Frequentistic approaches in physics: Fisher, Neyman-Pearson and beyond

Alessandro Palma

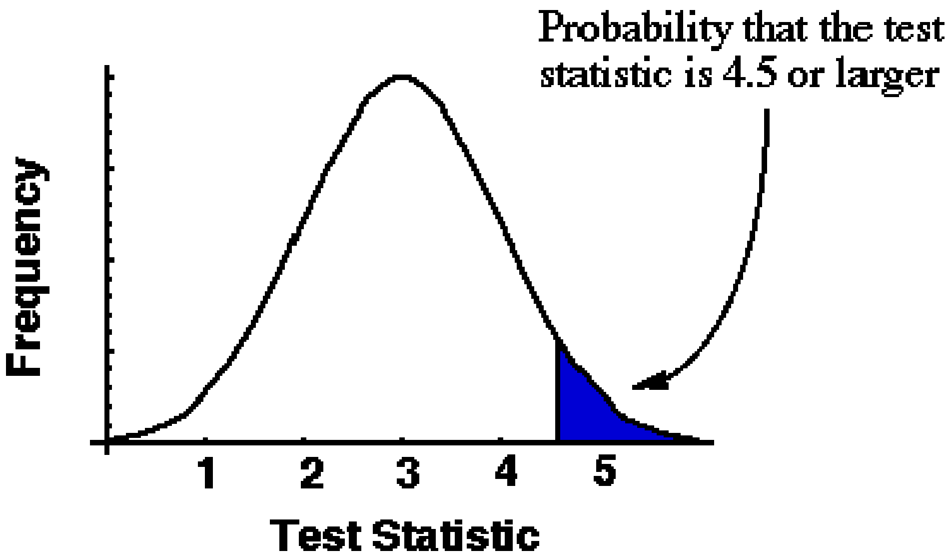
Dottorato in Fisica XXII ciclo
Corso di Probabilità e Incertezza di Misura
Prof. D'Agostini

Outlook

- Fisher's statistics and p-values
- Neyman-Pearson's statistics
- Fisher/NP: Differences and critics
- Usual misunderstandings
- "Taking the best from both": trying to merge the 2 approaches

Fisher's statistics (1)

- We have a single hypothesis H0
- Consider a test statistics T distributed with a known pdf f(T|H0) under H0
- Compute T on data→get the value Tobs
- Define p-value as Prob(T≥Tobs|H0)

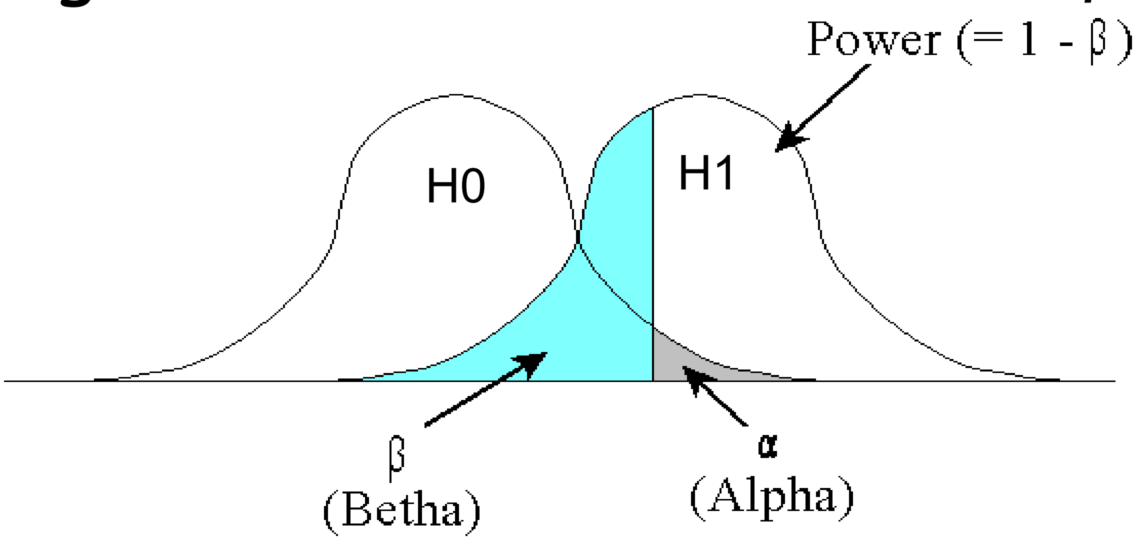


Fisher's statistics (2)

- According to Fisher, p-value is evidential for/against a theory
- if very small, "either something very rare has happened, or H0 is false"
- What happens typically?
  - one gets a p-value of 1% on data
  - the theory is rejected saying that "the probability that H0 is true is 1%"

Neyman-Pearson's statistics (1)

- We deal with (at least) TWO different hypotheses i.e. H0 and H1
- Test statistics T has different pdf f(T|H0) and f(T|H1)
- 2 integral tails are defined, α and β:



Neyman-Pearson's statistics (2)

- α is related to the probability of observing a H0 effect, given the data
- the value of α is chosen *a priori*, i.e. α=5%
- a value Tcut is chosen such that Prob(T≥Tcut)=α
- if T(data)>Tcut, one says that H0 is rejected at the 1-α (i.e. 95%) confidence level

Differences between Fisher and NP

- Number of Hypotheses
  - Fisher deals with only 1 hypotheses
  - NP needs at least 2
- Evidence against a theory:
  - for Fisher a p-value of 10^{-30} seems to reject much more strongly than a p-value of 10^{-2}
  - in NP, the very p-value does NOT matter, it only matters that $p < \alpha$ and the rejection is claimed always at the $1-\alpha$ level

Some critics to Fisher (1)

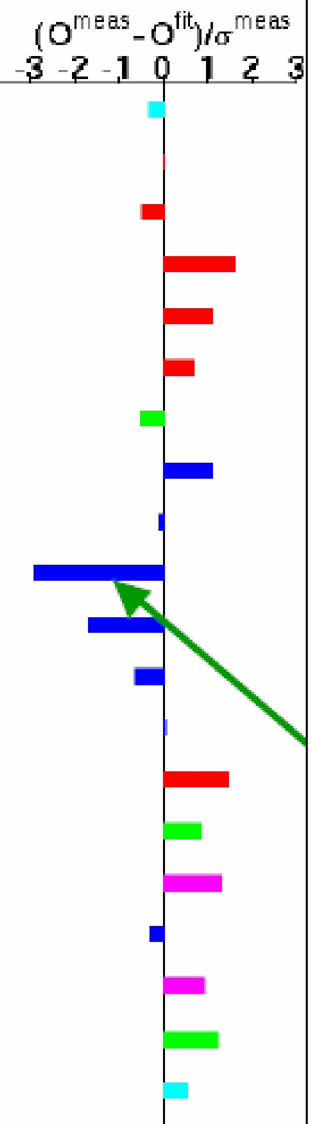
- p-value is Prob(T≥Tobs|H0): we are taking into account results which are very rare in H0, so...
  - "a hypothesis that may be true may be rejected because it has not predicted observable results thate have not occurred" (Jeffreys, 1961)
  - the p-value is strongly biased AGAINST the theory we are testing

Some critics to Fisher (2)

- A p-value of 1% does NOT state that the probability of the hypothesis H0 is 1%!
  - Probability of results ≠ probability of hypotheses (Bayes' theorem)
  - It can be seen using simulations (www.stat.duke.edu/~berger) that a p-value of 5% can result from data where H0 is true in 20-50% of cases!

Fisher in frontier physics: the LEP-SM pulls

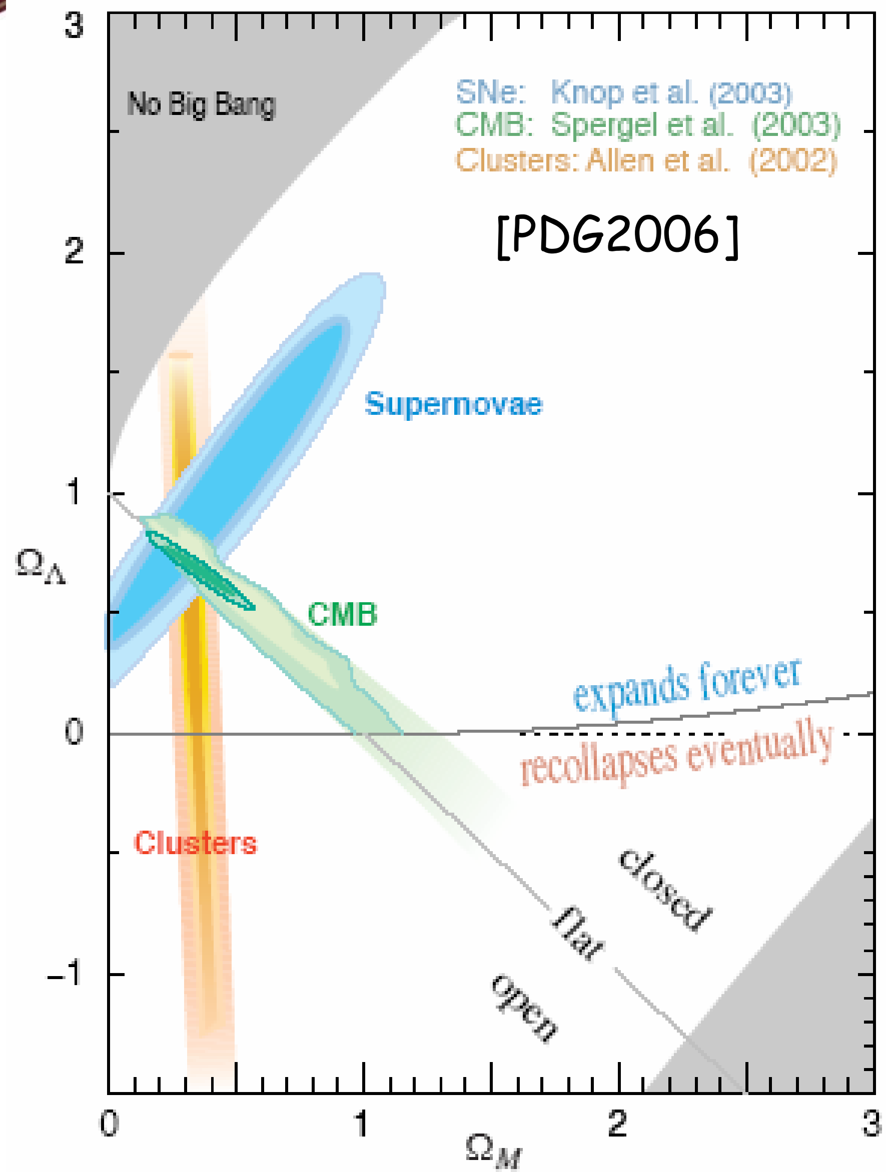
	Measurement	Pull
$\Delta\alpha^{(5)}_{had}(m_Z)$	0.02761 ± 0.00036	-.35
m_Z [GeV]	91.1875 ± 0.0021	.03
Γ_Z [GeV]	2.4952 ± 0.0023	-.48
σ^0_{had} [nb]	41.540 ± 0.037	1.60
R_l	20.767 ± 0.025	1.11
$A^{0,l}_{fb}$	0.01714 ± 0.00095	.69
$A_l(P_\tau)$	0.1465 ± 0.0033	-.54
R_b	0.21646 ± 0.00065	1.12
R_c	0.1719 ± 0.0031	-.12
$A^{0,b}_{fb}$	0.0990 ± 0.0017	-2.90
$A^{0,c}_{fb}$	0.0685 ± 0.0034	-1.71
A_b	0.922 ± 0.020	-.64
A_c	0.670 ± 0.026	.06
$A_l(SLD)$	0.1513 ± 0.0021	1.47
$\sin^2\theta^{lept}_{eff}(Q_{fb})$	0.2324 ± 0.0012	.86
$m_W^{(LEP)}$ [GeV]	80.450 ± 0.039	1.32
m_t [GeV]	174.3 ± 5.1	-.30
$m_W^{(TEV)}$ [GeV]	80.454 ± 0.060	.93
$\sin^2\theta_W(\nu N)$	0.2255 ± 0.0021	1.22
$Q_W(Cs)$	-72.50 ± 0.70	.56



- "Would Fisher accept SM"?
- Fisher would not be satisfied, low p-values from tensions in EW data
- SM is nonetheless "the" below-TeV theory of particle physics

pull is > 1 (2) in 7 (1) cases

Fisher in frontier physics: the cosmological constant

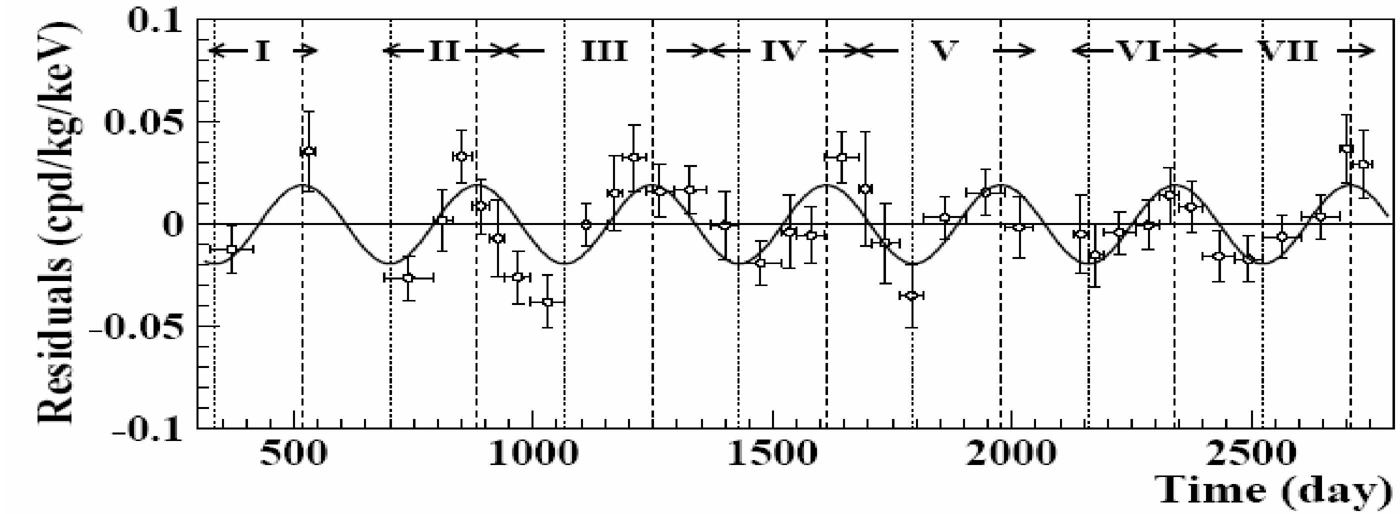


- All data point in the same ($\Lambda > 0$, $\Omega_\Lambda \approx 0.7$) direction
- Fisher is satisfied, but…
- …there is no physical idea for WHAT Λ is (i.e. no QFT vacuum energy)

DOUBT:
Is the cosmological constant a sensible issue?

Fisher in frontier physics: Dark Matter search

[Riv. N. Cim. 26 n.1 (2003) 1-73]



"The χ2 test on the (2–6) keV residual rate [...] disfavours the hypothesis of unmodulated behaviour giving a probability of 7 · 10–4 (χ2/d.o.f. = 71/37)."

- 6.3 σ signal, Fisher (and HEP community) would be greatly happy
- There is strong evidence for Dark Matter in cosmology: gravitation, BBN
- WHY "EVIDENCE"?! WHY NOT A WIDELY ACCLAIMED DISCOVERY?!
- Dark Matter particles not discovered yet → what about Higgs mechanism?
- Need confirmation by ther experiments? → what about W,Z in 1983?

Fisher in frontier physics: conclusions

- **SM**: HEP community hugely trusts SM at below-TeV scale, even with small p-values in precision data
- **Cosmological Constant**: physicists quite believe in Λ, even if there is NO idea about its physical meaning
- **Dark Matter**: physicists do NOT trust DAMA data, even if p-values strongly points towards DM discovery

…in frontier physics, Fisher's p-values and community beliefs seem NOT to be correlated!

Some critics to Neyman-Pearson

- NP statistics is nonevidential:
  - The level at which H0 is excluded does NOT depend on how rare the observed data is according to H0

Misunderstanding: p is NOT α !

- Frequent mistake on statistics textbooks
- What is usually done is the following "mix" of Fisher's and NP's approaches:
  - build NP distributions given H0 and H1, with α fixed (i.e. α = 5%)
  - compute p-value on data
  - if $p \leq \alpha$ exclude data at "p-value" confidence level

Take the best from NP and Fisher...

- Fisher's statistics has a number of bugs
  - it's a tail integral, gives no info on H0
- NP's statistics has no much sensitivity to data
  - rejection level is decided *a priori*
- Can we combine both, and get a sensible frequentistic approach?
  - evidential (based on data)
  - giving info on the "probability of hypotheses"

Proposal for unifying Fisher and NP (Jeffreys)

1. Take the case of 2 hypotheses H0 and H1 (NP)
2. Use ratio of likelihoods, NOT p-values ("debugged" Fisher) to address evidence from data:

$$B(x) = f(x|\theta_0)/f(x|\theta_1)$$

3. Reject H0 if B≤1, accept otherwise
4. Claim the following "objective" probabilities of the hypotheses:

$$\Pr(H_0|x) = \frac{B(x)}{1+B(x)} \qquad \Pr(H_1|x) = \frac{1}{1+B(x)}$$

Observations on Jeffreys' approach

- It is exactly a Bayesian approach with uniform priors f0(H0)=f0(H1)=1/2
- it is ON PRINCIPLE Bayesian since talks of probability of hypotheses
- "Objective" should be referred to the (arbitrary) choice of a uniform prior. But it's not a sensible choice that the scientist be forced to choose so!
- "Reject H0 if B≤1" is a pointless statement; it suffices to report the probability of hypotheses

Conclusions & (my) open issues

- Both Fisher's and NP's approaches are non-satisfactory among the frequentistic community
- Jeffreys' solution ends up in a Bayesian-like statement
- Physics community has beliefs which hugely contradict Fisher's approach (SM, cosmological const., DAMA)
- Frequentistics positions seem weak…

- General issue: do we really need to test at least 2 hypotheses?

References

- J. O. Berger, "Could Fisher, Jeffreys and Neyman have agreed on testing", Statistical Science vol. 18, n. 1, 1-32 (2003)
- R. Hubbard, M.J.Bayarri, "Confusion over measures of evidence (p's) versus errors (α's) in classical statistical testing", The American Statistician vol. 57, n.3 (2003)
- H. Jeffreys, "Fisher and inverse probability", International Statistical Review vol. 42, 1-3 (1974)
- [PDG2006]: W.-M. Yao et al., J. Phys. G **33**, 1 (2006)
- R. Bernabei et al., "Dark Matter search", Riv. N. Cim. 26 n.1 (2003) 1-73