Likelihood Asymptotics and Beyond

PHYSTAT Dark Matter

Limit Setting Methods | August 1, 2019

Alessandra R. Brazzale

alessandra.brazzale@unipd.it

Department of Statistical Sciences | University of Padova

Conrad (2014)

Astroparticle Physics 62 (2015) 165–177

Contents lists available at ScienceDirect

Astroparticle Physics

journal homepage: www.elsevier.com/locate/astropart

Review

Statistical issues in astrophysical searches for particle dark matter

Jan Conrad

Oskar Klein Centre for Cosmoparticle Physics, Physics Department, Stockholm University, Albanova University Centre, SE-10691 Stockholm and Imperial College London, London SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 15 June 2014
Received in revised form 15 August 2014
Accepted 16 September 2014
Available online 27 September 2014

Keywords:
Dark matter
Statistical methods
Indirect detection
Direct detection

ABSTRACT

In this review statistical issues appearing in astrophysical searches for particle dark matter, i.e. indirect detection (dark matter annihilating into standard model particles) or direct detection (dark matter particles scattering in deep underground detectors) are discussed. One particular aspect of these searches is the presence of very large uncertainties in nuisance parameters (astrophysical factors) that are degenerate with parameters of interest (mass and annihilation/decay cross sections for the particles). The likelihood approach has become the most powerful tool, offering at least one well motivated method for incorporation of nuisance parameters and increasing the sensitivity of experiments by allowing a combination of targets superior to the more traditional data stacking. Other statistical challenges appearing in astrophysical searches are to large extent similar to any new physics search, for example at colliders, a prime example being the calculation of trial factors. Frequentist methods prevail for hypothesis testing and interval estimation, Bayesian methods are used for assessment of nuisance parameters and parameter estimation in complex parameter spaces. The basic statistical concepts will be exposed, illustrated with concrete examples from experimental searches and caveats will be pointed out.

$$n \longrightarrow +\infty$$

LARGE-SAMPLE THEORY

(asymptotic approximation)

Cash (1979)

THE ASTROPHYSICAL JOURNAL, **228**:939–947, 1979 March 15

PARAMETER ESTIMATION IN ASTRONOMY THROUGH APPLICATION OF THE LIKELIHOOD RATIO

WEBSTER CASH

Space Sciences Laboratory, Department of Physics, University of California, Berkeley

Received 1977 August 22; accepted 1978 August 24

ABSTRACT

Many problems in the experimental estimation of parameters for models can be solved through use of the likelihood ratio test. Applications of the likelihood ratio, with particular attention to photon counting experiments, are discussed. The procedures presented solve a greater range of problems than those currently in use, yet are no more difficult to apply. The procedures are proved analytically, and examples from current problems in astronomy are discussed.

Subject heading: functions: numerical methods

Cash (1979)

prove that Avni's technique was valid for nonlinear models as well. In this paper, I show that the theorem of Wilks has a far wider applicability that provides an answer for many of the outstanding problems of parameter estimation in astronomy.

Briefly, the theorem states that if one takes n samples $X_1, \ldots, X_n$ from a probability distribution $f(X; \theta_1, \ldots, \theta_p)$, where $\theta_1, \ldots, \theta_p$ are the parameters of the probability distribution, one should form the statistic

$$L = \frac{\max \prod_{i=1}^{n} f(X_i; \theta_1^T, \ldots, \theta_q^T, \theta_{q+1}, \ldots, \theta_p)}{\max \prod_{i=1}^{n} f(X_i; \theta_1, \ldots, \theta_p)},$$

where the maxima are found by varying the parameters. In the numerator, the θ^T terms represent parameters

939

which are not varied but held constant at their true values. If one then calculates the value of the quantity $-2 \ln L$, it will have a probability distribution like χ^2 with q degrees of freedom, except for a small additive term of size about $n^{-1/2}$.

Even more remarkable is the fact that essentially no constraints are set on the probability distributions $f(X; \theta_1, \ldots, \theta_p)$, only that they be well-behaved and convergent (see Paper I). *The probability distribution $f(X; \theta)$ need not be Gaussian. It can be almost anything*. In the limit of large n, $-2 \ln L$ will always be distributed as χ_q^2. This is the lever I shall use to generalize the approach to parameter estimation.

Cookbook | Ingredients

- sample: $y = (y_1, \ldots, y_n)$
- model function: $f(y; \theta)$, with $\theta \in \Theta \subseteq \mathbb{R}$

- log likelihood: $\ell(\theta) = \log\{f(y; \theta)\}$
- maximum likelihood estimate: $\hat{\theta} = \text{argsup}_\theta \ell(\theta)$

- score function: $u(\theta) = d\ell(\theta)/d\theta$
- observed information: $j(\theta) = -d^2\ell(\theta)/d\theta^2$
- expected information: $i(\theta) = E[j(\theta; Y)]$

Cookbook | Recipe 1

- likelihood root

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

- score statistic

$$s(\theta) = j(\hat{\theta})^{-1/2} u(\theta)$$

- Wald statistic

$$t(\theta) = j(\hat{\theta})^{1/2}(\hat{\theta} - \theta)$$

one-sided

$$\sim N(0,1) + O(n^{-1/2})$$

two-sided

$$\sim \chi^2_1 + O(n^{-1})$$

But. . . what if

$$n = 1$$

Toy example: exponential

observe $\quad y = 1 \quad$ from $\quad f(y;\theta) = \theta\exp(-\theta y), \quad \theta > 0$

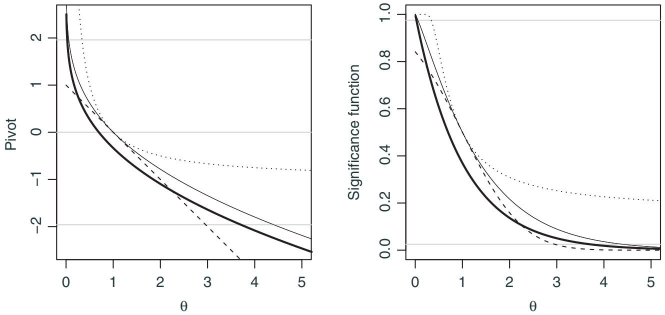


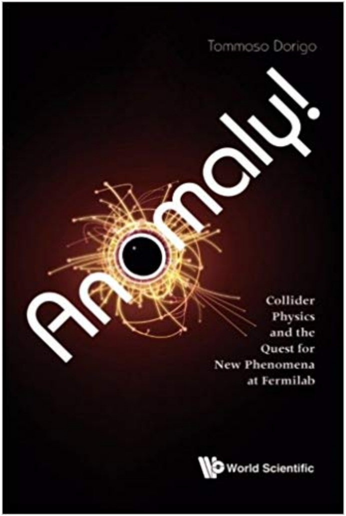
Figure 2.2 Approximate pivots and P-values based on an exponential sample of size $n = 1$. Left: likelihood root $r(\theta)$ (solid), score pivot $s(\theta)$ (dots), Wald pivot $t(\theta)$ (dashes), modified likelihood root $r^*(\theta)$ (heavy), and exact pivot $\theta\sum y_i$ indistinguishable from the modified likelihood root. The horizontal lines are at $0, \pm 1.96$. Right: corresponding significance functions, with horizontal lines at 0.025 and 0.975.

Toy example: Cauchy

$n=1$ observation from $f(y;\theta)=\frac{1}{\pi\{1+(y-\theta)^2\}}, \quad \theta \in \mathbb{R}$

Table 3.1 *The exact distribution function for a single observation from the Cauchy distribution with $\theta = 0$, with three first order approximations.*

y	0.1	1	3	5	10	15	25
$\Phi\{s(y)\}$	0.556	0.760	0.664	0.607	0.555	0.537	0.522
$\Phi\{t(y)\}$	0.556	0.921	0.999	1.000	1.000	1.000	1.000
$\Phi\{r(y)\}$	0.556	0.880	0.984	0.995	0.999	1.000	1.000
Exact	0.532	0.750	0.897	0.937	0.968	0.979	0.987



STATISTICS

$p < 0.05$

$\Downarrow$

$p < 0.00000029$

(5σ criterion)

(ASTRO)PHYSICS

Toy example: Cauchy

$n = 1$ observation from $f(y;\theta) = \frac{1}{\pi\{1+(y-\theta)^2\}}, \quad \theta \in \mathbb{R}$

Table 3.1 *The exact distribution function for a single observation from the Cauchy distribution with* $\theta = 0$*, with three first order and two third order approximations.*

y	0.1	1	3	5	10	15	25
$\Phi\{s(y)\}$	0.556	0.760	0.664	0.607	0.555	0.537	0.522
$\Phi\{t(y)\}$	0.556	0.921	0.999	1.000	1.000	1.000	1.000
$\Phi\{r(y)\}$	0.556	0.880	0.984	0.995	0.999	1.000	1.000
$\Phi\{r^*(y)\}$	0.535	0.772	0.918	0.953	0.979	0.987	0.993
Lugannani–Rice	0.535	0.768	0.909	0.944	0.972	0.981	0.989
Exact	0.532	0.750	0.897	0.937	0.968	0.979	0.987

3.3 Top quark

In 1995 the Fermi National Accelerator Laboratory announced the discovery of the top quark, the last of six quarks predicted by the ‘standard model’ of particle physics. Two simultaneous publications in *Physical Review Letters* described the experiments and the results of the analysis. Evidence for the existence of the top quark was based on signals from the decay of the top quark/antiquark pair created in proton/antiproton collisions. The analysis reported in one of the papers (Abe *et al.*, 1995) was based on a Poisson model:

$$f(y;\theta) = \theta^y e^{-\theta}/y!, \quad y = 0, 1, 2, \ldots, \quad \theta > b.$$

Here b is a positive background count of decay events, and we write $\theta = \mu + b$. Evidence that $\mu > 0$ would favour the existence of the top quark, if the underlying assumptions are correct. Table I of Abe *et al.* (1995) gives the observed values of y as $y^0 = 27$ and of b as 6.7. In fact b is also estimated, but with sufficient precision that this minor complication may be ignored here.

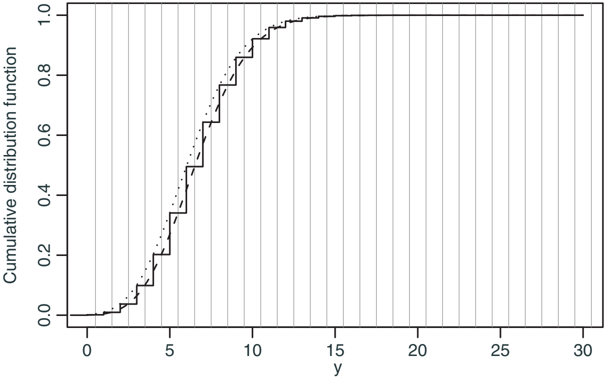


Figure 3.2 Cumulative distribution function for Poisson distribution with parameter 6.7 (solid), with approximations $\Phi\{r^*(y)\}$ (dashes) and $\Phi\{r^*(y + 1/2)\}$ (dots). The vertical lines are at 0.5, 1.5, 2.5, . . .

Statistical Science
2008, Vol. 23, No. 3, 354–364
DOI: 10.1214/08-STS260

The Banff Challenge: Statistical Detection of a Noisy Signal

A. C. Davison and N. Sartori

Abstract. Particle physics experiments such as those run in the Large Hadron Collider result in huge quantities of data, which are boiled down to a few numbers from which it is hoped that a signal will be detected. We discuss a simple probability model for this and derive frequentist and noninformative Bayesian procedures for inference about the signal. Both are highly accurate in realistic cases, with the frequentist procedure having the edge for interval estimation, and the Bayesian procedure yielding slightly better point estimates. We also argue that the significance, or p-value, function based on the modified likelihood root provides a comprehensive presentation of the information in the data and should be used for inference.

Key words and phrases: Bayesian inference, higher-order asymptotics, Large Hadron Collider, likelihood, noninformative prior, orthogonal parameter, particle physics, Poisson distribution, signal detection.

LARGE-SAMPLE THEORY

(asymptotic approximation)

$$\Downarrow$$

SMALL-SAMPLE THEORY

(higher order asymptotics)

Edgeworth expansions

*J*inst — Published by IOP Publishing for Sissa Medialab

Received: *July 27, 2018*
Revised: *October 17, 2018*
Accepted: *November 20, 2018*
Published: *December 6, 2018*

Improved inference for the signal significance

I. Volobouev[a,1] **and A. Trindade**[b]

[a] *Department of Physics & Astronomy, Texas Tech University, Lubbock, TX, 79409-1051, U.S.A.*

[b] *Department of Mathematics & Statistics, Texas Tech University, Lubbock, TX, 79409-1042, U.S.A.*

E-mail: i.volobouev@ttu.edu

Abstract: We study the properties of several likelihood-based statistics commonly used in testing for the presence of a known signal under a mixture model with known background, but unknown signal fraction. Under the null hypothesis of no signal, all statistics follow a standard normal distribution in large samples, but substantial deviations can occur at practically relevant sample sizes. Approximations for respective p-values are derived to various orders of accuracy using the methodology of Edgeworth expansions. Adherence to normality is studied, and the magnitude of deviations is quantified according to resulting p-value inflation or deflation. We find that

2018 JINST 13 P

Modified likelihood root

$$r^*(\theta) = r(\theta) + \frac{1}{r(\theta)} \log\left\{\frac{q(\theta)}{r(\theta)}\right\}$$

$$\sim N(0,1) + O(n^{-3/2})$$

Lugannani-Rice tail approximation

$$\Phi^*(r) = \Phi(r) + \phi(r)\left(\frac{1}{r} - \frac{1}{q}\right)$$

Saddlepoint approximation

- invariant wrt interest-respecting reparametrizations
- simplifies for special classes of models (exponential family, regression-scale model, ...)

- nuisance parameters (ψ, λ): several formulations (Skovgaard, 1996 | Fraser et al. 1999 | ...)

- vector θ: w^* (Bartlett, 1937 | Skovgaard, 2001)
- directional *p*-values (Fraser et al., 2014, 2016)

- Bayesian counterparts (Tierney & Kadane, 1986 | ...)

Amuse-bouche

$$r^*(\psi) = r(\psi) + \frac{1}{r(\psi)} \log \left\{ \frac{q(\psi)}{r(\psi)} \right\}$$

$$q(\psi) = \frac{|\varphi(\hat{\theta}) - \varphi(\hat{\theta}_\psi) \ \varphi_\lambda(\hat{\theta}_\psi)|}{|\varphi_\theta(\hat{\theta})|} \left\{ \frac{|j(\hat{\theta})|}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|} \right\}^{1/2}$$

$$\varphi(\theta)^\top = \sum_{k=1}^{K} \frac{\partial \ell(\theta; y)}{\partial y_k} \, |_{y=y^0} \ V_k$$

$$V_k = \frac{\partial E[Y_k; \theta]}{\partial \theta^\top} \, |_{\theta=\hat{\theta}}$$

Today's special: two-for-one

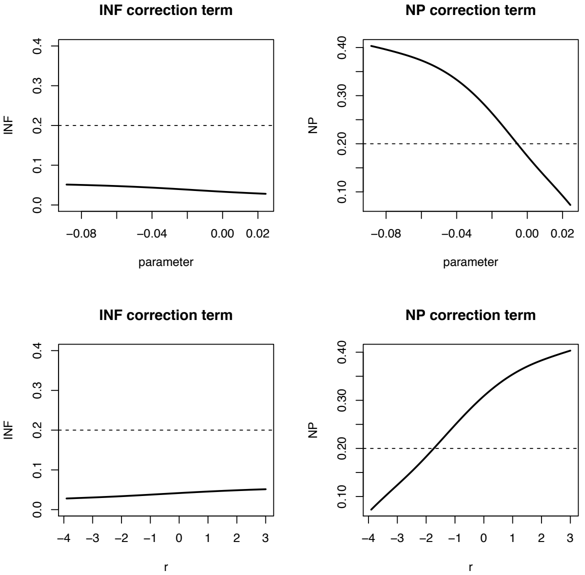
- **nuisance parameters**: $\theta = (\psi, \lambda)$

INCIDENTAL PARAMETERS !

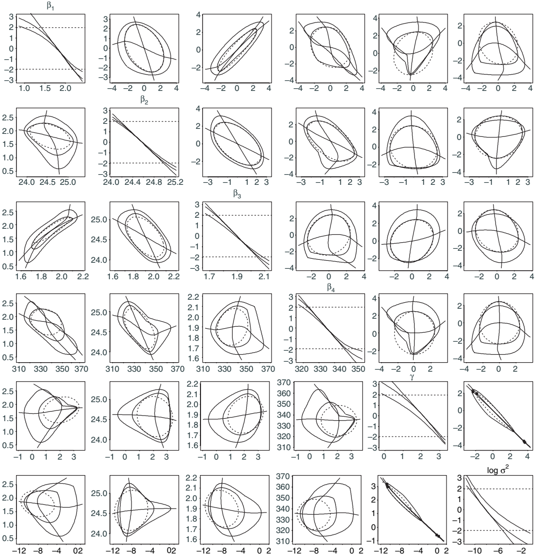
$$\frac{p}{n} \longrightarrow 0$$

$$r^*(\psi) = r(\psi) + r_{inf}(\psi) + r_{np}(\psi)$$

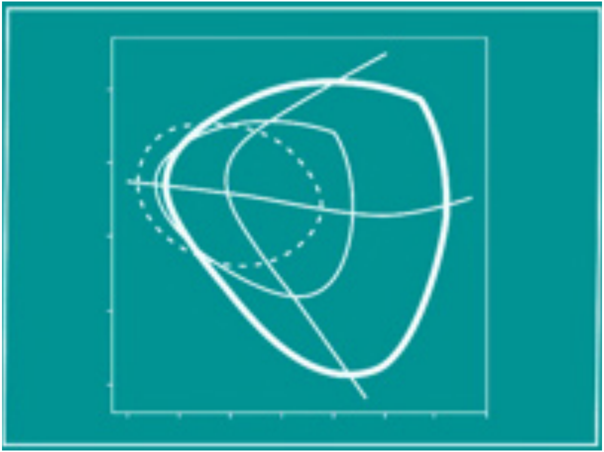
INF & NP correction terms



Model diagnostics



Contour plots



Rolke et al. (2005)

ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Nuclear Instruments and Methods in Physics Research A 551 (2005) 493–503

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH
Section A

www.elsevier.com/locate/nima

Limits and confidence intervals in the presence of nuisance parameters

Wolfgang A. Rolke[a,*], Angel M. López[b], Jan Conrad[c]

[a]*Department of Mathematics, University of Puerto Rico - Mayagüez, P.O. Box 5959, Mayagüez, PR 00681, USA*
[b]*Department of Physics, University of Puerto Rico - Mayagüez, Mayagüez, PR 00681, USA*
[c]*PH-Department, CERN, CH-1211, Geneva 23, Switzerland*

Received 7 February 2005; received in revised form 13 May 2005; accepted 18 May 2005
Available online 11 July 2005

Abstract

We study the frequentist properties of confidence intervals computed by the method known to statisticians as the Profile Likelihood. It is seen that the coverage of these intervals is surprisingly good over a wide range of possible parameter values for important classes of problems, in particular whenever there are additional nuisance parameters with statistical or systematic errors. Programs are available for calculating these intervals.

Conrad (2014)

alue of the primary param-
equiring:

(11)

ce of $[X_1, X_2]$ is not unique,
ondition (for example that
bove X_2 should be $\alpha/2$). It
ns [10] in 1998 that exper-
e after the observation is
erage property, otherwise
o include observations into

(12)

of the first kind (reject the null hypothesis though it is true), whereas the second property is called the power, i.e. the probability to accept the alternative hypothesis when it is true. In general, the most efficient way to perform a hypothesis test is to transform the data into a one dimensional *test statistic*, which preferably has a known distribution under the null hypothesis. The *p*-value of an experimental observation is defined as the probability of obtaining an observation as likely or less likely than the experimental observation itself, and the null hypothesis is rejected if the *p*-value $< \alpha$, the most common choice for α being $\sim 3 \cdot 10^7$, i.e. the 5σ tail probability of a Gaussian distribution. The most common hypothesis test used in searches for dark matter is based on (variants) of the likelihood ratio test. For simple (i.e. fully specified) hypotheses the likelihood ratio:

[5] It is in principle conceivable that this best estimate is not obtained from maximizing the likelihood, though we are not aware of any studies made what the implications of such an approach would be.

Adjusted likelihood function

$$L_a(\psi) = L_p(\psi) M(\psi)$$

derived from profile likelihood function

$$L_p(\psi) = \max_{\lambda} L(\psi, \lambda)$$

- approximates exact conditional/marginal likelihood function (where available) to the order $O(n^{-1})$
- behaves like an ordinary likelihood function

Adjusted MLE

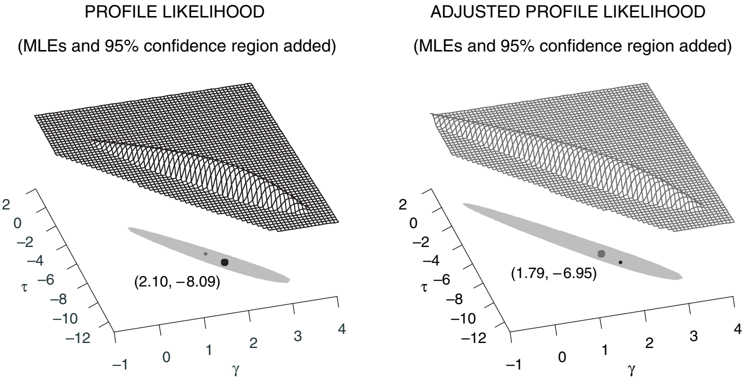


Figure 5.7 RIA data: profile and adjusted profile log likelihoods for the variance parameters γ and $\tau = \log \sigma^2$, maximum likelihood estimates, and corresponding 95% confidence regions. Based on the output of the `mpl` fitting routine of the `nlreg` package.

Code

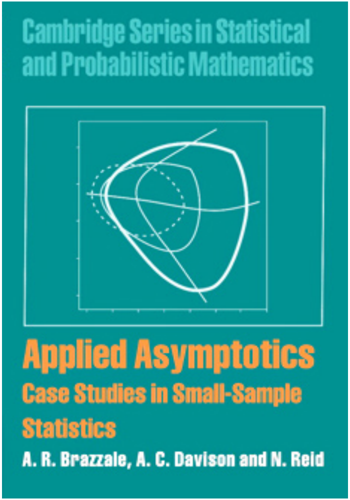
R packages

- cond | marg | nlreg | csampling (Bellio & Brazzale)
- hoa (Fortin & Davison, Bellio & Brazzale)

https://ruggerobellio.weebly.com/software.html

- likelihoodAsy
- hoacoxph
- …

Want to know more?



Thank you for your attention!