Bayesian and Frequentist Methods in Particle Physics

Roger Huang

November 8, 2017

Outline

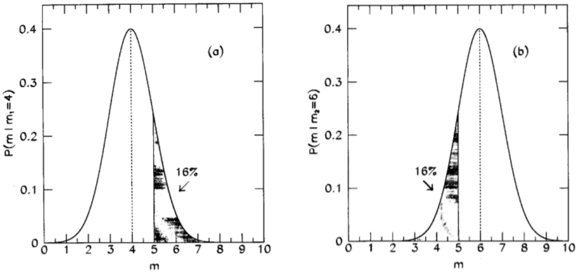
- Frequentist Interpretations and Constructions
- Bayesian Interpretations and Constructions
- Incorporating Physical Constraints
    - Historical Example: Neutrino Mass measurements
- Comparison of performances: Dark Photon Search at BaBar
- A Detailed Example using both methods: GERDA

Interpretations of Data

- What does it mean when an experiment reports a 68% confidence interval of 5 ± 1 MeV for some value m?
- The 68% number in the confidence interval is referred to as its *coverage*
    - *Undercoverage* is when a method actually covers the true value of m less often than it's supposed to
    - *Overcoverage* is when a method covers the true value of m more often than it says it does
    - Overcovering gives a *conservative* interval, which we're generally ok with
    - Technically the concept of coverage only refers to the frequentist interpretation
- In all constructions, we must accurately know $P(m_{data}|m_{true})$, and this what most experimental design and calibration goes into

Frequentist Interpretation

- Frequentist construction of 68% CI of 5 ± 1 MeV: 68% of an ensemble of similar experiments that construct an interval using this method will contain the true value of m in their constructed interval
  - Not a statement of degree of belief that m is in the given interval 4-6 MeV. The interval from any one experiment is technically meaningless to interpret on its own
  - "Ensemble of experiments" $\rightarrow$ actually happens in real life with many experiments with measurements of the same physical quantity



[1]

Frequentist Construction - Profile Likelihood

- Profile Likelihood method is used when our data depends on some parameters of interest $\boldsymbol{\pi}$, but also on nuisance parameters $\boldsymbol{\theta}$
- Given observed data **X** of size n, the likelihood function is

$$L(\boldsymbol{\pi}, \boldsymbol{\theta} | X) = \prod_{i=1}^{n} f(X_i | \boldsymbol{\pi}, \boldsymbol{\theta})$$

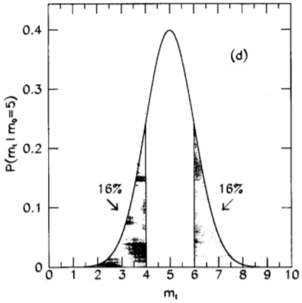
- Construct a likelihood ratio test statistic that is independent of the nuisance parameters:

$$\lambda(\boldsymbol{\pi_0} | X) = \frac{\sup[L(\boldsymbol{\pi_0}, \boldsymbol{\theta} | X); \boldsymbol{\theta}]}{\sup[L(\boldsymbol{\pi}, \boldsymbol{\theta} | X); \boldsymbol{\pi}, \boldsymbol{\theta}]}$$

- Minimize $-2 \log \lambda$ and treat as χ^2 statistic to extract confidence interval (Rolke Limit)

Bayesian Interpretation

- We model our beliefs about any physical value with some probability distribution
- Bayesian construction of 68% interval of 5 ± 1 MeV: this interval contains 68% of the pdf we have constructed for m using our data and some prior pdf for m
  - We are 68% certain that m falls in this region, having updated our beliefs in accordance with the data



[1]

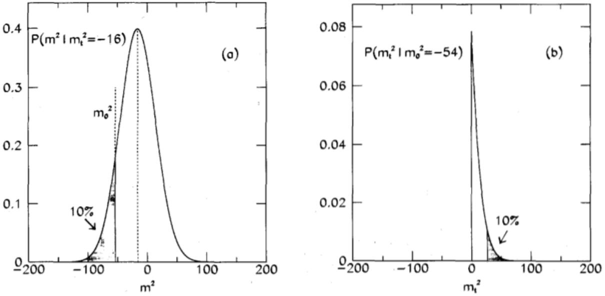
Bayesian Construction

- Must select some prior pdf for m, often chosen to be a uniform distribution
  - Other choices for an "uninformed" prior exist
- Posterior distribution obtained with Bayes' Theorem:

$$P(m_t|X) = \frac{L(X|m_t)P(m_t)}{\int L(X|m)P(m)dm}$$

Setting Limits - A Historical Example

- In 1994, the PDG weighted average of measurements of the electron neutrino mass was $m^2 = (-54 \pm 30)$ eV^2
- Naive frequentist construction (left) gives a 90% CL upper limit of $m^2 < -16$ eV^2 - clearly unphysical
- Bayesian construction (right) has intuitive way to incorporate physical constraints - set prior pdf to be 0 for $m^2 < 0$ and uniform for $m^2 > 0$
  - Problem: what variable to set the uniform prior in? $m, m^2, \ln m$, etc.



[1]

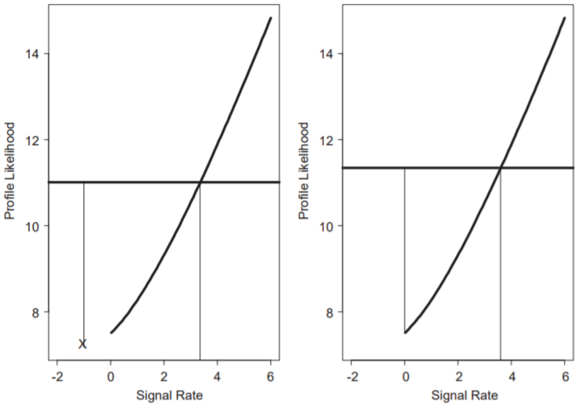
Adjusting Frequentist Limits

- How to incorporate physical constraints into frequentist limits?
- Unbounded likelihood: compute limit ignoring physical constraints. If the limit is unphysical, increase signal events by 1 until it becomes physical
- Bounded likelihood: compute limit using increase from maximum likelihood in physical region only, or by shifting the whole curve so that the maximum likelihood is at the edge of the physical region
- No proper justification for these adjustments, but we can accept them because they can only make the limit worse than the "proper" frequentist limit

Adjusting Frequentist Limits

Example: arbitrary rare event search

Unbounded likelihood   Bounded likelihood



[2]

Performance of Bayesian vs Frequentist Limits

- Differences resulting from choice of statistical method are only likely to be significant for rare event searches or in experiments that are setting limits due to lack of discovery
- For historical reasons, frequentist methods are considered "classical", and so serve as a point of comparison for other methods
- Bayesian methods tend to set more conservative limits in cases of negative fluctuations, and stricter limits in case of positive fluctuation

Performance of Bayesian vs Frequentist Limits - Dark Photon Searches at BaBar

- BaBar is an asymmetric e^+e^- collider at SLAC, permitting the search for dark photons by
$$e^+e^- \to \gamma A'$$
- Dark photon A' invisibly decays by $A' \to \chi\bar{\chi}$, making the signature a single photon with large missing energy/momentum
- Primary backgrounds are $e^+e^- \to \gamma\gamma$ with an escaping photon and $e^+e^- \to e^+e^-\gamma$ where the electron and positron escape the detector
- Signal fit with a Crystal Ball pdf centered on the expected value of $m^2_{A'}$

Performance of Bayesian vs Frequentist Limits - Dark Photon Searches at BaBar

- Results rule out dark-photon coupling as explanation for $(g-2)_\mu$ anomaly

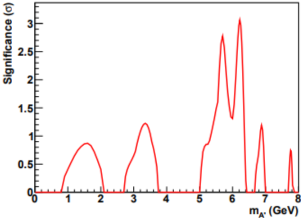


FIG. 2: Signal significance $\mathcal{S}$ as a function of the mass $m_{A'}$.

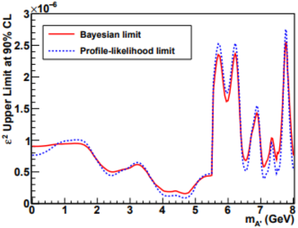
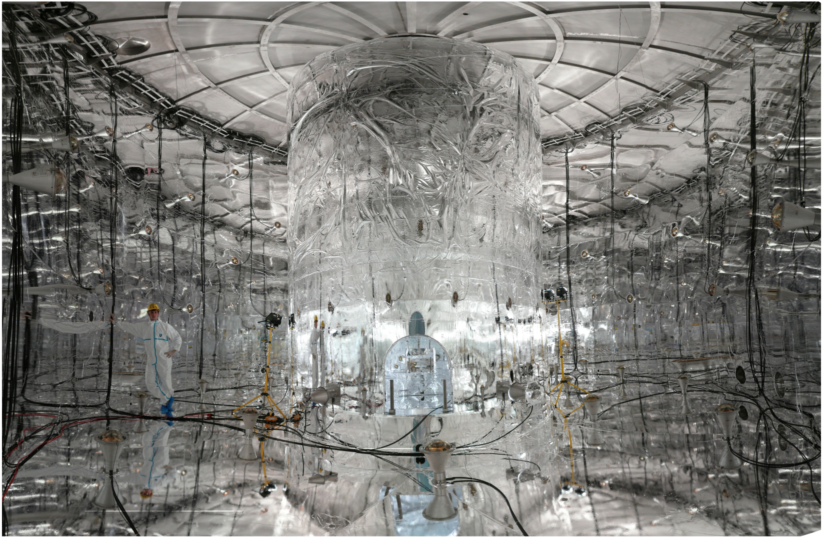


FIG. 4: Upper limits at 90% CL on A' mixing strength squared ε^2 as a function of $m_{A'}$. Shown are the Bayesian limit computed with a uniform prior for $\varepsilon^2 > 0$ (solid red line) and the profile-likelihood limit (blue dashed line).

[3]

A Detailed Example: GERDA



The GERDA Experiment

- GERDA (GERmanium Detector Array) is located at underground laboratory LNGS in Italy and is searching for $0\nu\beta\beta$ decay of ^{76}Ge
- Enriched Germanium detectors (87% ^{76}Ge) submerged in a 64 m^3 liquid argon cryostat, which is submerged in a 590 m^3 water tank
- 66 PMTs located in the water tank and 16 low-background cryogenic PMTs in the cryostat allow active rejection of background events
- Very low background of 0.001 cts/(keV*kg*yr), and very high energy resolution of 2.6-4.4 keV at 2.6 MeV
- Current total exposure of 34.4 kg*yr

The GERDA Experiment

GERDA Results

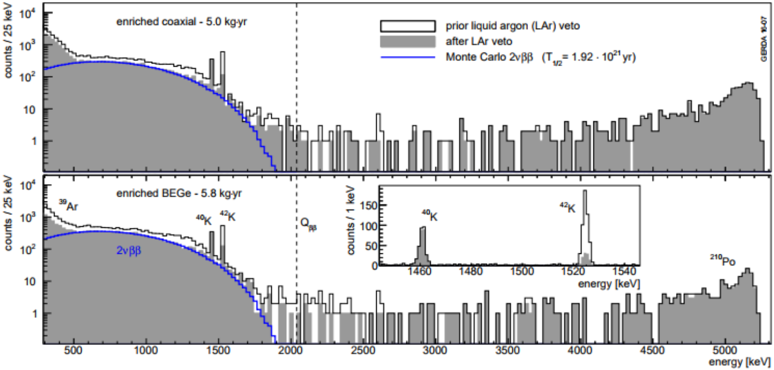


FIG. 2. Energy spectra of Phase II data sets before (open histogram) and after argon veto cut (filled histogram). The blue lines are the expected $2\nu\beta\beta$ spectra for our recent half-life measurement. The inset shows the BEGe spectrum in the energy region around the two potassium lines. Various background contributions are labeled in the bottom panel.

[4]

GERDA Results

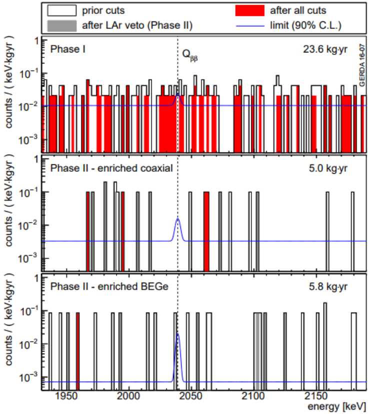


FIG. 4. Combined Phase I data (top), Phase II coaxial (middle) and BEGe detector spectra (bottom) in the analysis window. The binning is 2 keV. The exposures are given in the panels. The red histogram is the final spectrum, the filled grey one without pulse shape discrimination and the open one in addition without argon veto cut. The blue line is the fitted spectrum together with a hypothetical signal corresponding to the 90 % C.L. limit of $T^{0\nu}_{1/2} > 5.3 \cdot 10^{25}$ yr.

[4]

- Performed both Frequentist and Bayesian analysis using an unbinned extended likelihood function
- Fit the region of interest for each dataset with a flat background and a Gaussian centered at $Q_{\beta\beta}$ with width corresponding to the calibrated energy resolution
- Frequentist 90% CL limit: $T^{0\nu}_{1/2} > 5.3 \cdot 10^{25}$ yr
- Bayesian 90% CI limit: $T^{0\nu}_{1/2} > 3.5 \cdot 10^{25}$ yr

GERDA Analysis

- Likelihood function $L(D|S, BI, \theta)$ obtained by fitting the flat background + Gaussian to the data
  - D: Dataset of event energies
  - S: Signal rate $= 1/T_{1/2}^{0\nu}$
  - BI: Background index. Total number of background events in ROI is $\mu^B =$ (exposure)*(BI)*(width ΔE of ROI)
  - θ: Nuisance parameters with systematic uncertainties. Includes global signal efficiency ϵ, energy resolution σ, and a possible systematic energy offset δ
- Total likelihood constructed as Poisson-weighted product of each individual datasets' likelihood functions

$$L(D|S, BI, \theta) = \prod_i \left[\frac{e^{-(\mu_i^S + \mu_i^B)} \cdot (\mu_i^S + \mu_i^B)^{N_i^{obs}}}{N_i^{obs}!} \cdot L_i(D_i|S, BI_i, \theta_i)\right]$$

GERDA Analysis - Frequentist Construction

- Test statistic based on profile likelihood. $\hat{\hat{B}}I, \hat{\hat{\theta}}$ maximize L for a given S. $\hat{S}, \hat{B}I, \hat{\theta}$ give the absolute maximum for L.

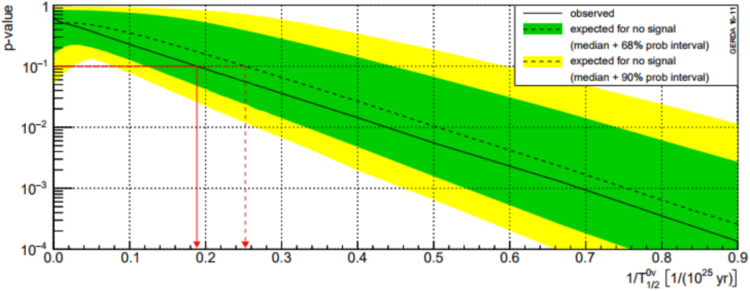
$$t_S = -2 \ln \lambda(S) = -2 \ln \frac{L(S, \hat{\hat{B}}I, \hat{\hat{\theta}})}{L(\hat{S}, \hat{B}I, \hat{\theta})}$$

- A discrete set of possible values S_j are considered for S. For each S_j, possible realizations of the experiment are generated with Monte Carlo methods. For each realization, the test statistic t_{S_j} is calculated
- The p-value of the data for each S_j is given by the below, where t_{obs} is the test statistic of the actual data for S_j

$$p_{S_j} = \int_{t_{obs}}^{\infty} f(t_S|S_j) d(t_{S_j})$$

- 90% CL interval given by all S_j values with $p_{S_j} > 0.1$ (values where $<$ 10% of simulated experiments had test statistic more unlikely than the one in data)

GERDA Analysis - Frequentist Construction



Extended Data Fig. 5. p-value for the hypothesis test as a function of the inverse half-life $1/T^{0\nu}_{1/2}$ for the data (full black line) and the median sensitivity (dashed black line). The 68 (90) % interval is given by the green (yellow) range. The red arrows indicate the results at 90 % confidence level: the limit for $T^{0\nu}_{1/2}(^{76}\mathrm{Ge}) > 5.3 \cdot 10^{25}$ yr (full red line), the median sensitivity for $T^{0\nu}_{1/2}(^{76}\mathrm{Ge}) > 4.0 \cdot 10^{25}$ yr (dashed red line).

[4]

GERDA Analysis - Bayesian Construction

- Posterior pdf given by Bayes' theorem over the combined datasets:

$$P(S, BI|D, \theta) \propto L(D|S, BI, \theta) P(S) \prod_i P(BI_i)$$

- Priors are flat pdf between 0 to 0.1 cts/(keV*kg*yr) for backgrounds $P(BI_i)$, and flat from 0 to 10^{-24}/yr for signal $P(S)$
- Nuisance parameters θ are then averaged into the final pdf with Gaussian distributions:

$$< P(S|D) >= \int P(S|D, \theta) \prod_i g(\theta_i) d\theta_i$$

- Other possible choices for flat priors: Majorana neutrino mass $(1/\sqrt{S})$, scale invariance in counting rate $(\log S)$

Summary

- Frequentist methods
  - Pros: "Objective" by not requiring usage of prior beliefs, so that beliefs can be incorporated later in decision theory
  - Cons: Results are prone to misinterpretation due to the strangeness of the correct way to look at them. Methods can rely on ad hoc adjustments
- Bayesian methods
  - Pros: Intuitive results that answer the questions we want to ask. Straightforward and well-justified way to incorporate nuisance parameters and physical constraints
  - Cons: Results depend on choice of prior pdf's
- Effects from experimental design and operation will always far outweigh any differences from choice of statistical method
- Understanding methods still important to know how to make sense of an experiment's results and compare against other experiments

References

(1) R. D. Cousins: Why isn't every physicist a Bayesian? http://www.astro.princeton.edu/ strauss/AST303/bayesian_paper.pdf

(2) W. Rolke, A. Lopez, J. Conrad: Limits and Confidence Intervals in the presence of nuisance parameters https://doi.org/10.1016/j.nima.2005.05.068

(3) Search for invisible decays of a dark photon produced in e^+e^- collisions at BaBar arxiv:1702.03327

(4) Background free search for neutrinoless double beta decay with Gerda Phase II arxiv:1703.00570

(5) S. Biller, S. Oser: Another look at confidence intervals: Proposal for a more relevant and transparent approach https://doi.org/10.1016/j.nima.2014.11.081