL CHALLENGES FOR SEARCHES FOR NEW PHYSICS AT THE LHC

KYLE CRANMER
Brookhaven National Laboratory, Upton, NY 11973, USA
e-mail: Kyle.Cranmer@cern.ch

Because the emphasis of the LHC is on 5σ discoveries and the LHC environment induces high systematic errors, many of the common statistical procedures used in High Energy Physics are not adequate. I review the basic ingredients of LHC searches, the sources of systematics, and the performance of several methods. Finally, I indicate the methods that seem most promising for the LHC and areas that are in need of further study.

Z_{Bi}: Binomial Solution to on/off problem

- Discussed by Linnemann at PhyStat 2003 and a (very few) references therein, but hardly ever used.
- Null hypothesis H_0: signal mean μ_s is zero.
- Key point: Equivalent way to formulate H_0: the ratio of Poisson means in the sideband region and the signal region is τ (i.e., the ratio expected for pure background).
- Then: hypothesis test for ratio of Poisson means has standard frequentist solution, expressed in terms of binomial probabilities, available in ROOT.
- One ROOT line to get p-value, one line to convert to equivalent number of standard deviations (Z-value)

ROOT implementation for n_{on} = 140, n_{off} = 100, τ = 1.2:

```
double n_on = 140.
double n_off = 100.
double tau = 1.2
double P_Bi = TMath::BetaIncomplete(1./(1.+tau),n_on,n_off+1)
double Z_Bi = sqrt(2)*TMath::ErfInverse(1 - 2*P_Bi)
```

yielding $p_{\mathrm{Bi}} = 4.19 \times 10^{-5}$ and $Z_{\mathrm{Bi}} = 3.93$.

By construction, Z_{Bi} never under-covers. However, due to discreteness of *n*, it over-covers, especially at small *n*.

Amazingly, the same numerical answer is obtained in a Bayesian method using a Gamma-function pdf for the background mean, as advocated by Linnemann at (pre)PhyStat 2000, FNAL Confidence Limits workshop.

Recipes for Gaussian-mean background problem

N.B. one needs to specify whether the experimenter knows the absolute Gaussian uncertainty σ_b or the relative uncertainty σ_b / μ_b.

Then it is common in HEP to integrate out the nuisance parameter (unknown background mean) in an otherwise frequentist calculation (in some cases citing Cousins and Highland 1992, who integrated out an unknown luminosity). This Z-value is denoted by Z_N (N for normal).

In fact this was the recommendation out of the CMS Higgs group for the Physics TDR, adopted by CMS, as presented in a poster at PhyStat 2005 by Bityukov, "Program for evaluation of the significance confidence intervals and limits by direct probabilities calculations."

But! Frequentist coverage not guaranteed, and Cranmer gave examples at PhyStat 2005 where it was poor.

Integrating out nuisance parameters

Poisson probability for n_{on} or more background events:

$$p_P = \sum_{j=n_{\mathrm{on}}}^{\infty} \mathrm{e}^{-\mu_{\mathrm{b}}} \, \mu_{\mathrm{b}}^{j}/j!.$$

Weighted average over pdf for background mean:

$$\int p_P \, p(\mu_{\mathrm{b}}) \, d\mu_{\mathrm{b}}$$

If pdf for μ_b is Gaussian, leads to p_N, Z_N.

If pdf for μ_b is Gamma function (flat prior times likelihood function from Poisson sideband observation of n_{off}), leads to p_Γ, Z_Γ

Profile Likelihood

approximate confidence interval, one begins with the likelihood function; for the on/off problem, this is

$$\mathscr{L}_{\mathrm{P}} = \frac{(\mu_{\mathrm{s}} + \mu_{\mathrm{b}})^{n_{\mathrm{on}}}}{n_{\mathrm{on}}!} \mathrm{e}^{-(\mu_{\mathrm{s}}+\mu_{\mathrm{b}})} \frac{(\tau\mu_{\mathrm{b}})^{n_{\mathrm{off}}}}{n_{\mathrm{off}}!} \mathrm{e}^{-\tau\mu_{\mathrm{b}}} \tag{20}$$

while for the Gaussian-mean background problem with either absolute or relative σ_{b}, it is

$$\mathscr{L}_{\mathrm{G}} = \frac{(\mu_{\mathrm{s}} + \mu_{\mathrm{b}})^{n_{\mathrm{on}}}}{n_{\mathrm{on}}!} \mathrm{e}^{-(\mu_{\mathrm{s}}+\mu_{\mathrm{b}})} \frac{1}{\sqrt{2\pi\sigma_{\mathrm{b}}^2}} \exp\left(-\frac{(\hat{\mu}_{\mathrm{b}} - \mu_{\mathrm{b}})^2}{2\sigma_{\mathrm{b}}^2}\right) \tag{21}$$

where as discussed below we have explored the effect of truncating the Gaussian pdf in $\hat{\mu}_{\mathrm{b}}$ and renormalizing prior to forming $\mathscr{L}_{\mathrm{G}}$.

Using either $\mathscr{L}_{\mathrm{P}}$ or $\mathscr{L}_{\mathrm{G}}$, one obtains the log-likelihood ratio

$$\Lambda(\mu_{\mathrm{s}}) = \frac{\mathscr{L}(\mu_{\mathrm{s}}, \tilde{\tilde{\mu}}_{\mathrm{b}}(\mu_{\mathrm{s}}))}{\mathscr{L}(\tilde{\mu}_{\mathrm{s}}, \tilde{\mu}_{\mathrm{b}})} \tag{22}$$

Table 1
Test cases and significance results

Reference	[40]	[41]	[42]	[43]	[44]	[44]	[45]	[46]	[47]	[48]
n_{on}	**4**	**6**	**9**	**17**	**50**	**67**	**200**	**523**	**498 426**	**2 119 449**
n_{off}	**5**	18.78	17.83	40.11	**55**	**15**	**10**	**2327**	**493 434**	23 650 096
τ	**5.0**	14.44	4.69	10.56	**2.0**	**0.5**	**0.1**	**5.99**	**1.0**	**11.21**
$\hat{\mu}_{\mathrm{b}}$	1.0	**1.3**	**3.8**	**3.8**	27.5	30.0	100.0	388.6	493 434	**2 109 732**
$s = n_{\mathrm{on}} - \hat{\mu}_{\mathrm{b}}$	3.0	4.7	5.2	13.2	22.5	37	100	**134**	**4992**	**9717**
σ_{b}	0.447	**0.3**	**0.9**	**0.6**	3.71	7.75	31.6	8.1	702.4	433.8
$f = \sigma_{\mathrm{b}}/\hat{\mu}_{\mathrm{b}}$	0.447	0.231	0.237	0.158	0.135	0.258	0.316	0.0207	0.00142	0.000206
Reported p		**0.003**	**0.027**	**2E−06**						
Reported Z		2.7	1.9	**4.6**				**5.9**	**5.0**	**6.4**
See conclusion										
$Z_{\mathrm{Bi}} = Z_{\Gamma}$ binomial	**1.66**	**2.63**	**1.82**	**4.46**	**2.93**	**2.89**	**2.20**	**5.93**	**5.01**	**6.40**
Z_{N} Bayes Gaussian	1.88	2.71	1.94	4.55	3.08	*3.44*	*2.90*	**5.93**	**5.02**	**6.40**
Z_{PL} profile likelihood	1.95	2.81	1.99	4.57	3.02	3.04	2.38	**5.93**	**5.01**	**6.41**
Z_{ZR} variance stabilization	1.93	2.66	1.98	4.22	3.00	3.07	2.39	5.86	**5.01**	**6.40**
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	**5.01**	**6.41**
$Z_{\mathrm{nn}} = s/\sqrt{n_{\mathrm{on}} + n_{\mathrm{off}}/\tau^2}$	1.46	*1.90*	1.66	*3.17*	2.82	3.28	*2.89*	*5.54*	**5.01**	**6.40**
$Z_{\mathrm{ssb}} = s/\sqrt{\hat{\mu}_{\mathrm{b}} + s}$	1.50	*1.92*	1.73	*3.20*	3.18	*4.52*	*7.07*	5.88	*7.07*	6.67
$Z_{\mathrm{bo}} = s/\sqrt{n_{\mathrm{off}}(1+\tau)/\tau^2}$	*2.74*	*3.99*	*2.42*	*6.47*	*3.50*	*3.90*	*3.02*	6.31	**5.03**	**6.41**
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*	*7.09*	6.69
$Z_{\mathrm{sb}} = s/\sqrt{\hat{\mu}_{\mathrm{b}}}$	*3.00*	*4.12*	*2.67*	*6.77*	*4.29*	*6.76*	*10.00*	*6.82*	*7.11*	6.69
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	*6.09*	**6.39**
$s/\sqrt{\hat{\mu}_{\mathrm{b}} + \sigma_{\mathrm{b}}}$	*2.49*	*3.72*	*2.40*	*6.29*	*4.03*	*6.02*	*8.72*	*6.75*	*7.10*	6.69

In the top section, the primary input numbers from the papers are in boldface, with derived numbers (using Eqs. (5)–(8)) in normal font. The test cases are ordered in data counts; Refs. [44,45,47] have small values of τ, troublesome for some methods. Below the top section, Z-values in boldface are nearly equal to the reference Z_{Bi}, while Z-values in italics differ by more than 0.5.

Scan of coverage for the three main recipes applied to the two problems, when no signal.

For each recipe and each problem:

- Fix true background mean μ_b and the other experimental setup parameter (off/on ratio τ, or relative $f = \sigma_b / \mu_b$).
- Choose a "claimed" Z-value such as 1.28, 3, or 5.
- Calculate frequency, in absence of signal, that claimed Z is exceeded for an ensemble of experiments with chosen true background mean. Convert to "true" Z-value.

For representative claimed Z's, our paper contains 2D plots of true Z-value, as function of μ_b and the other experimental setup parameter.

Z_{Bi} applied to the on/off problem with claimed Z=1.28

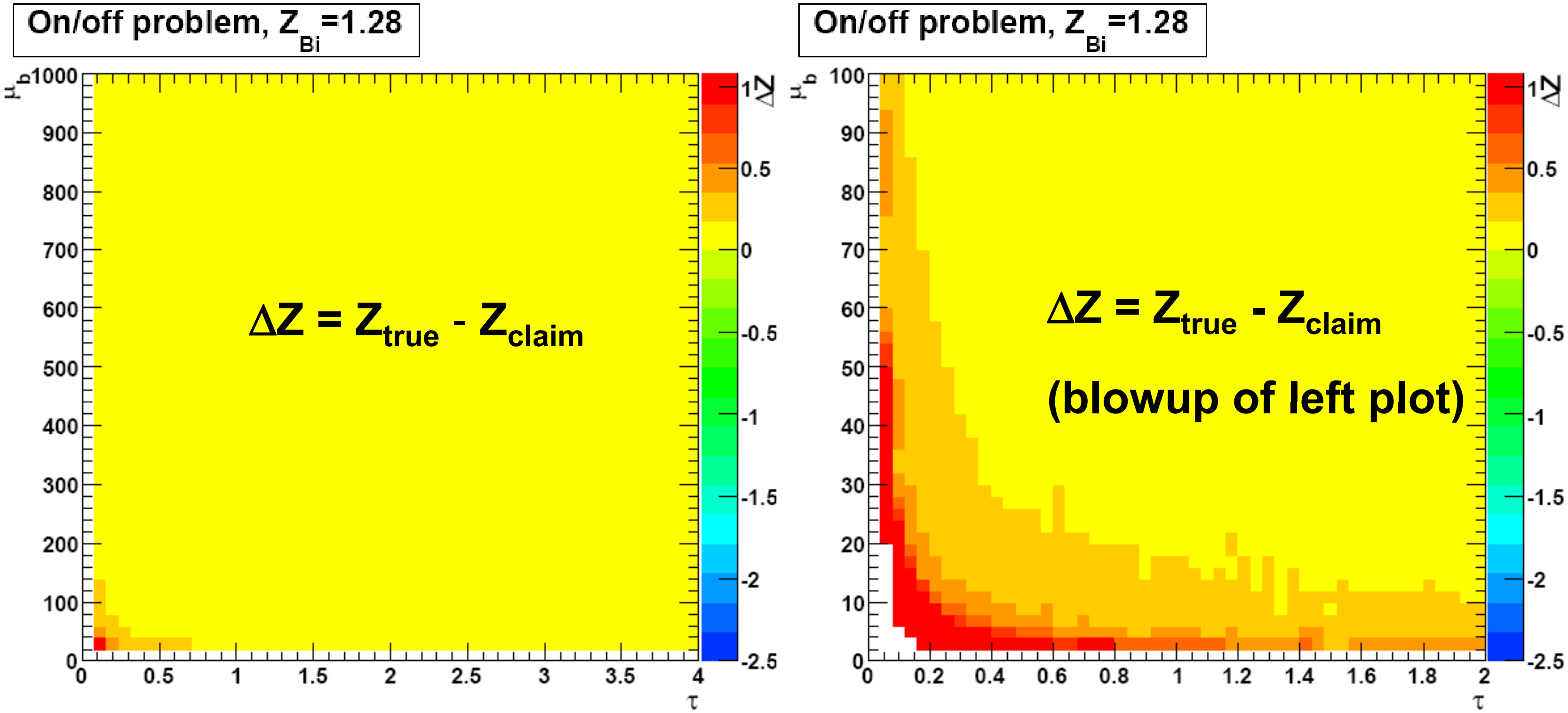


Fig. 1. For the on/off problem analyzed using the Z_{Bi} recipe, for each fixed value of τ and μ_{b}, the plot indicates the calculated $Z_{\text{true}} - Z_{\text{claim}}$ for the ensemble of experiments quoting a $Z_{\text{claim}} \geq 1.28$, i.e., a p-value of 0.1 or smaller.

$\Delta Z \geq 0$ always, noticeably so for small n.

Z_N applied to the on/off problem with claimed Z=1.28

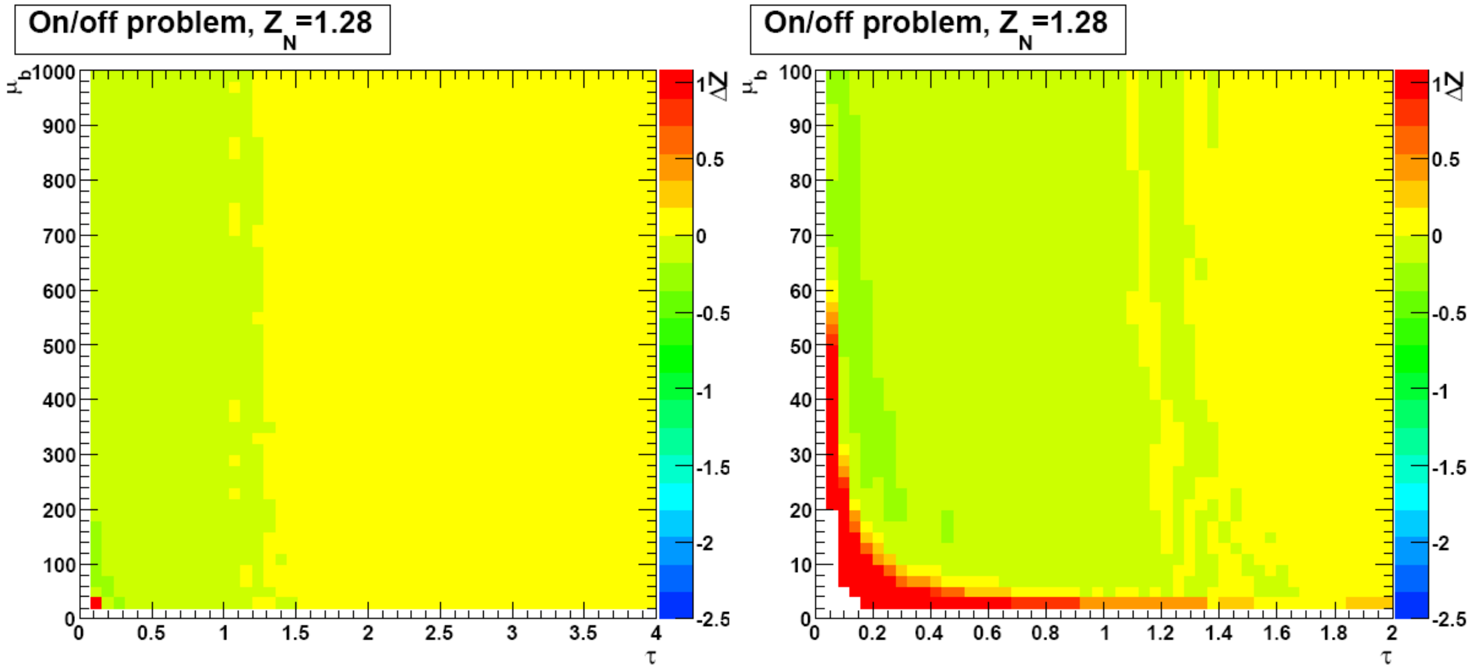


Fig. 4. For the on/off problem analyzed using the Z_N recipe, for each fixed value of τ and μ_b, the plot indicates the calculated $Z_{true} - Z_{claim}$ for the ensemble of experiments quoting $Z_{claim} \geq 1.28$, i.e., a p-value of 0.1 or smaller.

Z_{PL} applied to the on/off problem with claimed Z=1.28

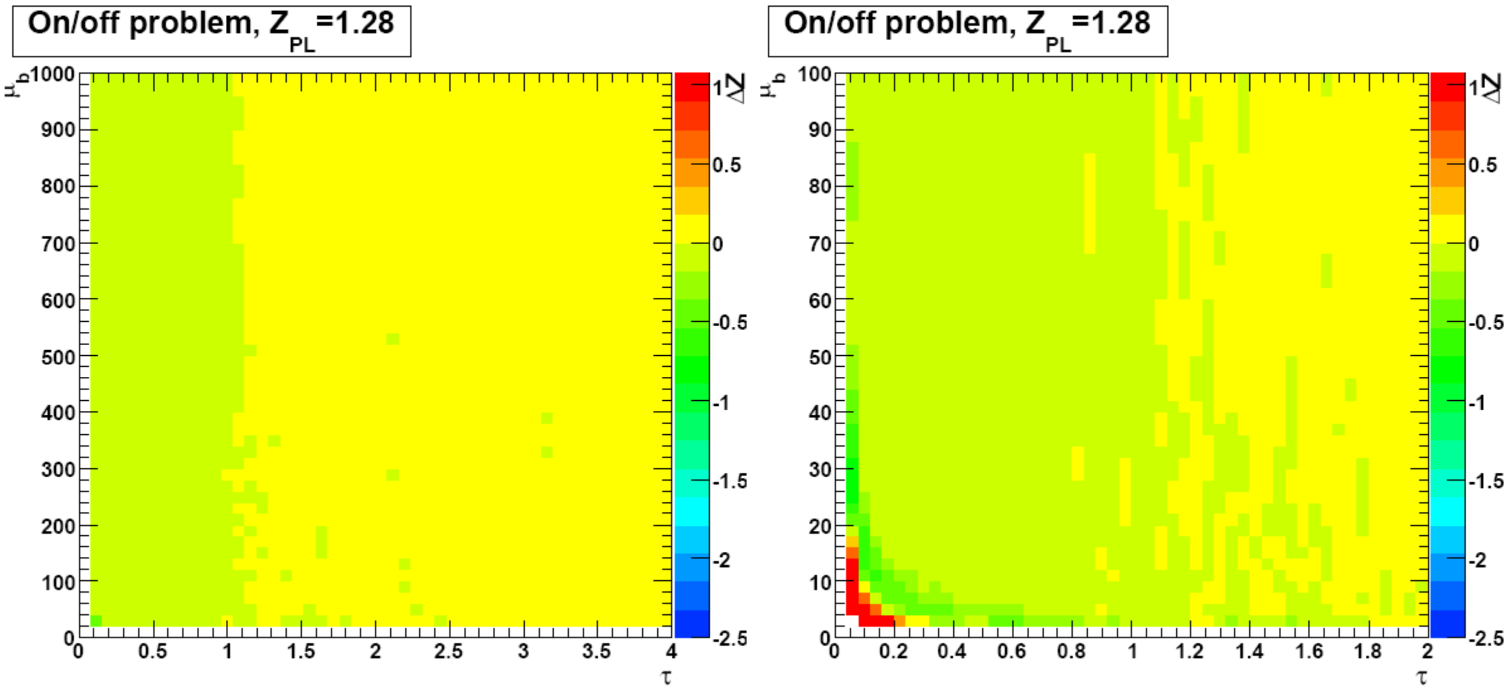


Fig. 7. For the on/off problem analyzed using the profile likelihood method, for each fixed value of τ and μ_b, the plot indicates the calculated $Z_{\text{true}} - Z_{\text{claim}}$ for the ensemble of experiments quoting $Z_{\text{claim}} \geq 1.28$, i.e., a p-value of 0.1 or smaller.

Z_{Bi} applied to the on/off problem with claimed Z=5

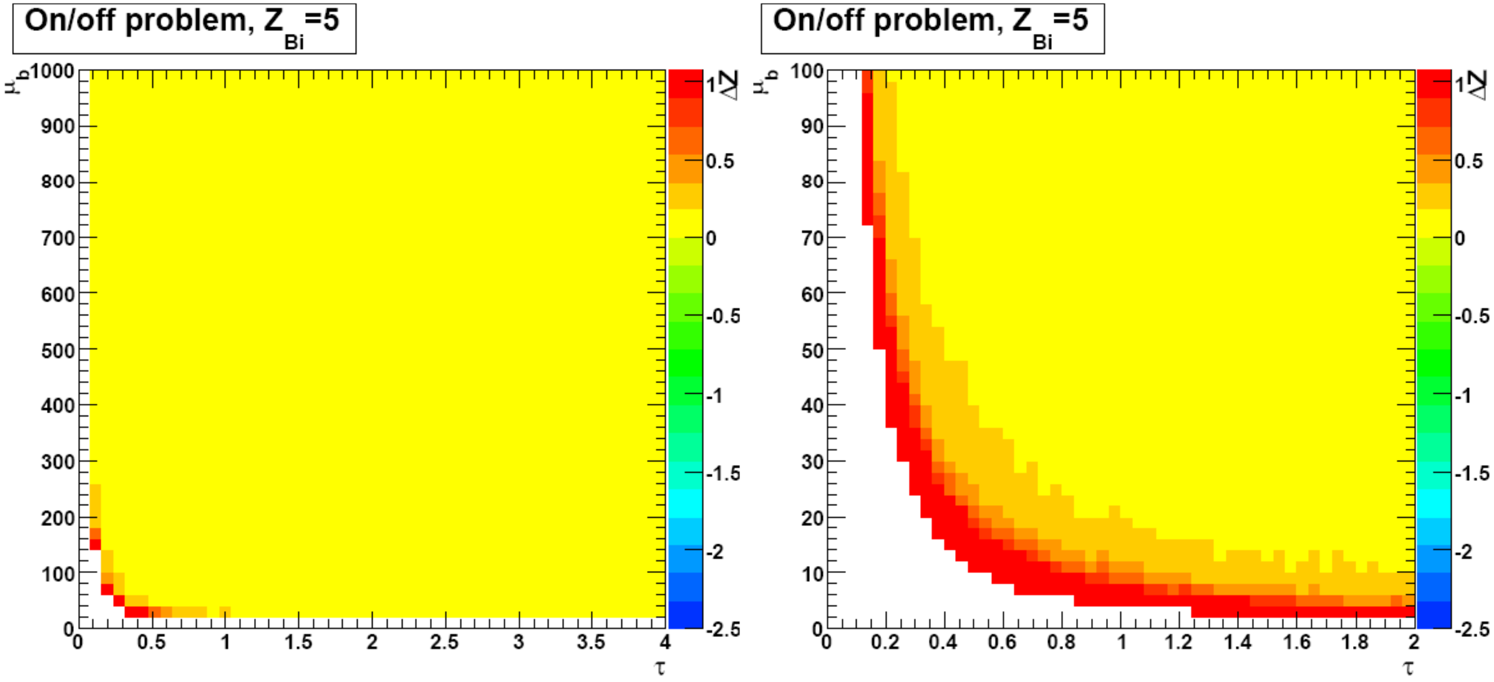


Fig. 3. For the on/off problem analyzed using the Z_{Bi} recipe, for each fixed value of τ and μ_{b}, the plot indicates the calculated $Z_{\mathrm{true}} - Z_{\mathrm{claim}}$ for the ensemble of experiments quoting $Z_{\mathrm{claim}} \geq 5$, i.e., a p-value of 2.87×10^{-7} or smaller.

Z_N applied to the on/off problem with claimed Z=5

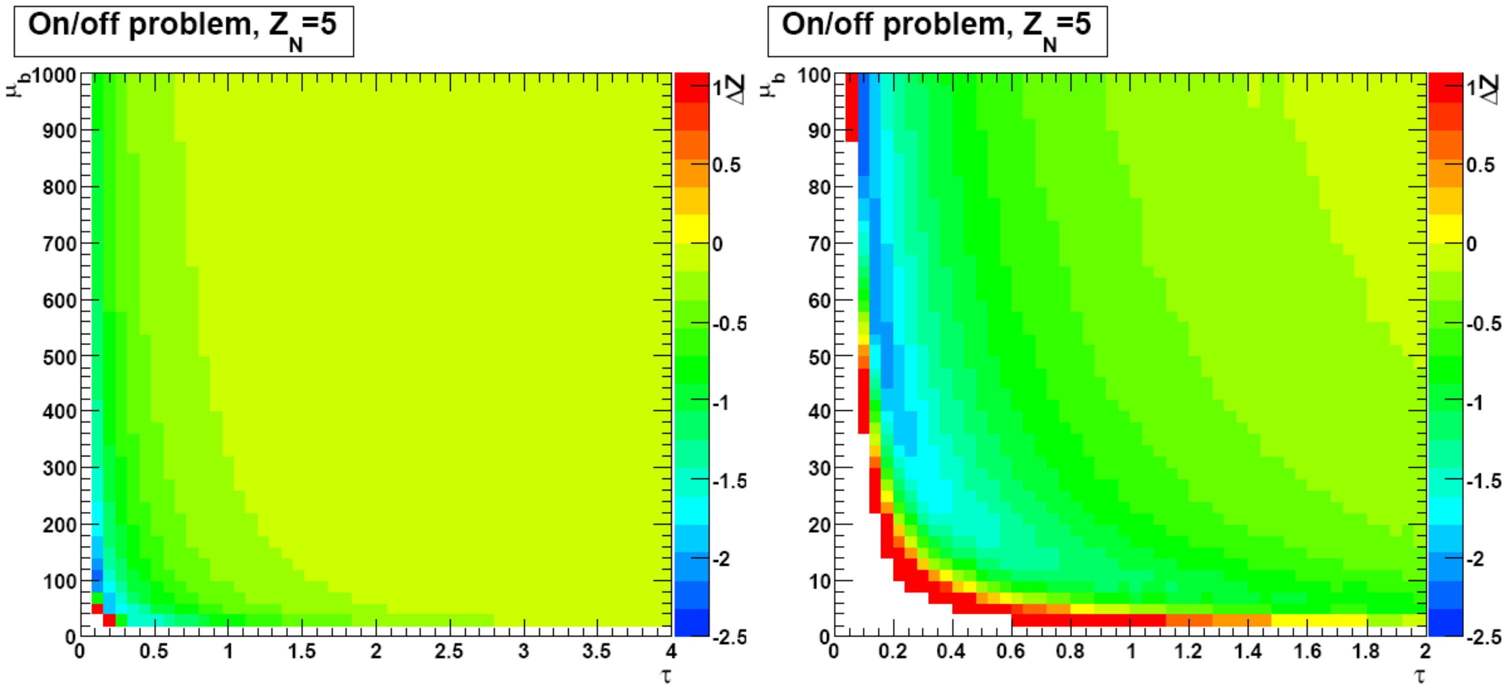


Fig. 6. For the on/off problem analyzed using the Z_{N} recipe, for each fixed value of τ and μ_{b}, the plot indicates the calculated $Z_{\mathrm{true}} - Z_{\mathrm{claim}}$ for the ensemble of experiments quoting $Z_{\mathrm{claim}} \geq 5$, i.e., a p-value of 2.87×10^{-7} or smaller.

Z_{PL} applied to the on/off problem with claimed Z=5

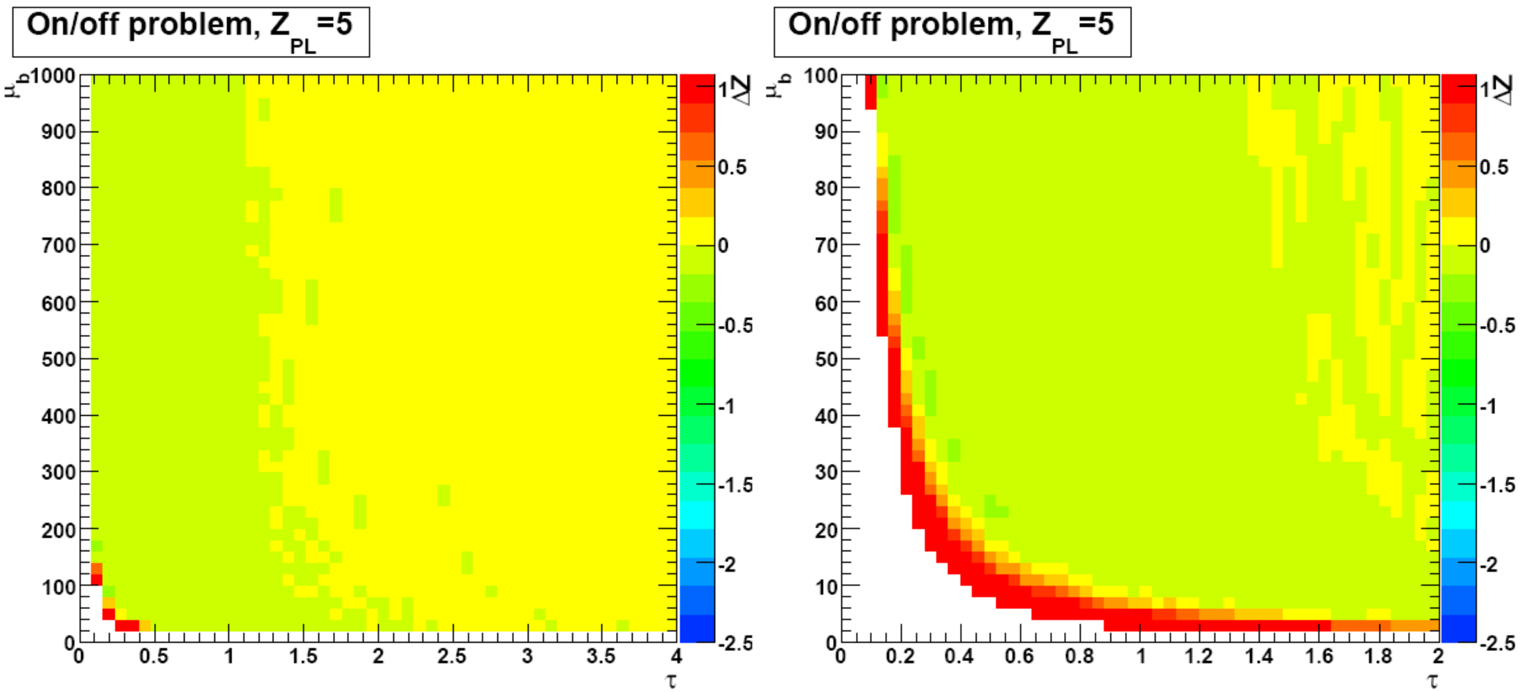


Fig. 9. For the on/off problem analyzed using the profile likelihood method, for each fixed value of τ and μ_b, the plot indicates the calculated $Z_{\rm true} - Z_{\rm claim}$ for the ensemble of experiments quoting $Z_{\rm claim} \geq 5$, i.e., a p-value of 2.87×10^{-7} or smaller.

Z_{Bi} applied to the Gaussian-mean background problem (relative σ_b) with claimed Z=5

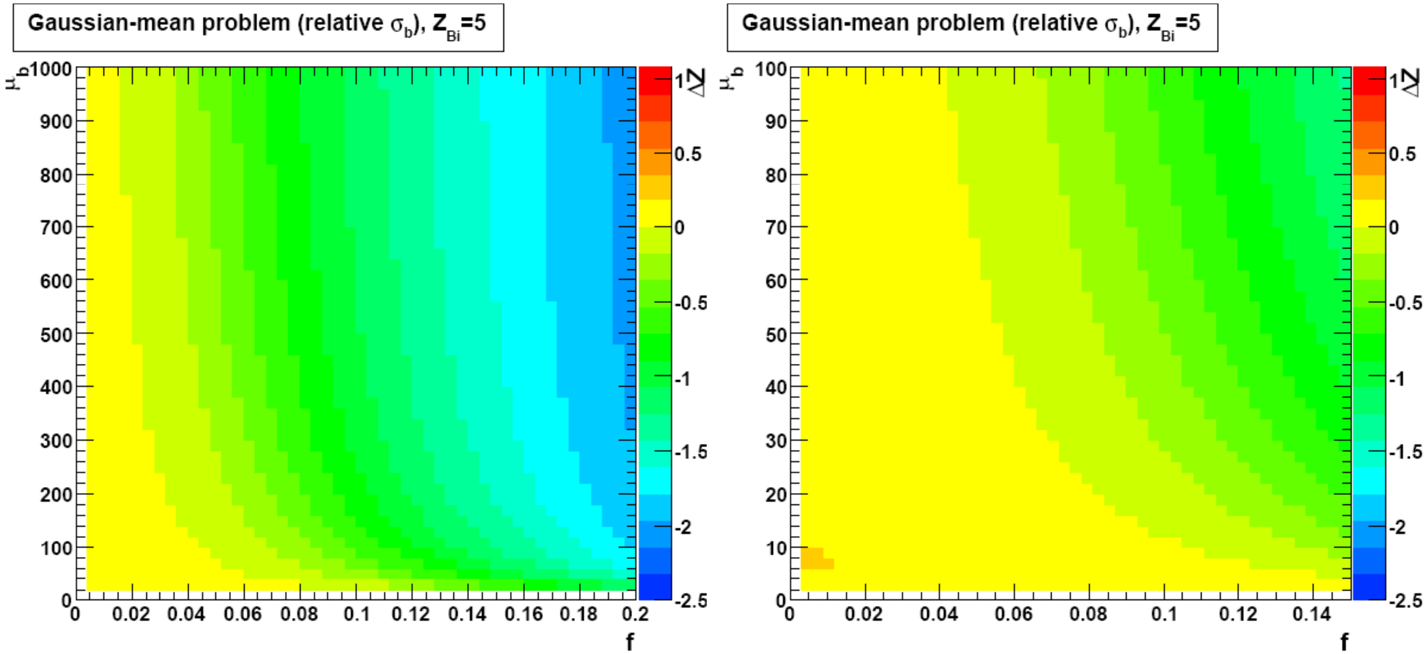
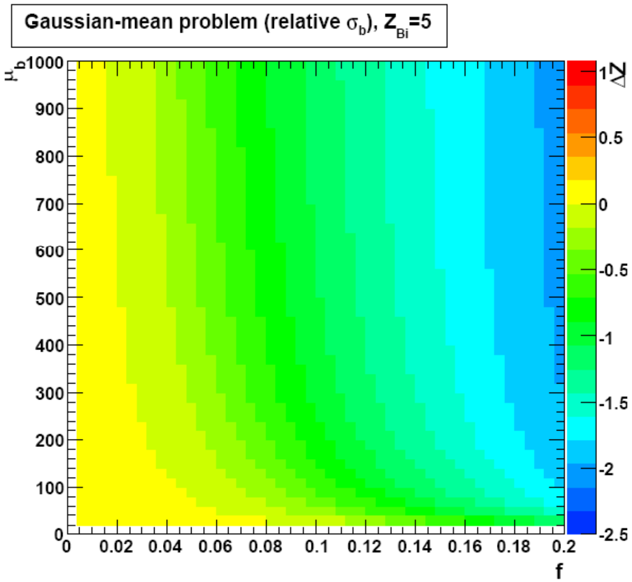
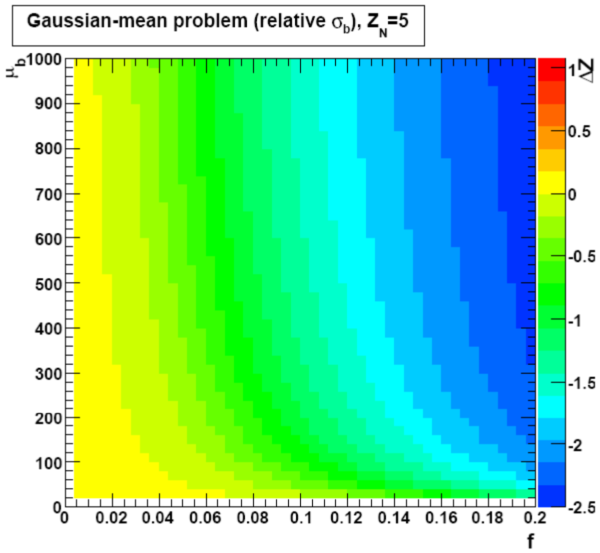


Fig. 21. For the Gaussian-mean background problem with exactly known relative uncertainty f, analyzed using the Z_{Bi} recipe, for each fixed value of f and μ_{b}, the plot indicates the calculated $Z_{\mathrm{true}} - Z_{\mathrm{claim}}$ for the ensemble of experiments quoting $Z_{\mathrm{claim}} \geq 5$, i.e., a p-value of 2.87×10^{-7} or smaller.

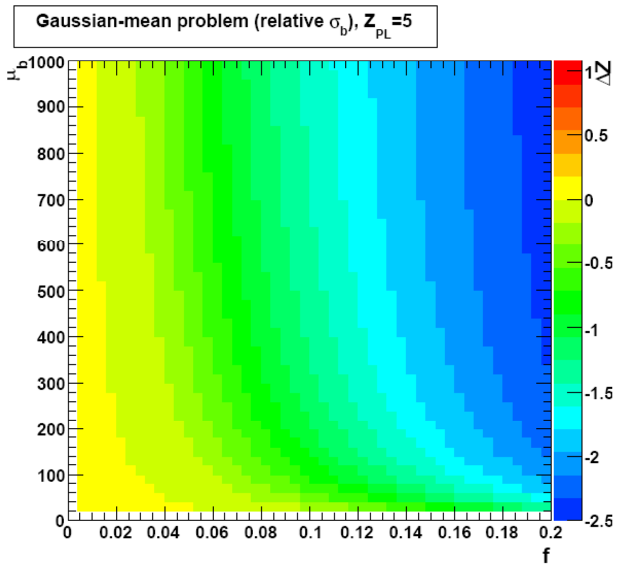
$$f = \sigma_b / \mu_b$$



Gaussian-mean problem (relative σ_b), Z_{Bi}=5
μ_b
f
ΔZ



Gaussian-mean problem (relative σ_b), Z_N=5
μ_b
f
ΔZ



Gaussian-mean problem (relative σ_b), Z_{PL}=5
μ_b
f
ΔZ

Summary and Conclusions

Sorry, ran out of time. See paper!

Acknowledgments

We thank Kyle Cranmer and Luc Demortier for numerous enlightening discussions, pointers to references, and for insightful comments on earlier versions of this work. J.L. wishes to thank LANL for hospitality and financial support during his sabbatical; Tom Loredo for Ref. [6]; and James Berger for hospitality at the SAMSI 2006 Institute, and acknowledges useful conversations there with professors John Hartigan and Joel Heinrich, which helped him toward the proof that $Z_{\mathrm{Bi}} = Z_{\Gamma}$. This work was partially supported by the U.S. Department of Energy and the National Science Foundation.