Statistical Issues for Dark Matter Searches

Louis Lyons

Imperial College

DMUK, Edinburgh
July 2017

Theme:

Using data to make judgements about H1 (Bgd+ DM) versus
H0 (just Bgd)

Why Statistics?

Experiments are expensive and time-consuming
so
Worth investing effort in statistical analysis
→ better information from data

Possible Topics:

- Blind Analysis
- Why 5σ for discovery?
- Significance
- P(A|B) ≠ P(B|A)
- Meaning of p-values
- Wilks' Theorem
- LEE = Look Elsewhere Effect
- Background Systematics
- Coverage
- p_0 v p_1 plots
- Upper Limits

(N.B. Several of these topics have no unique solutions from Statisticians)

Conclusions

Statistical Procedures

Parameter Determination / Upper Limits

e.g. $M_{Higgs} = 80 \pm 2$

Flux of WIMPs < ? in given mass range

Goodness of Fit

Is data consistent with 'No WIMPs' ?

Hypothesis Testing

Which theory fits data better?

e.g. D.M. or no D.M. = Discovery or Exclusion (or cannot decide)

Decision Theory

What expt should I do next?

Involves cost functions

Data

1) Counting expt = 1 bin

N_{obs} counts, with estimated bgd b ($\pm\sigma_b$)

2) On-off problem = 2 bins

N counts in signal region, M counts in bgd

3) Distribution F(x) = many bins

Fit with $B(x) + \mu S(x)$

BAYES and FREQUENTISM:
The Return of an Old Controversy

WHAT IS PROBABILITY?

MATHEMATICAL

Formal

Based on Axioms

FREQUENTIST

Ratio of frequencies as n→ infinity

Repeated “identical” trials

Not applicable to **single event** or **physical constant**

BAYESIAN Degree of belief

Can be applied to single event or physical constant

(even though these have unique truth)

Varies from person to person ***

Quantified by “fair bet”

LEGAL PROBABILITY

Bayesian

$$P(A;B) = \frac{P(B;A) \times P(A)}{P(B)}$$

Bayes' Theorem

p(param | data) α p(data | param) * p(param)

↑ posterior ↑ likelihood ↑ prior

Problems: p(param) Has particular true value

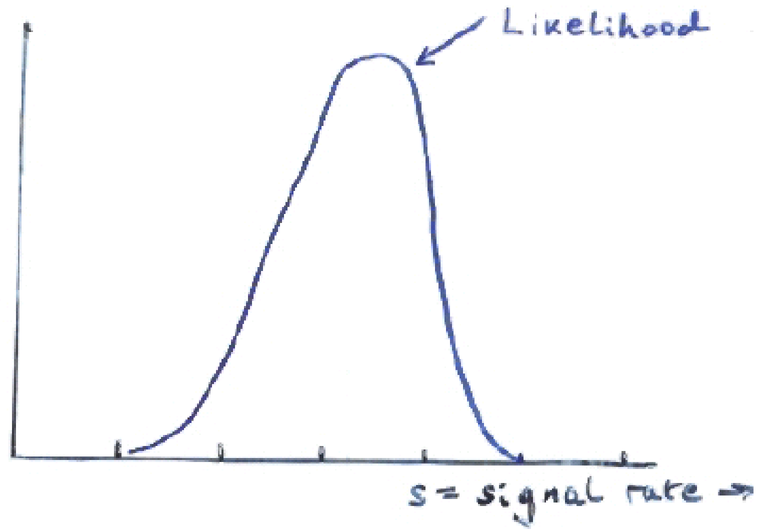
"Degree of belief"

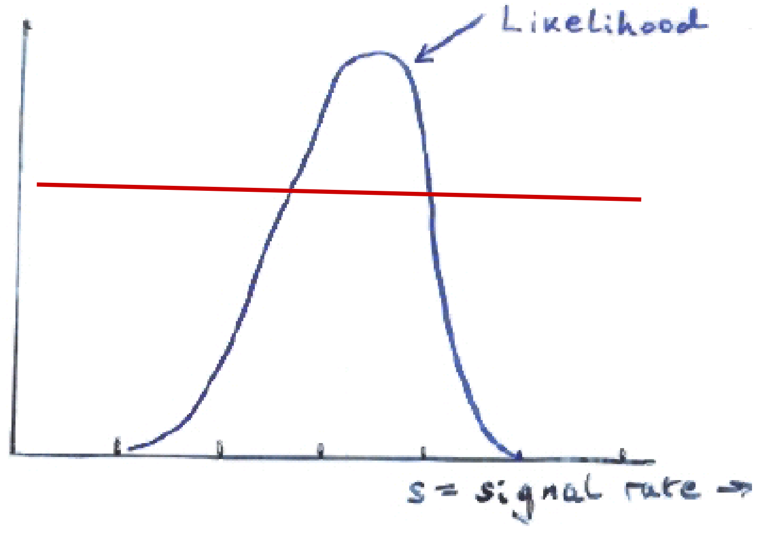
Prior What functional form?

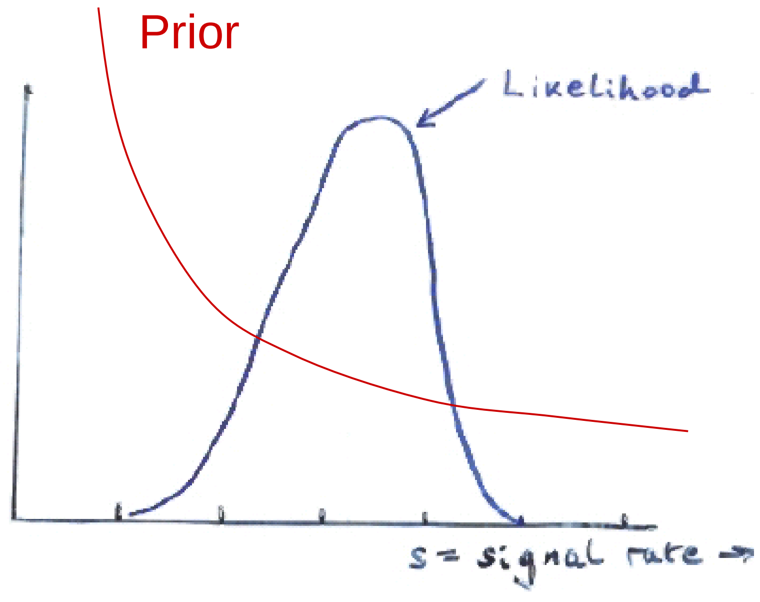
Uninformative prior: flat?

In which variable? e.g. m, m^2, ln m,….?

"Priors may be OK for parametrising prior knowledge, but not really for prior ignorance"







Even more important for UPPER LIMITS

Classical (Neyman) Confidence Intervals

Uses only P(data|theory)

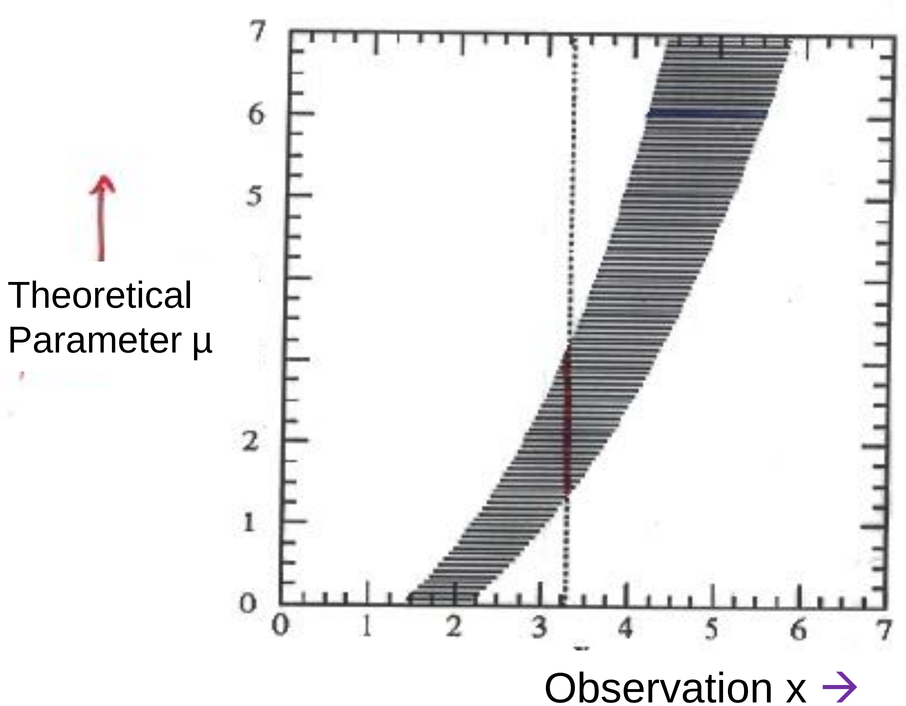


FIG. 1. A generic confidence belt construction and its use. For each value of μ, one draws a horizontal acceptance interval $[x_1, x_2]$ such that $P(x \in [x_1, x_2] | \mu) = \alpha$. Upon performing an experiment to measure x and obtaining the value x_0, one draws the dashed vertical line through x_0. The confidence interval $[\mu_1, \mu_2]$ is the union of all values of μ for which the corresponding acceptance interval is intercepted by the vertical line.

Specific Example

μ = Temp at centre of Sun

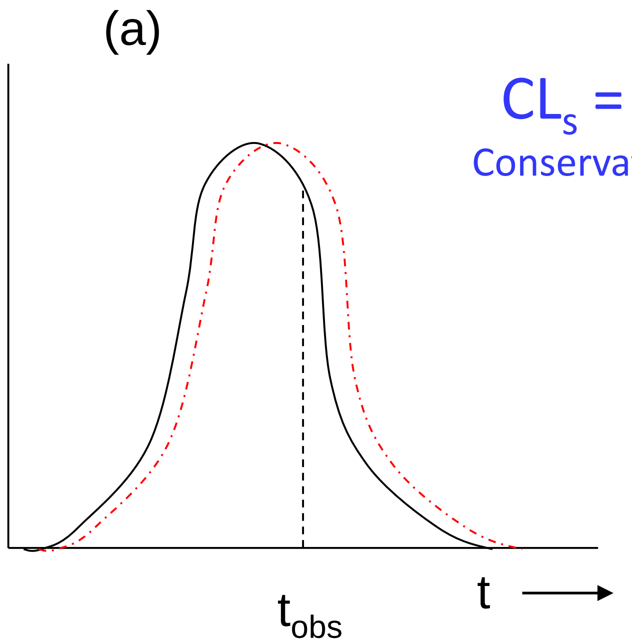
x = Measured solar ν flux

μ≥0

No prior for μ

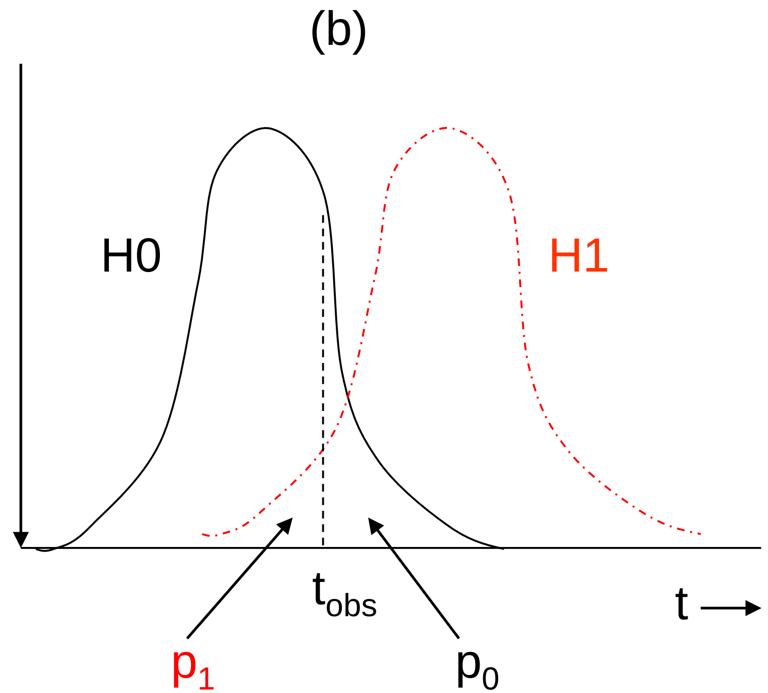
Methods for Upper Limits

(a) p-values
(b) Likelihoods
(c) χ^2 (Neyman or Pearson versions)
(d) Bayesian methods
    Sensitive to priors
(e) Neyman construction for Upper Limits
(f) Feldman-Cousins
(g) $CL_s = p_1/(1-p_0)$

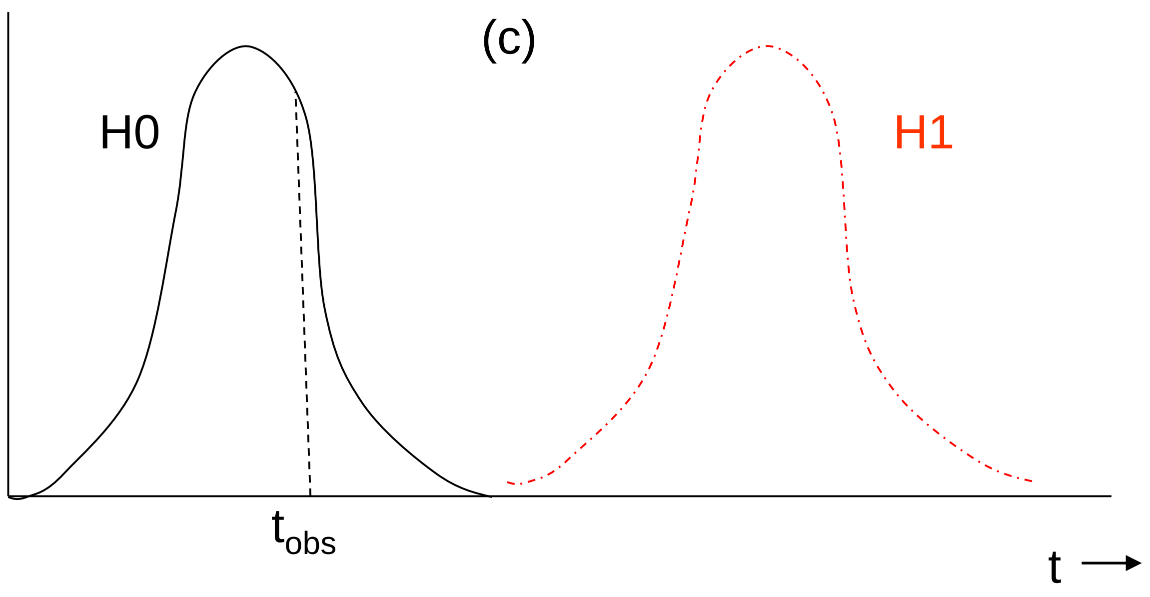


$CL_s = p_1/(1-p_0)$

Conservative frequentist



With 2 hypotheses, each with own pdf, p-values are defined as tail areas, pointing in towards each other



90% Classical 2-sided interval for Gaussian

σ = 1 $\mu \geq 0$ e.g. $m^2(\nu_e)$

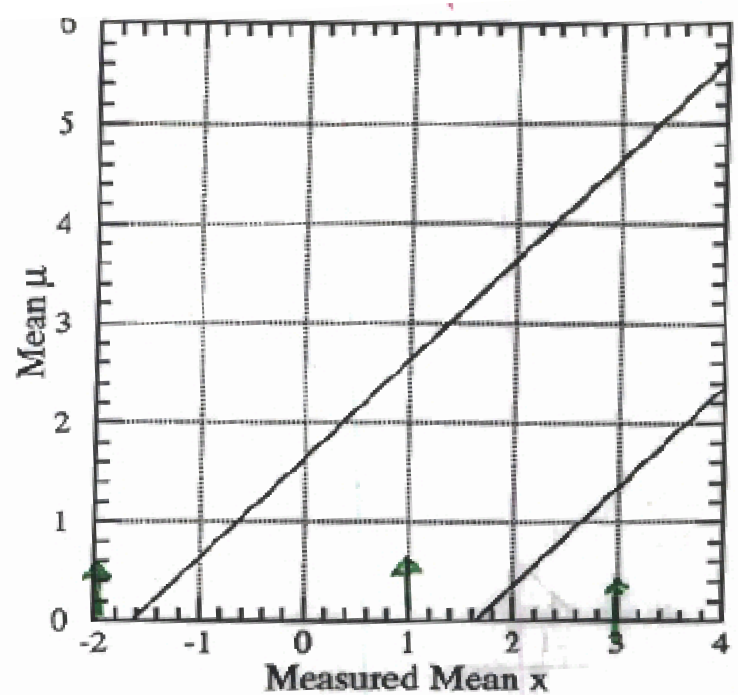


FIG. 3. Standard confidence belt for 90% C.L. central confidence intervals for the mean of a Gaussian, in units of the rms deviation.

$X_{obs} = 3$ Two-sided range for μ

$X_{obs} = 1$ Upper limit for μ =2.6

$X_{obs} = -2$ No region for μ

90% Classical Upper Limit for Gaussian

$\sigma = 1 \qquad \mu \geq 0 \qquad$ e.g. $m^2(\nu_e)$

Conclusion:

Be very explicit about what your procedure is

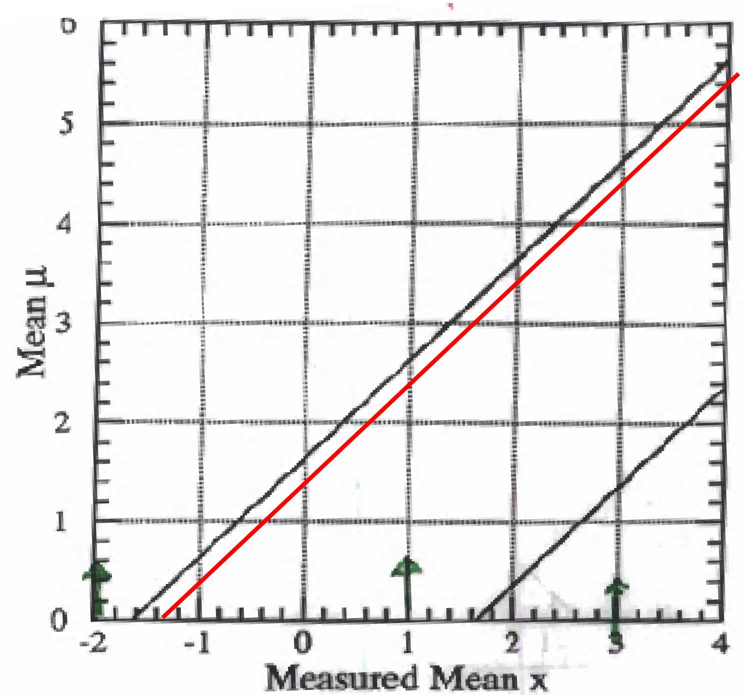
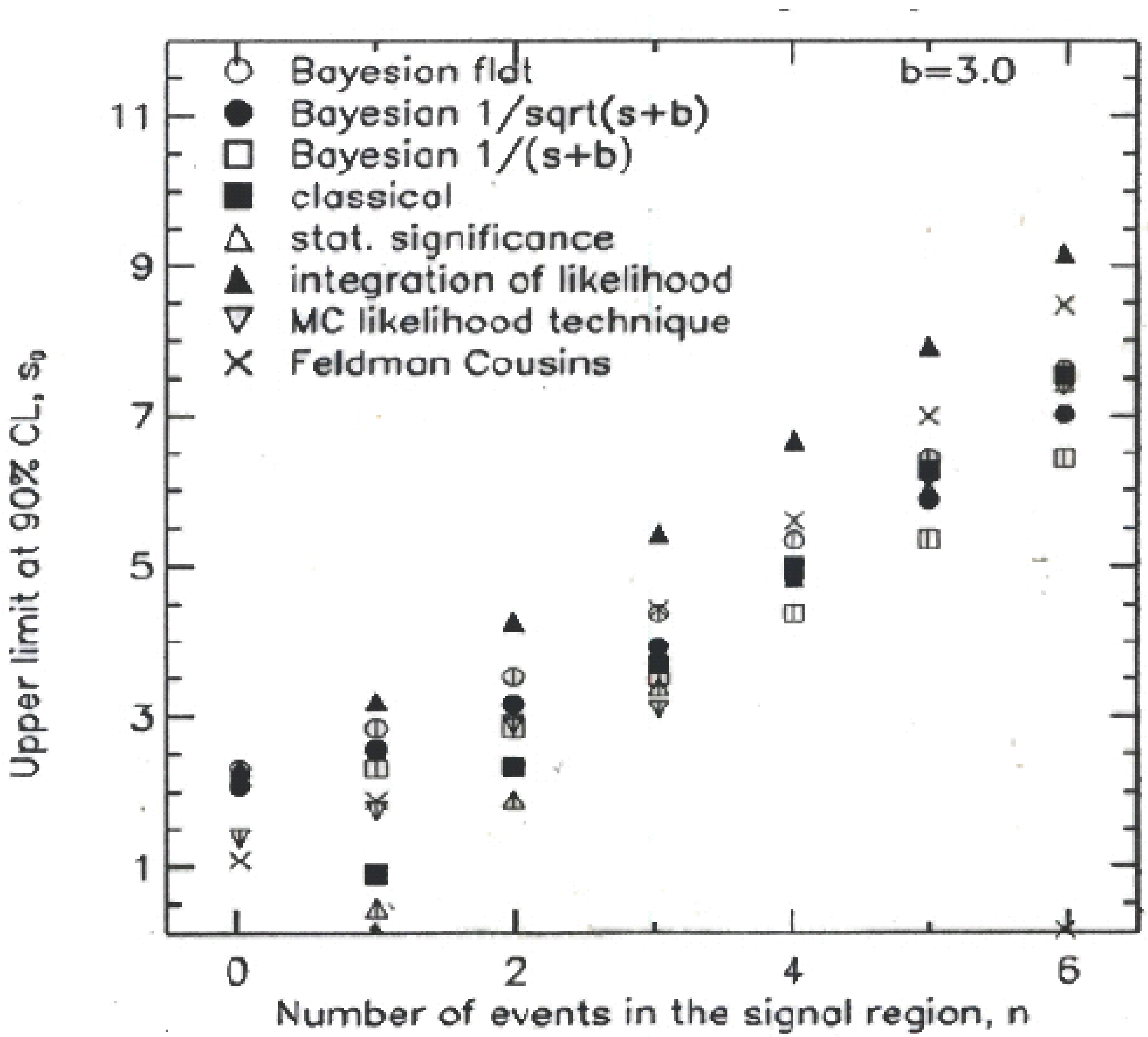


FIG. 3. Standard confidence belt for 90% C.L. central confidence intervals for the mean of a Gaussian, in units of the rms deviation.

$X_{obs} = 1 \qquad$ Upper limit = 2.3

Ilya Narsky, FNAL CLW 2000

(No systematics)



Upper Limit is very sensitive to method when n < b

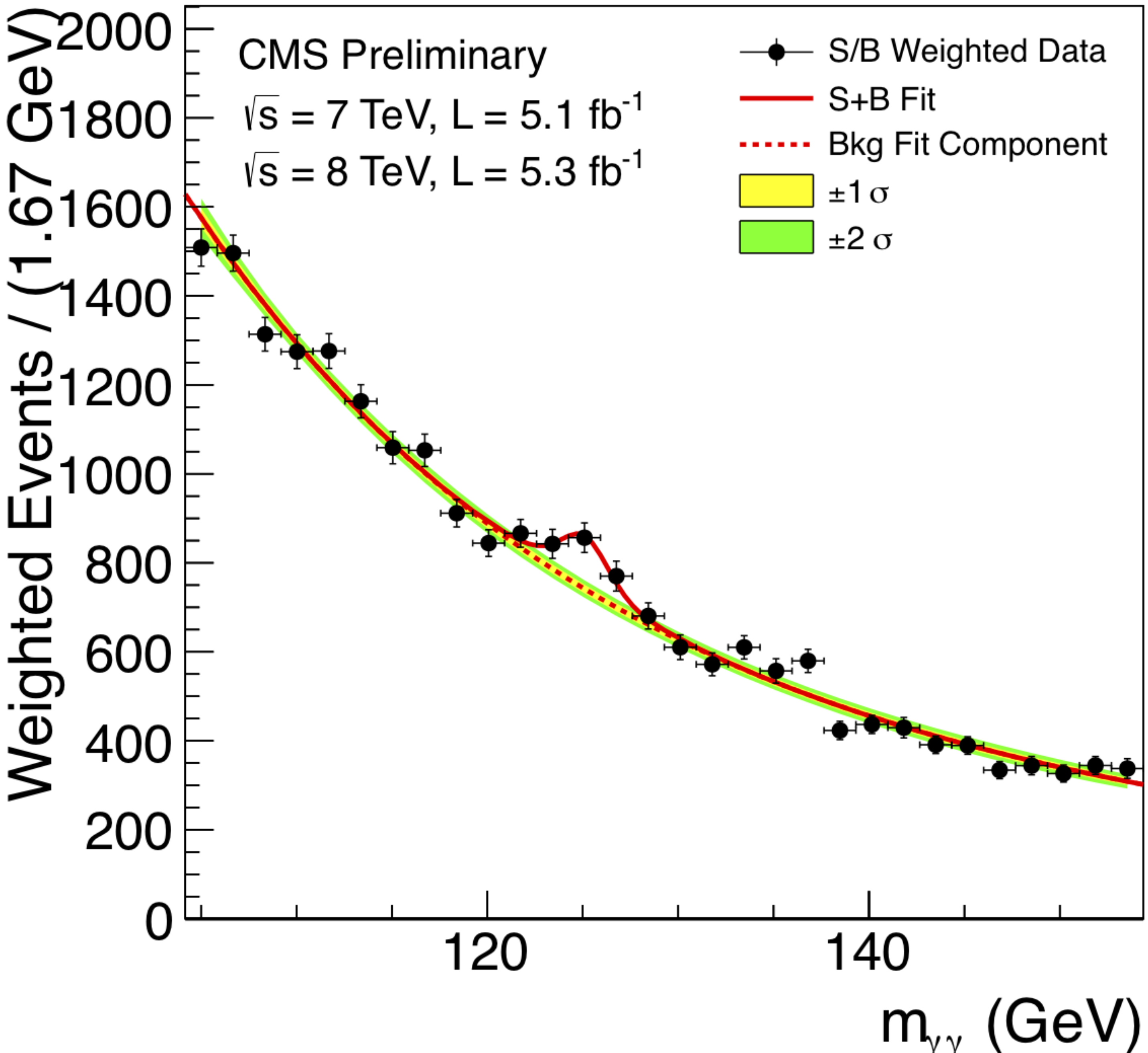
Including systematics

Bayes: Uses priors to model uncertainties

Cousins-Highland: Bayesian systematics for frequentist ULs

Profile Likelihood: $\mathcal{L}_{\text{profile}}(\varphi) = \mathcal{L}(\varphi, \nu_{\text{best}}(\mu))$

Dauncey, Kenzie, Wardle & Davies (IC, CMS):
"Handling uncertainties in background shapes: the **discrete profiling method**", JINST 10 (2015) no.04, 04015
Has been used in CMS analysis of $H \rightarrow \gamma\gamma$



CMS Preliminary
√s = 7 TeV, L = 5.1 fb-1
√s = 8 TeV, L = 5.3 fb-1
S/B Weighted Data
S+B Fit
Bkg Fit Component
±1 σ
±2 σ
Weighted Events / (1.67 GeV)
0
200
400
600
800
1000
1200
1400
1600
1800
2000
120
140
mγγ (GeV)

Sensitivity

Expected Upper Limit

Expected = Median, Mean, Asimov

(Can also give 68% and/or 90% bands)

Useful for:

a) Planning stage of experiment
b) Optimise search procedure
c) See if observed limit is plausible
d) Compare different experiments

$\mu_l \le \mu \le \mu_u$ at 90% confidence

Frequentist

μ_l and μ_u known, but random

μ unknown, but fixed

Probability statement about μ_l and μ_u

Bayesian

μ_l and μ_u known, and fixed

μ unknown, and random

Probability/credible statement about μ

Why 5σ for Discovery?

Statisticians ridicule our belief in extreme tails (esp. for systematics)

Our reasons:

1) Past history (Many 3σ and 4σ effects have gone away)

2) LEE = Look Elsewhere Effect

3) Worries about underestimated systematics

4) Subconscious Bayes calculation

$$\frac{p(H_1|x)}{p(H_0|x)} = \frac{p(x|H_1)}{p(x|H_0)} * \frac{\pi(H_1)}{\pi(H_0)}$$

Posterior prob   Likelihood ratio   Priors

"Extraordinary claims require extraordinary evidence"

N.B. Points 2), 3) and 4) are experiment-dependent

Alternative suggestion:

L.L. "Discovering the significance of 5σ"   http://arxiv.org/abs/1310.1284

How many σ's for discovery?

SEARCH	SURPRISE	IMPACT	LEE	SYSTEMATICS	No. σ
Higgs search	Medium	Very high	M	Medium	5
Single top	No	Low	No	No	3
SUSY	Yes	Very high	Very large	Yes	7
B_s oscillations	Medium/Low	Medium	Δm	No	4
Neutrino osc	Medium	High	$\sin^2 2\vartheta$, Δm^2	No	4
$B_s \rightarrow \mu\mu$	No	Low/Medium	No	Medium	3
Pentaquark	Yes	High/V. high	M, decay mode	Medium	7
$(g-2)_\mu$ anom	Yes	High	No	Yes	4
H spin ≠ 0	Yes	High	No	Medium	5
4^{th} gen q, l, ν	Yes	High	M, mode	No	6
Dark energy	Yes	Very high	Strength	Yes	5
Grav Waves	No	High	Enormous	Yes	8

Suggestions to provoke discussion, rather than `delivered on Mt. Sinai'

Bob Cousins: "2 independent expts each with 3.5σ better than one expt with 5σ"

Resources

Books by Particle Physicists

Barlow, Benkhe, Cowan, James, Lista, Lyons, Roe,…..
PDG: Sections on Probability, Statistics and Monte Carlo simulation.

PHYSTAT meetings

CERN and FNAL 2000 for U.L.
PhyStat-nu, Japan and FNAL 2016
PhyStat-DM in 2018?

Statistics Committees

Collider expts: BaBar, CDF, ATLAS, CMS
Maybe for neutrino expts
Perhaps for DM

RooStats

e.g. Lyons + Moneta at CERN (2016) and at IPMU (2017)
"Too easy to use"

Conclusions

Do your homework:

Before re-inventing the wheel, try to see if Statisticians have already found a solution to your statistics analysis problem.

Don’t use your own square wheel if a circular one already exists.

Try to achieve consensus

Good luck.

Move from U.L. → Discovery and Measurements