Statisticians Summary

PHYSTAT Dark Matter

Summaries | August 2, 2019

Alessandra R. Brazzale
alessandra.brazzale@unipd.it

Department of Statistical Sciences | University of Padova

Premise

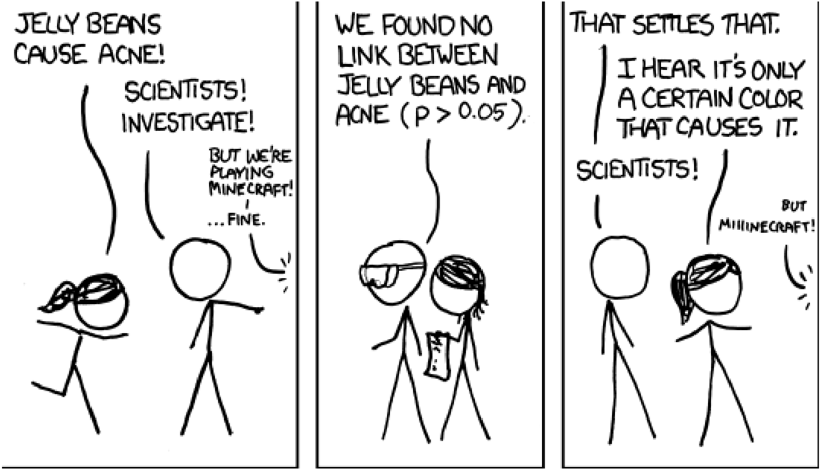
Credits: David van Dyk (ICL), PhyStat-ν 2016, 2019

- Statisticians “discuss” talks.
- I’m sure I missed some important points. And I may have misinterpreted your work.
- So, feel free to correct me!
- And, please, be patient. I know you aim for answers… but I have far more questions.

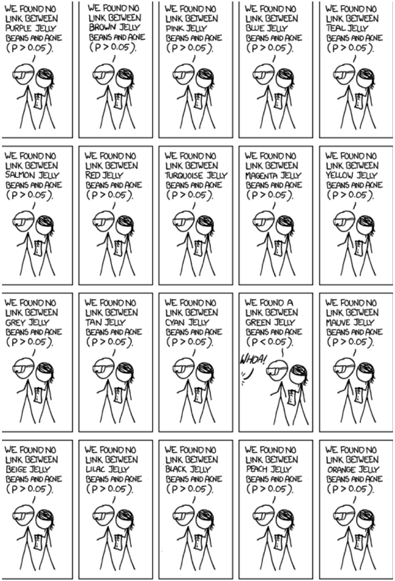
Physics/Statistics glossary

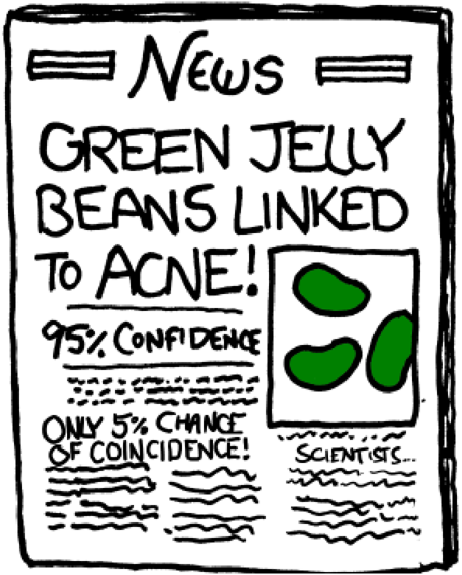
- parameter determination = estimation
- look elsewhere effect = **multiple comparison**

Look elsewhere effect



Look elsewhere effect




NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

Physics/Statistics glossary

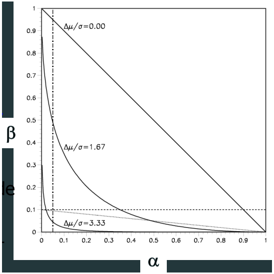
- parameter determination = estimation
- look elsewhere effect = multiple comparison
- systematic uncertainty = nuisance parameters
- profiled value = constrained estimate
- toy simulation = (boostrap) resampling
- error propagation = delta method
- half chi-square distribution = chi-square bar distribution
- chi-square method = least squares criterion
- M(achine) L(earning) = M(aximum) L(ikelihood)
- …

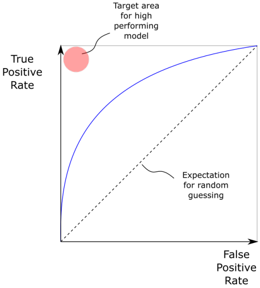
Some possible false friends

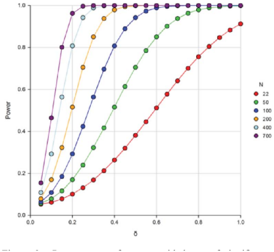
- marginal likelihood = integrated likelihood
- ancillary information $\neq$ ancillary statistic
- recoiling(?) $\neq$ correlation
- ...

ROC & power curves

- CLs = p1/(1-p0) = ROC curve + power curve







[Lyons]

$$(\alpha, \beta, \delta, \sigma, n)$$

Still looking for a match

- blind analysis [Lyons, Manzani, Loer]
  $\rightarrow$ (validation in machine learning?) finite population sampling? (other?)
- salting [Manzani]
  $\rightarrow$?
- bias?
  $\rightarrow$ confounding?
- **power constrained limits**
  $\rightarrow$?
- …

Statistical Science
2002, Vol. 17, No. 2, 149–172

Setting Confidence Intervals for Bounded Parameters

Mark Mandelkern

Abstract. Setting confidence bounds is an essential part of the reporting of experimental results. Current physics experiments are often done to measure nonnegative parameters that are small and may be zero and to search for small signals in the presence of backgrounds. These are examples of experiments which offer the possibility of yielding a result, recognized a priori to be relatively improbable, of a negative estimate for a quantity known to be positive. The classical Neyman procedure for setting confidence bounds in this situation is arguably unsatisfactory and several alternatives have been recently proposed. We compare methods for setting Gaussian and Poisson confidence intervals for cases in which the parameter to be estimated is bounded. These procedures lead to substantially different intervals when a relatively improbable observation implies a parameter estimate beyond the bound.

Key words and phrases: Confidence bounds, Poisson-with-background, Gaussian-with-boundary.

Fraser et al. (2004)

PHYSICAL REVIEW D **69**, 033002 (2004)

Inference for bounded parameters

D. A. S. Fraser* and N. Reid†
Department of Statistics, University of Toronto, 100 St. George St., Toronto, Canada M5S 3G3

A. C. M. Wong‡
SASIT, Atkinson Faculty, York University, Toronto, Canada M3J 1P3
(Received 10 March 2003; published 23 February 2004)

The estimation of the signal frequency count in the presence of background noise has been widely discussed in recent physics literature, and Mandelkern [Stat. Sci. **17**, 149 (2002)] brings the central issues to the statistical community, leading in turn to extensive discussion by statisticians. The primary focus however of Mandelkern and the accompanying discussion is on the construction of a confidence interval. We argue that the likelihood function and *p*-value function provide a comprehensive presentation of the information available from the model and the data. This is illustrated for Gaussian and Poisson models with lower bounds for the mean parameter.

DOI: 10.1103/PhysRevD.69.033002 PACS number(s): 02.50.Tt, 06.20.Dk

Take home messages

Great conference!

I learnt that…

- experiments are expensive and time-consuming [Lyons]
- computing time matters [Kahlhoefer]
- **complex models** and complex likelihoods [Kahlhoefer]

Conrad (2014)

is not unique,
example that
ıld be $\alpha/2$). It
98 that exper-
ıbservation is
ty, otherwise
ervations into

(12)

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.

Ongoings

Complex data structures

- larger/harder problems
- full joint distribution has complex dependence structure on the parameter and/or the data

  spatial/spatio-temporal processes, clustered/hierarchical data, time-course observations, . . .
- plausible models exists, but difficult to evaluate or not fully reliable

Pseudo likelihoods

- regular likelihood provides the starting point
- try to retain "good" properties of likelihood (efficiency, asymptotic normality, . . .)
- model-based: ignore part of the data (*partial likelihood*). . .
- . . . or of the parameter (*quasi likelihood*)
- derived from an estimating equation (GEE)

Examples of pseudo likelihood

- partial likelihood (Cox, 1975)
- quasi likelihood (Wedderburn, 1974 | Liang & Zeger, 1986)
- REML (Patterson & Thompson, 1971)
- composite likelihood (Besag, 1974 | Lindsay, 1988 | Varin, 2008)

Take home messages

Great conference!

I learnt that…

- experiments are expensive and time-consuming [Lyons]
- computing time matters [Kahlhoefer]
- complex models and complex likelihoods [Kahlhoefer]
- background is a problem [Conrad, …, THU + FRI, …, Algeri]
- nuisance parameters are (less?) a problem [Battey]
- statistical uncertainty vs systematic uncertainty (the first weights more than the second) [Agostini]
- …

Take home messages

You seem to like…

- Poisson distribution (+ COM-Poisson) [Durnford, Pollmann]
- profile likelihood [Dobson, Wardle, …]
- Wilks' theorem
- mixture models
- *p*-values (frequentist/Fisherian inference)…
- …though some seem to prefer **Bayes factors**
- …

Two or three?

- *p*-values vs. Bayes factor
  $\rightarrow$ D. van Dyk, PhyStat-ν 2016, 2019, *matching prior*

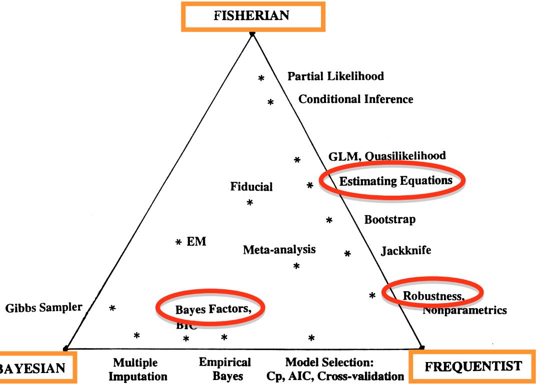


FIG. 8. *A barycentric picture of modern statistical research, showing the relative influence of the Bayesian, frequentist and Fisherian philosophies upon various topics of current interest.*

Take home messages

You know that…

- p-values difficult to calculate
- simulation not always feasible [Kahlhoefer]
- challenges with discrete data [Agostini]
- issues with Bayesian construction [Agostini]
- **nonregular problems** [Kahlhoefer]
- …

Cash (1979)

940 CASH Vol. 228

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 $\chi_q{}^2$. This is the lever I shall use to generalize the approach to parameter estimation.

. . . well behaved and convergent

The Astrophysical Journal, 571:545–559, 2002 May 20

STATISTICS, HANDLE WITH CARE: DETECTING MULTIPLE MODEL COMPONENTS WITH THE LIKELIHOOD RATIO TEST

Rostislav Protassov and David A. van Dyk
Department of Statistics, Harvard University, 1 Oxford Street, Cambridge, MA 02138; protasso@stat.harvard.edu, vandyk@stat.harvard.edu

Alanna Connors
Eureka Scientific, 2452 Delmer Street, Suite 100, Oakland, CA 94602-3017; connors@frances.astro.wellesley.edu

and

Vinay L. Kashyap and Aneta Siemiginowska
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138;
kashyap@head-cfa.harvard.edu, aneta@head-cfa.harvard.edu
Received 2001 June 1; accepted 2002 January 25

ABSTRACT

The likelihood ratio test (LRT) and the related *F*-test, popularized in astrophysics by Eadie and coworkers in 1971, Bevington in 1969, Lampton, Margon, & Bowyer, in 1976, Cash in 1979, and Avni in 1978, do not (even asymptotically) adhere to their nominal χ^2 and *F*-distributions in many statistical tests common in astrophysics, thereby casting many marginal line or source detections and nondetections into doubt. Although the above authors illustrate the many legitimate uses of these statistics, in some important cases it can be impossible to compute the correct false positive rate. For example, it has become common practice to use the LRT or the *F*-test to detect a line in a spectral model or a source above background despite the lack of

Protassov et al. (2002)

APPENDIX A

REGULARITY CONDITIONS FOR THE LRT

Here we state the regularity conditions required for the standard asymptotic behavior of the LRT. (Our presentation follows Serfling 1980, pp. 138–160, which should be consulted for details.) Let $X_1, \dots, X_n$ be independent identically distributed random variables with distribution $F(x;\ \boldsymbol{\theta})$ belonging to a family $\{F(x;\ \boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta\}$, where $\Theta \subset \mathscr{R}^k$ is open and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. $F(x;\ \boldsymbol{\theta})$ are assumed to possess densities or mass functions $f(x;\ \boldsymbol{\theta})$ that satisfy the following conditions:

1. For each $\boldsymbol{\theta} \in \Theta$, each $i = 1, \dots, k$, each $j = 1, \dots, k$, and each $l = 1, \dots, k$, the derivatives

$$\frac{\partial \log f(x;\boldsymbol{\theta})}{\partial \theta_i}, \qquad \frac{\partial^2 \log f(x;\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}, \qquad \frac{\partial^3 \log f(x;\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_l} \tag{A1}$$

exist, all x.[21]

2. For each $\boldsymbol{\theta}_* \in \Theta$, there exist functions $h_1(x)$, $h_2(x)$, and $h_3(x)$ (possibly depending on $\boldsymbol{\theta}_*$) such that for $\boldsymbol{\theta}$ in a neighborhood $N(\boldsymbol{\theta}_*) \subset \Theta$, the relations

$$\left|\frac{\partial f(x;\ \boldsymbol{\theta})}{\partial \theta_i}\right| \le h_1(x), \qquad \left|\frac{\partial^2 f(x;\ \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right| \le h_2(x), \qquad \left|\frac{\partial^3 \log f(x;\ \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_l}\right| \le h_3(x) \tag{A2}$$

hold, for all x and all $1 \le i, j, l \le k$, with

$$\int h_1(x)dx < \infty, \qquad \int h_2(x)dx < \infty, \qquad \int h_3(x) f(x;\ \boldsymbol{\theta})dx < \infty \text{ for } \boldsymbol{\theta} \in N(\boldsymbol{\theta}_*)\ . \tag{A3}$$

3. For each $\boldsymbol{\theta} \in \Theta$, the information matrix

Nonregular problems

Working Paper Series, N. 4, September 2018

iences

Likelihood Asymptotics in Nonregular Settings

A Review with Emphasis on the Likelihood Ratio

Brazzale Alessandra R.
Department of Statistical Sciences
University of Padua
Italy

Valentina Mameli
Department of Environmental Sciences, Informatics and Statistics
Ca' Foscari University of Venice
Italy

http://paduaresearch.cab.unipd.it/11306/

Regularity conditions

Condition 1 All components of θ are identifiable. That is, two model functions, $f(y;\theta^1)$ and $f(y;\theta^2)$, defined by any two different values $\theta^1 \neq \theta^2$ of θ, are distinct almost surely.

FINITE MIXTURE MODELS

Condition 2 The support of $f(y;\theta)$ does not depend on any component of θ.

CHANGE-POINT DETECTION

Condition 3 The parameter space Θ is a compact subspace of $\mathbb{R}^p$, for a fixed value of $p \in \mathbb{N}\backslash\{0\}$, and the true value θ^0 of θ is an interior point of Θ.

BOUNDARY PROBLEMS

Regularity conditions

Condition 4 The partial derivatives of the log-likelihood function $\ell(\theta; y)$ with respect to θ up to the order three exist in a neighbourhood of the true parameter value θ^0 almost surely. Furthermore, in such a neighbourhood, n^{-1} times the absolute value of the log-likelihood derivatives of order three are bounded above by a function of Y whose expectation is finite.

NON GAUSSIANITY

Condition 5 The first two Bartlett identities hold, which imply that

$$E[u(\theta; Y)] = 0 \quad \text{and} \quad i(\theta) = \mathrm{Var}(u(\theta; Y)).$$

FINITE MIXTURE MODELS

Curiouser and curiouser!

Limiting distributions

- chi-bar squared

$$\Pr(\bar{\chi}^2 \leq c) = \sum_{\nu=0}^{N} \omega_\nu \Pr(\chi^2_\nu \leq c)$$

- supremum of Gaussian random process

$$\{\sup_{|t| \leq M} Z(t)\}^2 + W, \quad W \sim \chi^2_1$$

- extreme value theory
- may need to bound parameter space

Algeri et al. (2016)

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **458,** L84–L88 (2016) doi:10.1093/mnrasl/slw025
Advance Access publication 2016 February 15

A method for comparing non-nested models with application to astrophysical searches for new physics

Sara Algeri,[1,2★] Jan Conrad[1,2,3†] and David A. van Dyk[1]

[1] *Statistics Section, Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK*
[2] *Department of Physics, Stockholm University, AlbaNova, SE-106 91 Stockholm, Sweden*
[3] *The Oskar Klein Centre for Cosmoparticle Physics, AlbaNova, SE-106 91 Stockholm, Sweden*

Accepted 2016 February 10. Received 2016 February 10; in original form 2015 December 25

ABSTRACT

Searches for unknown physics and decisions between competing astrophysical models to explain data both rely on statistical hypothesis testing. The usual approach in searches for new physical phenomena is based on the statistical likelihood ratio test and its asymptotic properties. In the common situation, when neither of the two models under comparison is a special case of the other i.e. when the hypotheses are non-nested, this test is not applicable. In astrophysics, this problem occurs when two models that reside in different parameter spaces are to be compared. An important example is the recently reported excess emission in astrophysical γ-rays and the question whether its origin is known astrophysics or dark matter. We develop and study a new, simple, generally applicable, frequentist method and validate its statistical properties using a suite of simulations studies. We exemplify it on realistic simulated data of the *Fermi*-Large Area Telescope γ-ray satellite, where non-nested hypotheses testing appears in the search for particle dark matter.

Key words: astroparticle physics – methods: data analysis – methods: statistical – dark matter.

To wind up

Some statistics landmarks

- Fisher (1925): contains most of the main elements of Fisherian inference

  sufficiency, likelihood, MLE, consistency, Fisher information, efficiency, asymptotic normality, . . .

- Fisher (1934): fully develops concept of *ancillarity*
- Wilks (1938): asymptotic distribution of the *likelihood ratio*

To wind up

Fisherian inference

- central role of likelihood function
- evaluation under repeated sampling
- conditioning on the relevant aspects of the data…
- … and to eliminate nuisance parameters
- practice-oriented

Neo-Fisherian inference

- Fisherian statistics is not a dead language: it continues to inspire new research.
- Yet, Fisher's language is not the only language in town…

Bayesian/Frequentis dialogue

- both use the likelihood
- empirical Bayes (meta analysis, hierarchical models, ...)
- matching priors

a posteriori credible intervals = standard confindence intervals

To wind up

Data mining/machine learning [Kieseler]

- different modes of attack on different types of inferential problems
  ("case-to-case expediency")

Simola et al. (2019)

Jinst

PUBLISHED BY IOP PUBLISHING FOR SISSA MEDIALAB

RECEIVED: *October 23, 2018*
REVISED: *January 11, 2019*
ACCEPTED: *February 22, 2019*
PUBLISHED: *March 6, 2019*

Machine learning accelerated likelihood-free event reconstruction in dark matter direct detection

U. Simola,[a] B. Pelssers,[b,1] D. Barge,[b] J. Conrad[b] and J. Corander[a]

[a] *Department of Mathematics and Statistics, University of Helsinki, Pietari Kalmin katu 5, Helsinki, Finland*

[b] *Department of Physics, Stockholm University, Roslagstullsbacken 21 A, Stockholm, Sweden*

E-mail: bart.pelssers@fysik.su.se