HEP

June 2[illegible]006

Abstract

We describe a statistical hypothesis test of a possible signal in the presence of nuisance parameters that are motivated by problems arising in high energy physics. Our approach is based on higher order likelihood inference to tail areas in the presence of nuisance parameters, using the third order approximations. Practical examples from particle physics are outlined, and simulation studies show that the third order likelihood based approximations provide nice solutions and are a most promising method for the examples.

Introduction

In high energy physics (HEP) experiments looking for some rare or perhaps non-existent process, the primary measurement involves counting particle physics interactions of interest, called "events". Parameters that are not of physical interest or not observable to the experimenter appear unavoidably due to the necessary calibration of the measuring apparatus. The uncertainty in the mean rate of background events contributes an uninteresting nuisance parameter. We consider the hypothesis test of a background-only null hypothesis in the presence of uncertainty in the background. The method used here is provided by the observed likelihood function along with the associated p-value function obtained to a high order approximation.

In Section 2 we give out some background notation, introduce the notion of profile likelihood, and higher order approximations. In Section 3 we illustrate the third order p-value function approach in models with nuisance parameters. Section 4 shows....

2 Third order p-values from Profile Likelihood

2.1 Notation

Suppose the observations y come from a model with probability density or mass function $f(y;\theta)$, where $\theta = (\psi, \lambda)$ is q-dimensional. Here, ψ is the parameter of interest, and λ is the nuisance parameter with ~~$(q-1)$~~ dimensions. The likelihood function is proportional to $f(y;\theta)$ only up to an arbitrary multiple depending on y but not θ. That is,

$$L(\theta) = L(\theta; y) = c(y) f(y;\theta) \tag{1}$$

Consider a random sample $\underset{\sim}{y} = (y_1, \ldots, y_n)$ of independent observations with each y_i following the model $f(y;\theta)$. The likelihood function is then proportional to $\prod f(y_i;\theta)$. Therefore, the log-likelihood function becomes a sum of independently and identically distributed components

$$l(\theta) = l(\theta; y) = \sum l(\theta, y_i) \tag{2}$$

The maximum likelihood estimator $\hat{\theta}$ is the solution to the score equation

$$l'(\hat{\theta}) = 0 \tag{3}$$

~~For further information,~~ the observed Fisher information function $j(\theta)$ is the curvature of the log-likelihood

$$j(\theta) = -l''(\theta) \tag{4}$$

and the expected Fisher information is the model quantity

$$i(\theta) = E\{-l''(\theta)\} \tag{5}$$

The normal approximation to the square root of the likelihood ratio is often considered more accurate in the tails, and is called first order approximations because the error in the approximation is $O(n^{-1/2})$. That is,

$$r(\theta) = \text{sign}(\hat{\theta} - \theta)[2\{l(\hat{\theta}) - l(\theta)\}]^{1/2} \quad \dot{\sim} \quad N(0,1) \tag{6}$$

2.2 Profile Likelihood

The basic idea of profile likelihood is rather straightforward. Instead of the inference based on full likelihood functions with both parameter of interest and nuisance parameters involved, eliminating nuisance parameters might be desirable. The profile likelihood function is thus defined as

$$L_p(\psi) = L(\psi, \hat{\lambda}_\psi) \tag{7}$$

where $\hat{\lambda}_\psi$ is the restricted maximum likelihood estimate obtained by maximizing the likelihood function over the nuisance parameter λ with ψ fixed. The first order approximation of the previous section generalizes to

$$r(\psi) = \text{sign}(\hat{\psi} - \psi)[2\{l_p(\hat{\psi}) - l_p(\psi)\}]^{1/2} \quad \dot{\sim} \quad N(0,1) \tag{8}$$

2.3 3rd Order Approximations from Profile Likelihood

The first order approximations can be improved to the third order, i.e. with relative error $O(n^{-3/2})$, using the so-called r^* approximation

$$r^*(\psi) = r(\psi) + 1/\{r(\psi)\} \log\{Q(\psi)/r(\psi)\} \tag{9}$$

where Q is a likelihood-based statistic~~s~~ and a generalization of the Wald statistics $(\hat{\theta} - \theta)j^{1/2}(\hat{\theta})$, ...

The approach to third-order likelihood theory involves two distinct steps in dimensional simplification of the variable being examined. The first step can be described as to find an appropriate reparameterization and obtain a canonical interest parameter. The second step is a subsequent reduction by marginalisation from the canonical interest parameter to the dimension 1 of a separate component of interest.

Again assume q-dimensional parameter $\theta = (\psi, \lambda)$ and λ is $(q-1)$-dimensional. ψ is the scalar parameter of interest, while λ is the nuisance parameter. As we said above for the inference concerning ψ we need to find an appropriate reparameterization and obtain a canonical interest parameter, say φ, and then a likelihood function appropriate to the component parameter. It has been shown in [3] that in a full exponential family model, the canonical parameter can be obtained by differentiating the log-likelihood function with respect to the minimal sufficient statistic, say t. And the approximation is completed through the profile likelihood function. Please refer to [1], [2], [3], and [4] for details and the derivation of the following r^* approximation formulas.

The result of the third order approximation r^* is given above by (9)

$$r^*(\psi) = r(\psi) + 1/\{r(\psi)\} \log\{Q(\psi)/r(\psi)\}$$

where $r(\psi)$ is the likelihood root based on profile likelihood function of first order approximation

$$r(\psi) = \text{sign}(\hat{\psi} - \psi)[2\{l_p(\hat{\psi}) - l_p(\psi)\}]^{1/2}$$

and

$$\begin{aligned} Q(\psi) &= \frac{|l_{;t}(\hat{\theta}) - l_{;t}(\hat{\theta}_\psi) \quad l_{\lambda;t}(\hat{\theta}_\psi))|}{|l_{\theta;t}(\hat{\theta})|} \cdot \frac{|j_{\theta\theta}(\hat{\theta})|^{1/2}}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|^{1/2}} && (10) \\ &= \{v(\hat{\theta}) - v(\hat{\theta}_\psi)\}\hat{\sigma}_v^{-1} && (11) \end{aligned}$$

where

$$\begin{aligned} v(\theta) &= e_\psi^T \varphi(\theta), && (12) \\ e_\psi &= \psi_{\varphi'}(\hat{\theta}_\psi)/|\psi_{\varphi'}(\hat{\theta}_\psi)|, && (13) \\ \hat{\sigma}_v^2 &= |j_{(\lambda\lambda)}(\hat{\theta}_\psi)|/|j_{(\theta\theta)}(\hat{\theta})|, && (14) \\ |j_{(\theta\theta)}(\hat{\theta})| &= |j_{\theta\theta}(\hat{\theta})||\varphi_{\theta'}(\hat{\theta})|^{-2}, && (15) \\ |j_{(\lambda\lambda)}(\hat{\theta}_\psi)| &= |j_{\lambda\lambda}(\hat{\theta}_\psi)||\varphi_{\lambda'}(\hat{\theta}_\psi)|^{-2} && (16) \end{aligned}$$

Starting from the log-likelihood function $l(\theta; y)$ with canonical parameter φ and the observed Fisher information $j(\theta)$, the key step in the calculation, values of Q, is not very difficult to implement algorithmically as we will show with examples in the next section.

3 The Signal Test

The case without nuisance parameters has been discussed fully in [1], including the recommended exact mid-p-value function, and p-value functions from both the first order approximation r and third order approximation r^*. We illustrate the method of the third order approximations from the profile likelihood using the examples of looking for rare particles with uncertainty in the background.

3.1 Poisson signal with Poisson noise: PP-model

Assume the signal measurement is a Poisson count x with mean $\mu + \lambda$, and the background estimate is obtained from an independent Poisson count y with mean $\tau\lambda$. The estimated precision of the background gives us a value of τ. Then Poisson signal with Poisson noise Model taken from [1] is given by

$$X \sim Pois(\mu + \lambda),\ Y \sim Pois(\tau\lambda),\ \tau \text{ known}(> 0)$$

The background rate is assumed unknown and thus considered as a nuisance parameter. The detailed procedure to calculate the r^* approximation is outlined in [1].

Figure 1 compares the p-value function using the mid-p-value assuming the background rate is known with the p-value from $\Phi(r^*)$ with background error adjustment. Both figures assume $x = 25$, $y = 6.7(??)$, and a background error of ± 1.75. The p-value for testing $\mu = 0$ is 0.00468, allowing for the uncertainty in the background, whereas 0.00038 ignoring this uncertainty.

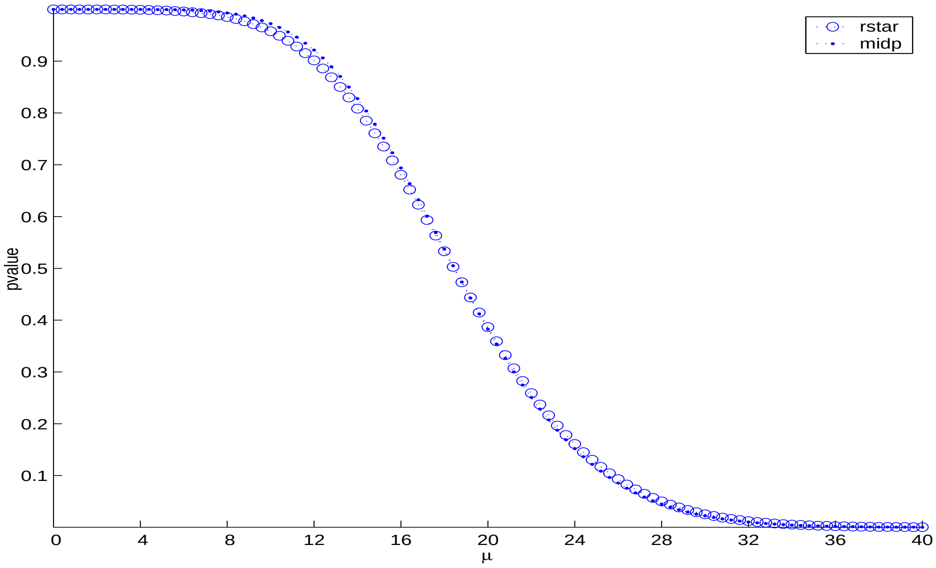

Figure 1: Comparison of the p-value functions computed assuming the background is known and using the mid-p-value with the third order approximation r^*.

~~Compared with the inference for the difference between two Poisson means, statistical inference about the ratio of two Poisson means is satisfactorily solved using the binomial distribution when the background imprecision is ignored. If a nuisance parameter is also included, the signal test can be implemented by modeling the mean of x as $\alpha\lambda$, say, and testing the value $\alpha = 1$. Numerical result shows that the p-value for testing $\alpha = 1$ is exactly the same as the p-value for testing $\mu = 0$, which is consistent with what we expect.~~

3.2 Poisson signal with Normal noise: PN-model

As a second example, assume that the background is modeled as an independent Normal distribution rather than Poisson. Then a reasonable model can be described as

$$X \sim Pois(\mu + \lambda),\, Y \sim N(\tau\lambda, \tau\lambda),\, \tau \text{ known}(> 0).$$

Following ~~the~~ similar steps as ~~done~~ with the PP-Model yields the third order r^* approximation. ~~First of all, t~~he log likelihood function is given by

$$l(\mu, \lambda) = x \log(\mu + \lambda) - \frac{1}{2\tau\lambda} y^2 - (\mu + \lambda) - \frac{1}{2} \log \lambda - \frac{\tau}{2} \lambda \tag{17}$$

~~Let $\theta = (\mu, \lambda)^T$.~~ The canonical parameter is ~~notated by~~ $\varphi = (\mu + \lambda), -\frac{1}{2\tau\lambda})^T$. The m.l.e. of θ is given by

$$\begin{cases} \hat{\mu} = x - \hat{\lambda} \\ \hat{\lambda} = \frac{-1 + \sqrt{1 + 4y^2}}{2\tau} \end{cases}.$$

~~It is also~~ possible to find the restricted m.l.e. $\hat{\lambda}_\mu$ analytically~~. Instead,~~ the calculation ~~is completed~~ by Newton-Raphson ~~method~~.

er necessary ingredients are

$$e_\psi^T = (\mu + \hat{\lambda}_\mu, -2\tau\hat{\lambda}_\mu^2)/\sqrt{(\mu+\hat{\lambda}_\mu)^2 + 4\tau^2\hat{\lambda}_\mu^4} \tag{18}$$

$$v(\hat{\theta}) = ((\mu+\hat{\lambda}_\mu)\log(\hat{\mu}+\hat{\lambda}) + \hat{\lambda}_\mu^2/\hat{\lambda})/\sqrt{(\mu+\hat{\lambda}_\mu)^2 + 4\tau\hat{\lambda}_\mu^4} \tag{19}$$

$$v(\hat{\theta}_\psi) = ((\mu+\hat{\lambda}_\mu)\log(\mu+\hat{\lambda}_\mu) + \hat{\lambda}_\mu)/\sqrt{(\mu+\hat{\lambda}_\mu)^2 + 4\tau\hat{\lambda}_\mu^4} \tag{20}$$

and finally

$$Q = (v(\hat{\theta}) - v(\hat{\theta}_\psi))/\hat{\sigma}_v \tag{21}$$

with

$$|j_{(\theta\theta)}(\hat{\theta})| = 4\tau^2(\frac{y^2}{\tau}\hat{\lambda} - \frac{\hat{\lambda}^2}{2}) \tag{22}$$

$$|j_{(\lambda\lambda)}(\hat{\theta}_\psi)| = [(\mu+\hat{\lambda}_\mu)(2\tau\hat{\lambda}_\mu^2)]^2 \cdot |\frac{x}{(\mu+\hat{\lambda}_\mu)^2} + \frac{y^2}{\tau\hat{\lambda}_\mu^3} - \frac{1}{2\hat{\lambda}_\mu^2}|/[(\mu+\hat{\lambda}_\mu)^2 + (2\tau\hat{\lambda})^2] \tag{23}$$

The likelihood root is

$$r = \mathrm{sign}(Q)\sqrt{2[l(\hat{\mu},\hat{\lambda}) - l(\mu,\hat{\lambda}_\mu)]} \tag{24}$$

e third order approximation to the p-value function is $\Phi(r^*)$, where

$$r^* = r + (1/r)\log(Q/r) \tag{25}$$

Figure 2 shows the p-value function using the mid-p-value assuming the background rate is known with the p-value from $\Phi(r^*)$ with background error adjustment, provided that $x = 25$, $y = 6.7$, and a background error of ± 1.75. The p-value for testing $\mu = 0$ is 0.00561, allowing for the uncertainty in the background, whereas 0.0004081, same as PP-Model, ignoring this uncertainty.

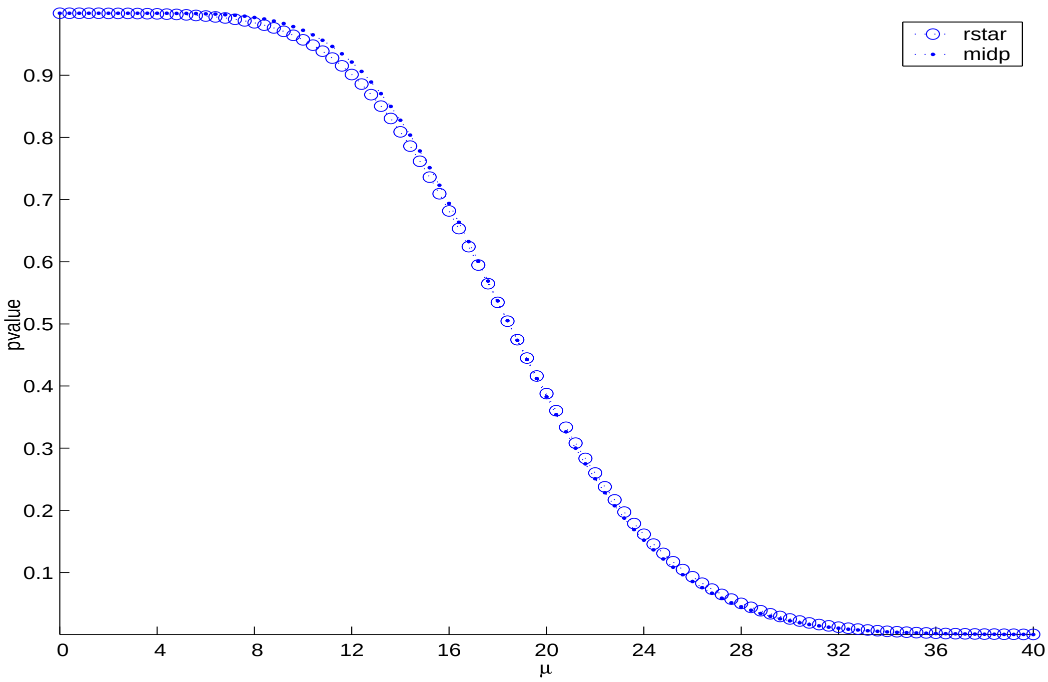

Figure 2: Comparison of the p-value functions computed assuming the background is known and using the lower and mid-p-value with the third order approximation r^*.

Table I p-values for testing $\mu = 0$ given $x = 17, y = 6.7$ and background error ± 1.75

mid p-value	0.0004081
$\Phi(r^*)$ PP-Mol	0.00468
$\Phi(r^*)$ PN-Mol	0.00561

If we have the two p-value functions from PP-Model and PN-Model plotted together on the same graph with $x = 25$, $y = 6.7$, and a background error of ± 1.75, we will see that the two p-value functions are nearly identical, which implies the background measurement is flexible. In addition, they are both close to the mid-p-value function with known background rate.

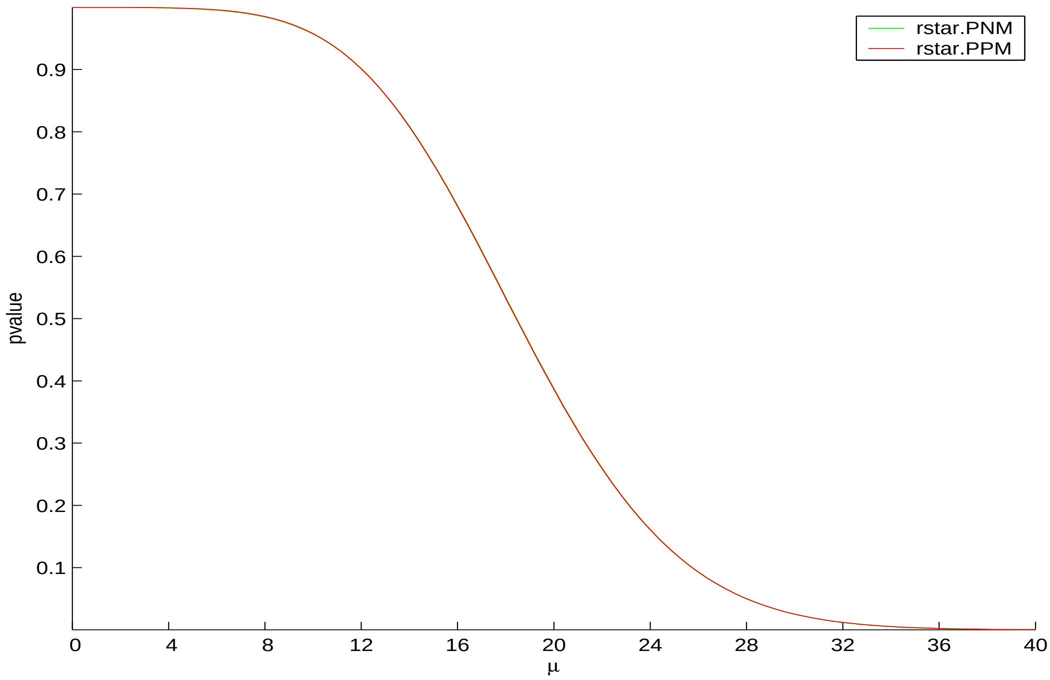

Figure 3: Comparison of the p-value functions computed by PP-model and PN-model.

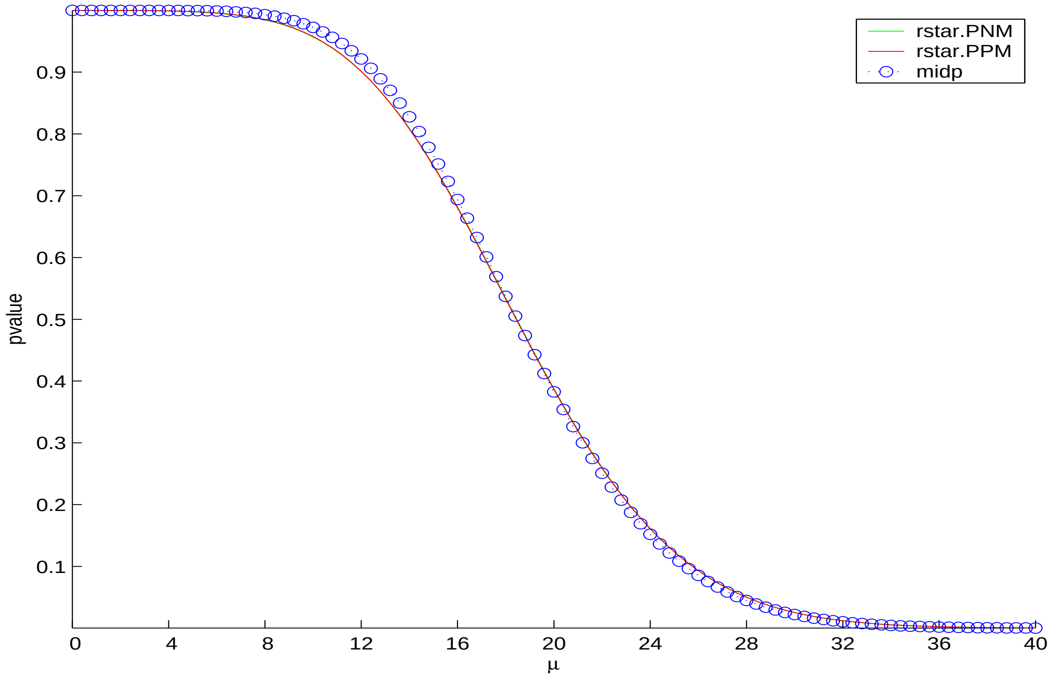

Figure 4: Comparison of the p-value functions computed by PP-model and PN-model in addition with the mid-p-value .

.3 Discussion....

1. In the cases when fewer events are observed in the signal measurement than background.

2. Extreme case: $x = 0/y = 0$, any adjustment necessary?

3. Confidence Intervals?

4. nuisance parameters?(uncertainty in the standard error of the background??)

.... some HEP and GRA datasets

We also look into several datasets from the HEP and High Energy Gamma Ray Astrophysics (GRA) literatures from [5], and we here follow his notations. Let $x = N_{on}$, called an on-source count, $y = N_{off}$, and $k = x + y$ for compactness. The background count mean's estimate is $b = \alpha N_{off}$, where α is the relative exposure of the two(on-source and off-source) observations, and the background count's uncertainty is $\delta b = \alpha\sqrt{N_{off}}$. The signal is estimated by $s = N_{on} - b = x - \alpha y$. The hypotheses test is $\mu_{on} = \alpha\mu_{off}$. Five different Z-statistics are used. They are Z_{Bi}, quantile from Binomial distribution; Z_P from Poisson measurement; Z_L, likelihood root, which is a 1st-order approximation; Z_3, 3rd-order approximation ignoring background error; and Z_{r^*}, the third order approximation with background error adjustment. Significant tests are operated and results are listed in Table II.

Table II Quantile Z's for testing $\mu_{on} = \alpha\mu_{off}$

$N_{on} = x$	4	6	9	17	50	67	200	523	16789	498426	2119449
$N_{off} = y$	5	18.78	17.83	40.11	55	15	10	2327	1864910	493434	23671193
α	0.2	0.0692	0.2132	0.0947	0.5	2.0	10.0	0.167	0.0891	1.000	0.891
b	1.0	1.3	3.8	3.8	27.5	30.0	100.0	388.6	166213	493434	2109732
s=x-b	3.0	4.7	5.2	13.2	22.5	37	100	134.4	1376	4992	9717
δb	0.45	0.3	0.9	0.6	3.71	7.75	31.6	8.1	121.7	702.4	433.6
Z_{Bi}	1.66	2.63	1.82	4.46	2.93	2.89	2.20	5.93	3.23	5.01	6.40
mid-Z_{Bi}	1.49	2.50	1.67	4.34	2.87	2.81	2.11	5.90	3.34	5.01	6.82
Z_P	2.08	2.84	2.14	4.87	3.80	5.76	INF	6.45	3.37	7.09	6.69
mid-Z_P	1.93	2.73	2.03	4.80	3.74	5.72	INF	6.44	3.37	7.09	6.68
1st Z_L	1.95	2.82	1.99	4.57	3.02	3.04	2.38	5.95	3.35	5.01	6.82
3rd Z_3	2.13	2.92	2.11	4.64	3.13	3.30	2.98	5.97	3.35	5.08	6.82
est'd bg Z_{r^*}	1.98	2.87	2.02	4.60	3.03	3.02	2.34	5.95	3.35	5.01	6.82

As a modification, mid-Z_{Bi} and mid-Z_P corresponding to mid-p-values are more appropriate to be considered. And Z-values are more consistent with large Astrophysics data compared with different test statistics, although our method is initially motivated by rare process with nuisance parameters. The 3rd-order approximation Z_3 without background error adjustment might not be good here due to unignorable imprecisions. When nuisance parameter is incorporated into measurements, Z-values calculated based on r^* are appreciated. Surprisingly, the Z-values calculated from 3rd-order approximations Z_{r^*} are close to Z-values calculated from likelihood root statistics Z_L.(??)

5 Summary

We describe the third order approximations provided by the observed likelihood associated with p-value function as a treatment of nuisance parameters arising in background measurements. For the case of Poisson signal with either Poisson noise or Normal noise, we have given out explicit results and shown that the method yields p-value functions with good performance, ...

6 Acknowledgements

References

[1] N. Reid, D.A.S. Fraser, *Likelihood Inference in the Presence of Nuisance Parameters*, PhyStat2003, 2003.

[2] D.A.S. Fraser, *Statistical Inference: Likelihood to Significance, J.Am. Statist. Assoc*, **86**, 258, 1990.

[3] D.A.S. Fraser, N. Reid, and J. Wu, *A Simple General Formula for Tail Probabilities for Frequentist and Bayesian Inference*, *Biometrika*, **86**, 246, 1999.

[4] N. Reid, *Asymptotics and the Theory of Inference*, *Ann.Statist.*, 2004.

[5] J.T.Linnemann, *Measure of Significance in HEP and Astrophysics*, PhyStat2003, 2003.