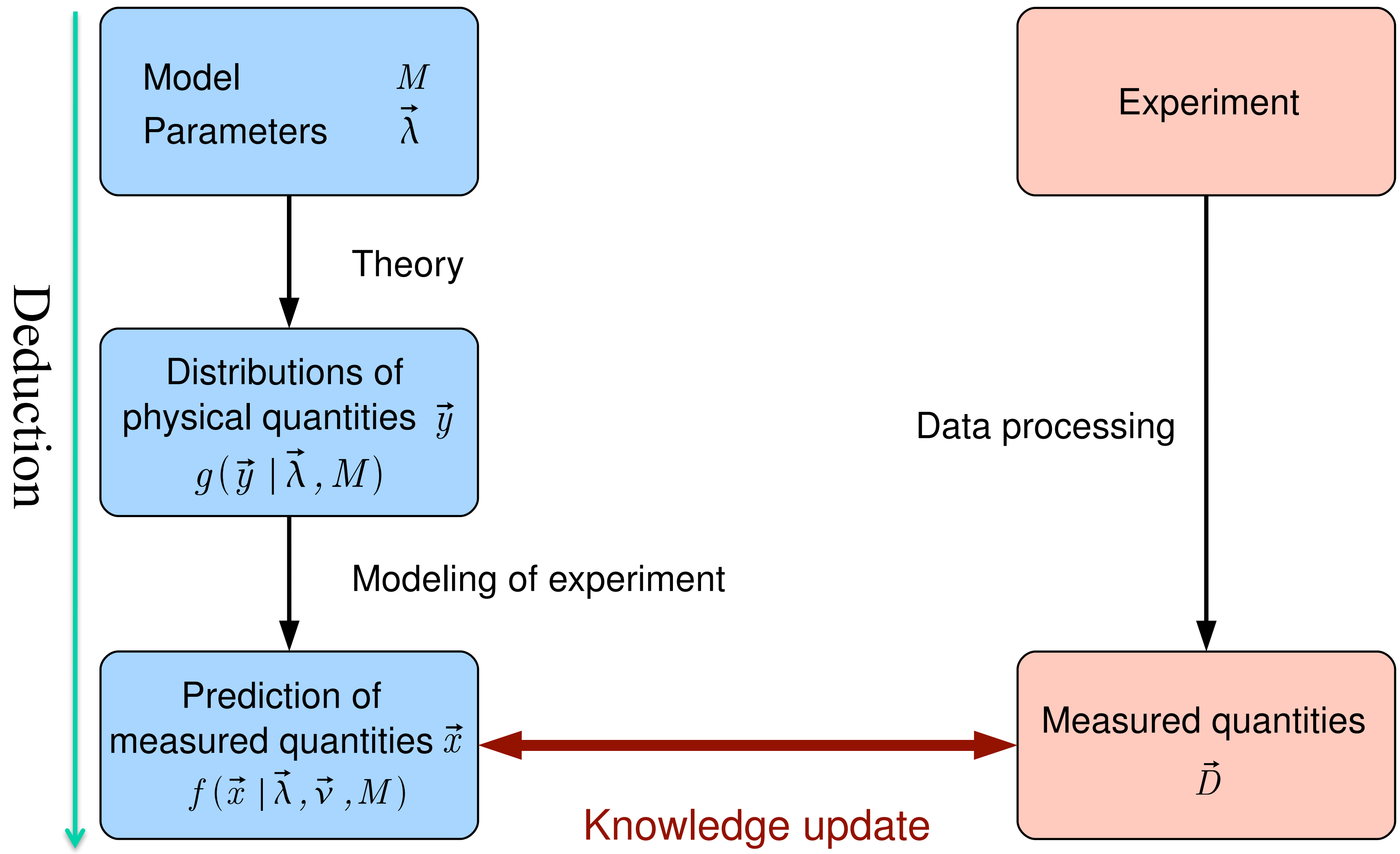
Signal Discovery in Sparse Spectra

F. Beaujean[1], **A. Caldwell[1],** D. Kollar[2], K. Kröninger[3], S. Pashapour[3]

1-Max Planck Institute for Physics
2-CERN
3-II. Physikalisches Institut, Universität Göttingen

1. Logical framework for data analysis
2. Discovery in sparse spectra
3. Discussion

How we learn


Model M
Parameters $\vec{\lambda}$
Theory
Distributions of physical quantities $\vec{y}$
$g(\vec{y} \mid \vec{\lambda}, M)$
Modeling of experiment
Prediction of measured quantities $\vec{x}$
$f(\vec{x} \mid \vec{\lambda}, \vec{\nu}, M)$
Experiment
Data processing
Measured quantities
$\vec{D}$
Knowledge update
Deduction

Logical Basis

Model building and making predictions from models follows deductive reasoning:

Given A➔B (with some frequency)
Given B➔C (with some frequency)
Then, given A you can conclude that C is possible with some probability (frequency)

etc.

Everything is clear, we can make frequency distributions of possible outcomes within the model, etc. This is math, so it is correct …

Logical Basis

However, in physics what we want to know is the validity of the model given the data. i.e., logic of the form:

Given A➔B (with some frequency)
Given B➔C (with some frequency)
Measure C, what can we say about A ? Well, can say A is a possibly correct model. What else ? Need to know about other models

maybe A_1➔C, A_2➔C, …

We now need inductive logic to decide how much we want to believe each possible model. We can never say anything absolutely conclusive about A unless we can guarantee a complete set of alternatives A_i and only one of them can give outcome C. This does not happen in science, so we can never say we found the true model.

Logical basis

Purpose of science is to increase knowledge, where

Knowledge = **justified ~~true~~ belief**

Justification comes from the data.

Start with some knowledge or maybe plain belief

Data analysis gives updated knowledge. Need Bayes' Theorem to make coherent statements on what we believe.

nb: if don't use Bayes' theorem, belief update becomes maximally subjective.

Example: Double Beta Decay

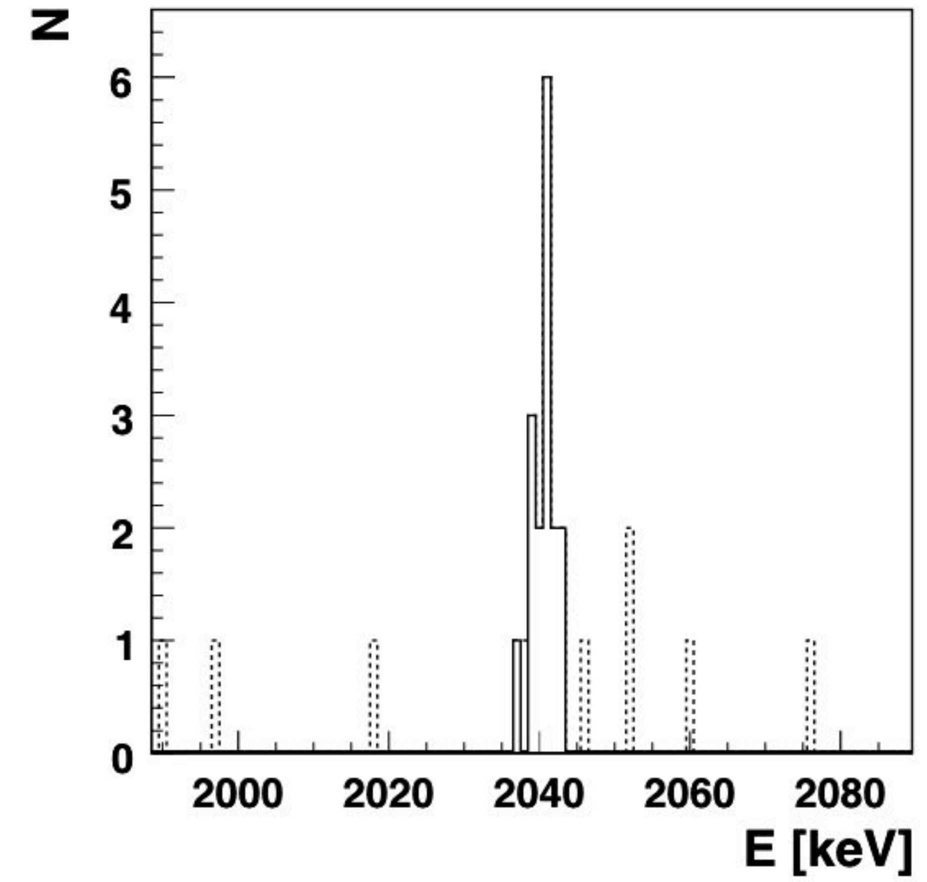
One of the outstanding questions in Particle Physics is whether the neutrino is its own antiparticle (so-called Majorana particle).

The only practical way which has been found to search for the Majorana nature of neutrinos (particle same as antiparticle) is double beta decay (because of the light mass of neutrinos, helicity flip is very unlikely unless the neutrinos have very low energy).

For us, what is interesting is that we are looking for a peak at a well-defined energy in a sparse spectrum.

A. Caldwell, K. Kröninger, Phys. Rev. D 74 (2006) 092003

Discovery or not ?



Analyze energy spectrum and decide if there is evidence for a signal.
Counting experiment – Poisson statistics.

Sparse Spectra

Define the proposition:

H = The observed spectrum is due to background only

If p(H|spectrum)<cut, can claim ‘evidence’ for something beyond background. If we assume that **what is not background is signal**, then we claim evidence for the signal.
E.g.:

p(H|spectrum)<0.01, ‘evidence’ (better >99% belief in `new physics’)

p(H|spectrum)<0.0001, ‘discovery’ (better >99.99% belief in `new physics’) (very stringent, DoB contains our belief in the new physics)

Note: intended to be the real ‘degree-of-belief’. **No fudging allowed afterwards – otherwise it implies you did not really believe your prior.**

Sparse Spectra

What we know how to calculate:
p(spectrum|H) - the probability to observe the spectrum given H
(We assume Poisson statistics are valid)

How do we go from p(spectrum|H) to p(H|spectrum) ?

Certainly p(A|B) ≠ p(B|A)
(e.g., 1% probability of signal assuming Standard Model does not mean Standard Model model ruled out with 99% certainty)

Need to use Bayes' theorem to reach a conclusion

$$p(H|spectrum) = \frac{p(spectrum|H)p_0(H)}{p(spectrum)}$$

Double Beta Decay Example

$p_0(H)$ is the prior belief in H (before we do the experiment). It is a critical part of the Bayesian analysis. Our posterior belief in the truthfulness of H always depends on prior beliefs. E.g.,

The existing limits are $T_{1/2}>4\ 10^{25}$ yr; a positive claim for a signal exists at the level $T_{1/2}=1.2\ 10^{25}$ yr; my favorite theorist believes strongly that neutrinos are Majorana particles, but he wont tell me the neutrino mass; the theorist at a neighboring university says that he believes strongly in Leptogenesis, and in that context the neutrino is a Majorana particle but it must be very light, such that neutrinoless double beta decay is unobservable,...

What is p(spectrum) ? Expand (law of total probability)

$$p(spectrum) = p(spectrum \mid H)p(H) + p(spectrum \mid \overline{H})p(\overline{H})$$

DBD example

We need also the probability of the negation of H. In our case, we assume knowledge concerning the background, so

$\overline{H}$ = **The spectrum is due to background + signal (neutrinoless double beta decay).**

I.e., we assume backgrounds are known up to normalization and some smoothly varying shape, and the only possibility other than known background is signal from neutrinoless double beta decay.

$$p(H \mid spectrum) + p(\overline{H} \mid spectrum) = 1$$

so

$$p(H \mid spectrum) = \frac{p(spectrum \mid H)\, p_0(H)}{p(spectrum \mid H)\, p_0(H) + p(spectrum \mid \overline{H})\, p_0(\overline{H})}$$

$$p(\overline{H} \mid spectrum) = \frac{p(spectrum \mid \overline{H})\, p_0(\overline{H})}{p(spectrum \mid H)\, p_0(H) + p(spectrum \mid \overline{H})\, p_0(\overline{H})}$$

DBD example

Now we know how to perform all calculations:

$$p(spectrum \mid H) = \int p(spectrum \mid B)\, p_0(B)\, dB$$

$$p(spectrum \mid \overline{H}) = \int p(spectrum \mid S,B)\, p_0(S)\, p_0(B)\, dB$$

Where B is the expected number of background events and S is the expected number of signal events. These quantities come with their own priors.

$n_i =$ observed number of events in bin i

$\lambda_i =$ expected number of events in bin i

$$\lambda_i = S \int_{\Delta E_i} f_S(E)\, dE + B \int_{\Delta E_i} f_B(E)\, dE$$

Where f_S and f_B are the normalized signal and background probability densities as functions of energy.

DBD example

then
$$p(spectrum \mid B) = \prod_{i=1}^{N} \frac{\lambda_i(0,B)^{n_i}}{n_i!} e^{-\lambda_i(0,B)}$$

$$p(spectrum \mid S,B) = \prod_{i=1}^{N} \frac{\lambda_i(S,B)^{n_i}}{n_i!} e^{-\lambda_i(S,B)}$$

To determine parameter values or set limits, we need

$$p(S,B \mid spectrum) = \frac{p(spectrum \mid S,B)\, p_0(S)\, p_0(B)}{\int p(spectrum \mid S,B)\, p_0(S)\, p_0(B)\, dS\, dB}$$

and then marginalize

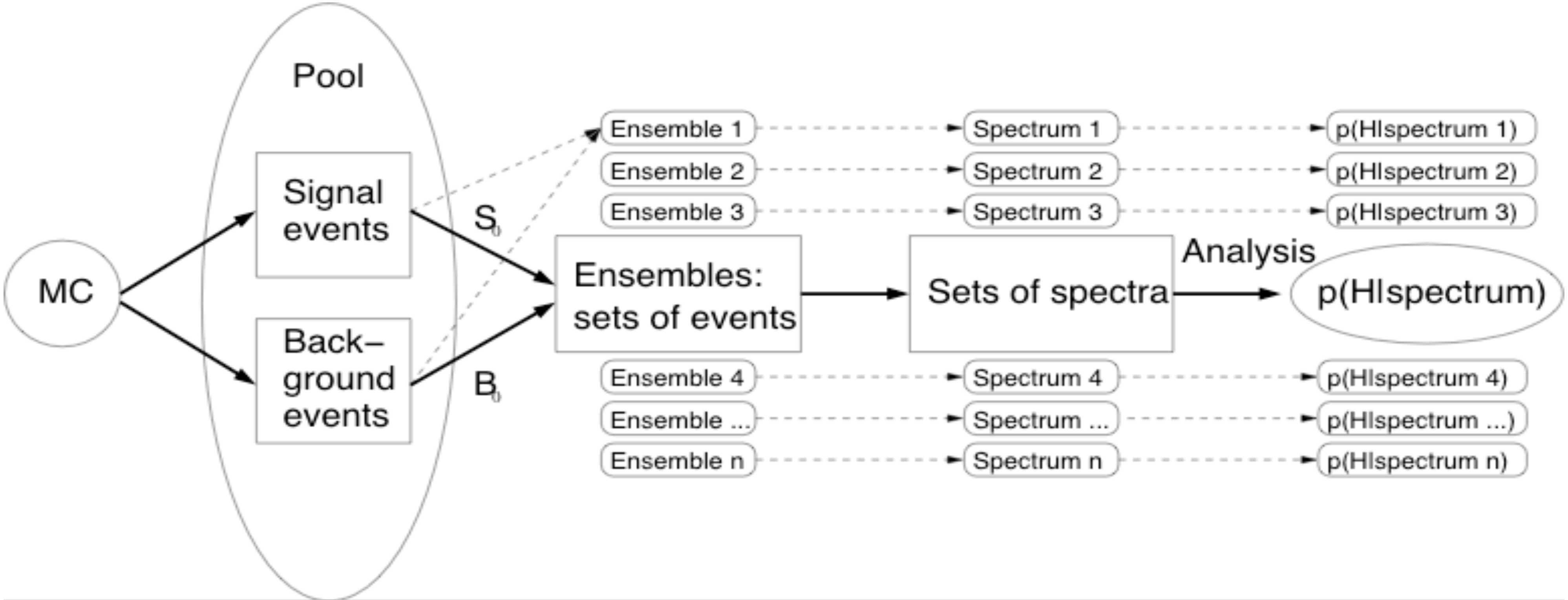
$$p(S \mid spectrum) = \int p(S,B \mid spectrum)\, dB$$

e.g., 90% probability upper limit, S_{90} from solving

$$\int_0^{S_{90}} p(S \mid spectrum)\, dS = 0.90$$

DBD example

So we know how to calculate probabilities given an experimental outcome. What do we do to check the sensitivity of the experiment ? We generate ensembles of possible experimental results, which will depend on particular choices of background and signal, B_0 and S_0. Then we can make distributions of the probabilities which could result under these conditions.



GERDA example

Assumptions for GERDA:

$$p_0(H) = p_0(\overline{H}) = 1/2$$

$$p_0(S) = \frac{1}{S_{\max}} \quad 0 \le S \le S_{\max} \quad p_0(S) = 0 \text{ otherwise}$$

$$p_0(B) = \frac{e^{-\frac{(B-\mu_B)^2}{2\sigma_B^2}}}{\int_0^\infty e^{-\frac{(B-\mu_B)^2}{2\sigma_B^2}} dB} \quad B \ge 0; \quad p_0(B) = 0 \; B < 0$$

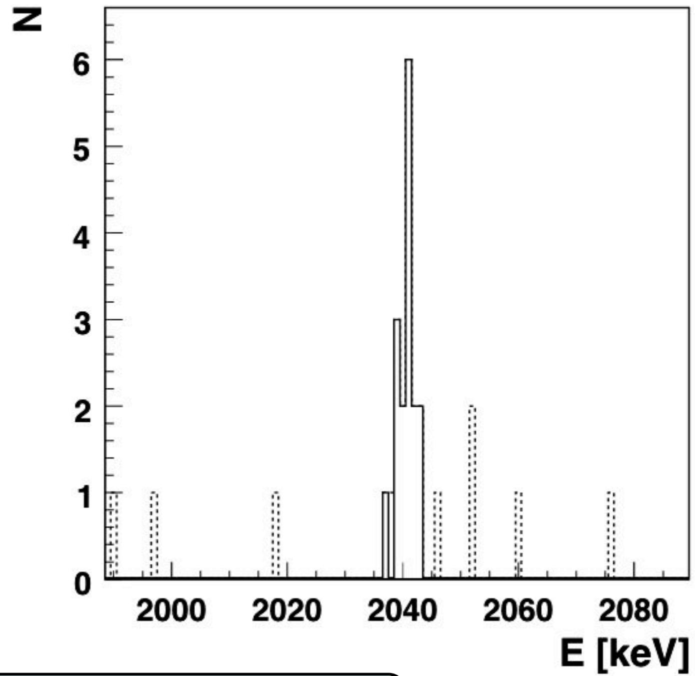
S_{max} was calculated assuming $T_{1/2}$=0.5 10^{25} yr

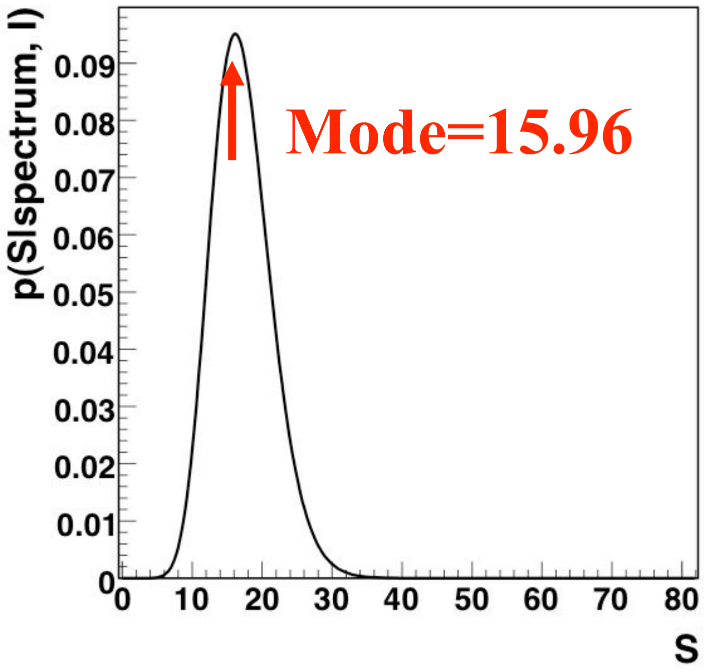
$\mu_B=B_0$, $\sigma_B=B_0/2$

100 keV window analyzed. B_0 total background in this window.

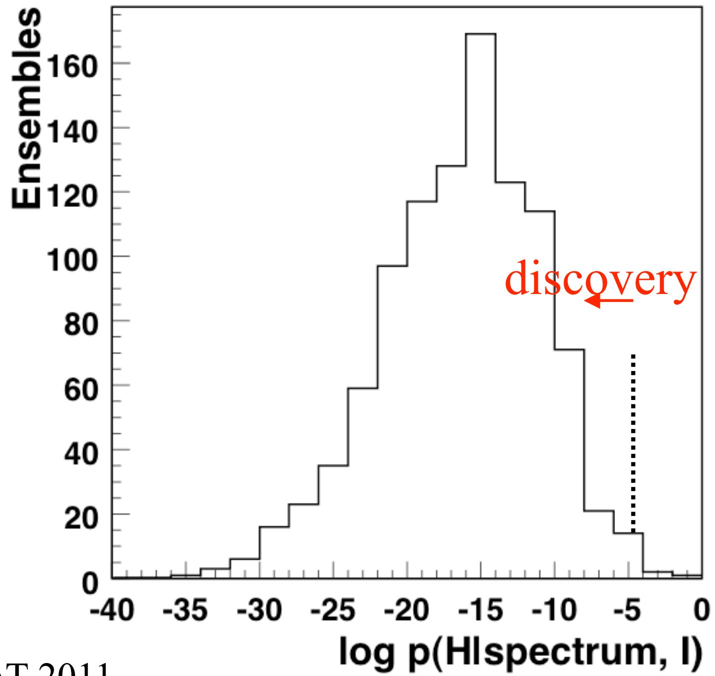
Example:

$S_{true}=16$, $B_{true}=9$





$$p(H \mid spectrum) = 2.2 \cdot 10^{-12}$$



1000 experiments simulated with

$T_{1/2}=2\ 10^{25}$ yr, 10^{-3}/(kg keV yr)

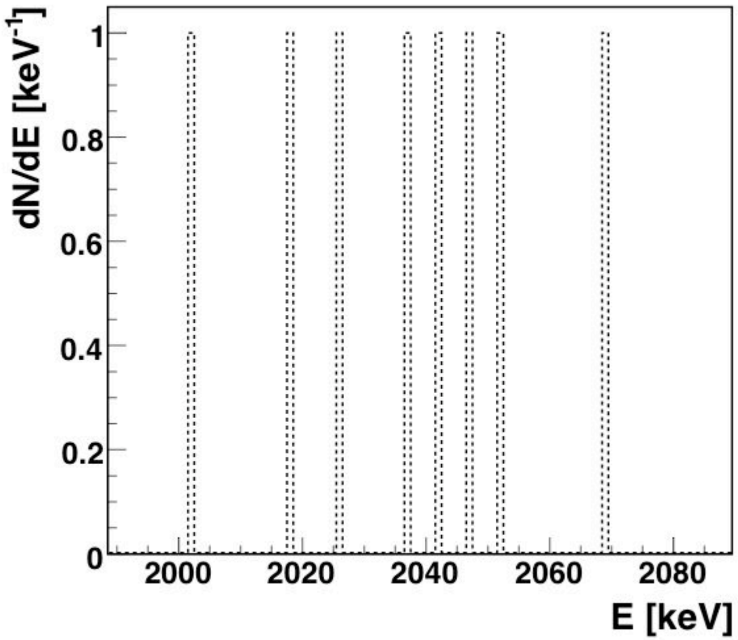
Exposure 100 kg-yr

About 95% chance a discovery could be claimed

Example:

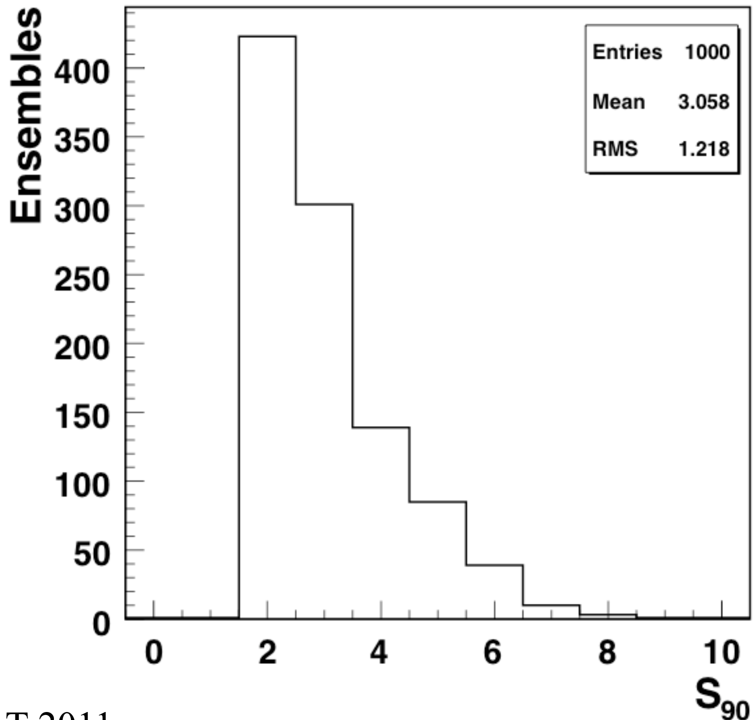
$S_{true}=0$, $B_{true}=8$

$10^{-3}/(\text{kg keV yr})$
Exposure 100 kg-yr



$$p(H\,|\,spectrum) = 0.93$$

$$S_{90} = 3.99$$



1000 experiments simulated

0 false claims of a discovery

GERDA example

To translate the event numbers into lifetimes, we use

$$S = \ln 2 \cdot \kappa \cdot M \cdot \varepsilon_{sig} \cdot \frac{N_A}{M_A} \cdot \frac{T}{T_{1/2}}$$

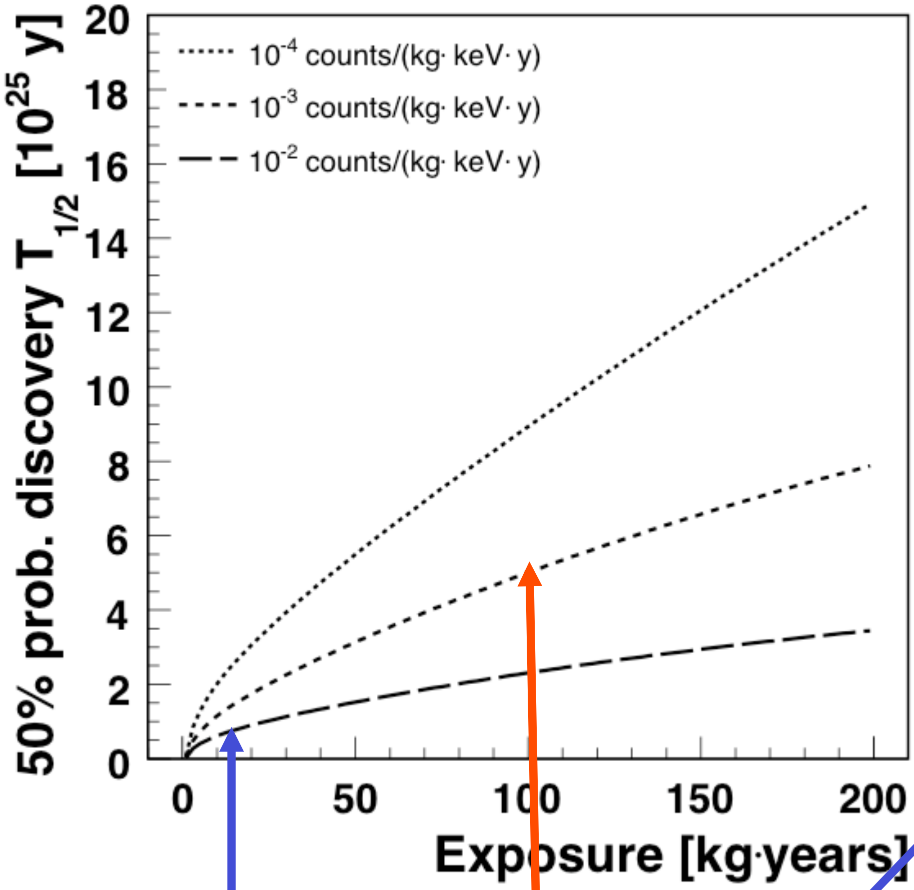
Where:
N_A is Avogadro's number
M_A is the atomic mass of ^{enr}Ge
M is the total mass of Germanium
κ is the enrichment factor (by atom, 0.86 used)
ε_{sig} is the signal efficiency (taken to be 87%)

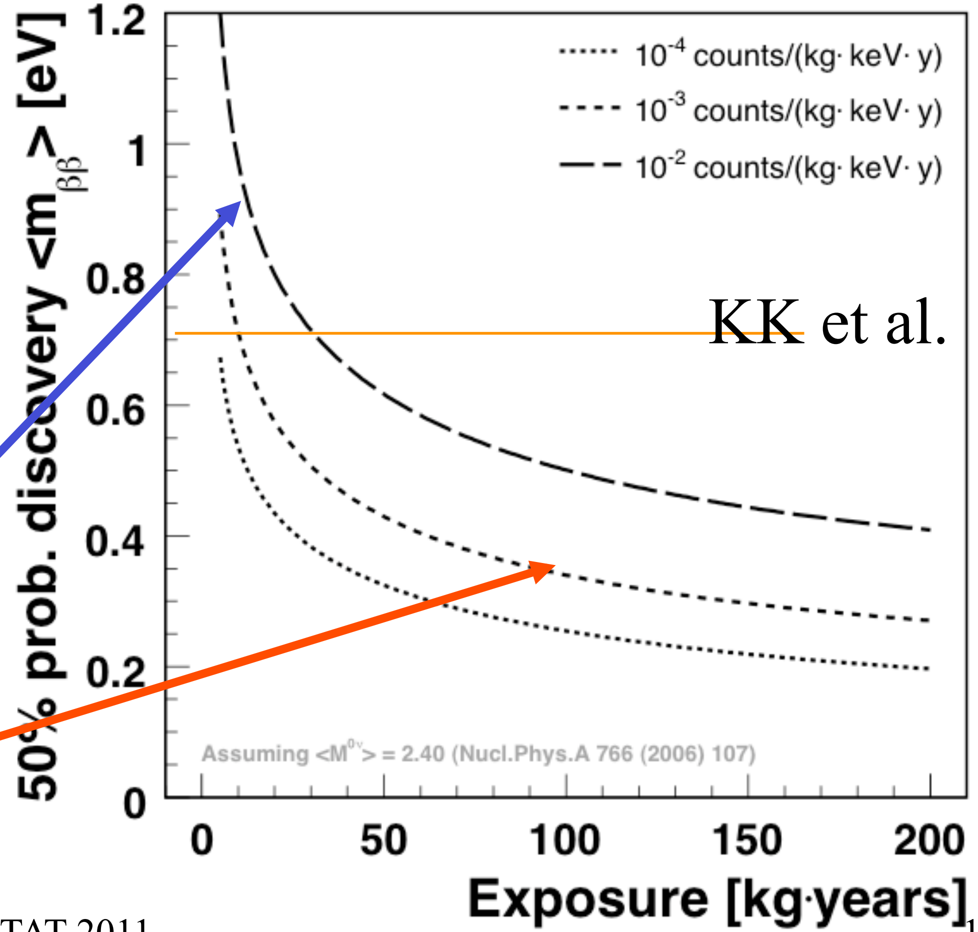
To translate $T_{1/2}$ to a mass $$\langle m_\nu \rangle = \left(T_{1/2} G^{0\nu} \right)^{-1/2} \cdot \frac{1}{M_{0\nu}}$$

$G^{0\nu}$ and $T_{1/2}$ from Rodin, Faessler, Simkovic, Vogel nucl-th/0503063

GERDA example

Bayesian analysis: discovery defined as P(background only|spectrum)<0.0001





Phase I: 15 kg-yr, existing ^{enr}Ge crystals

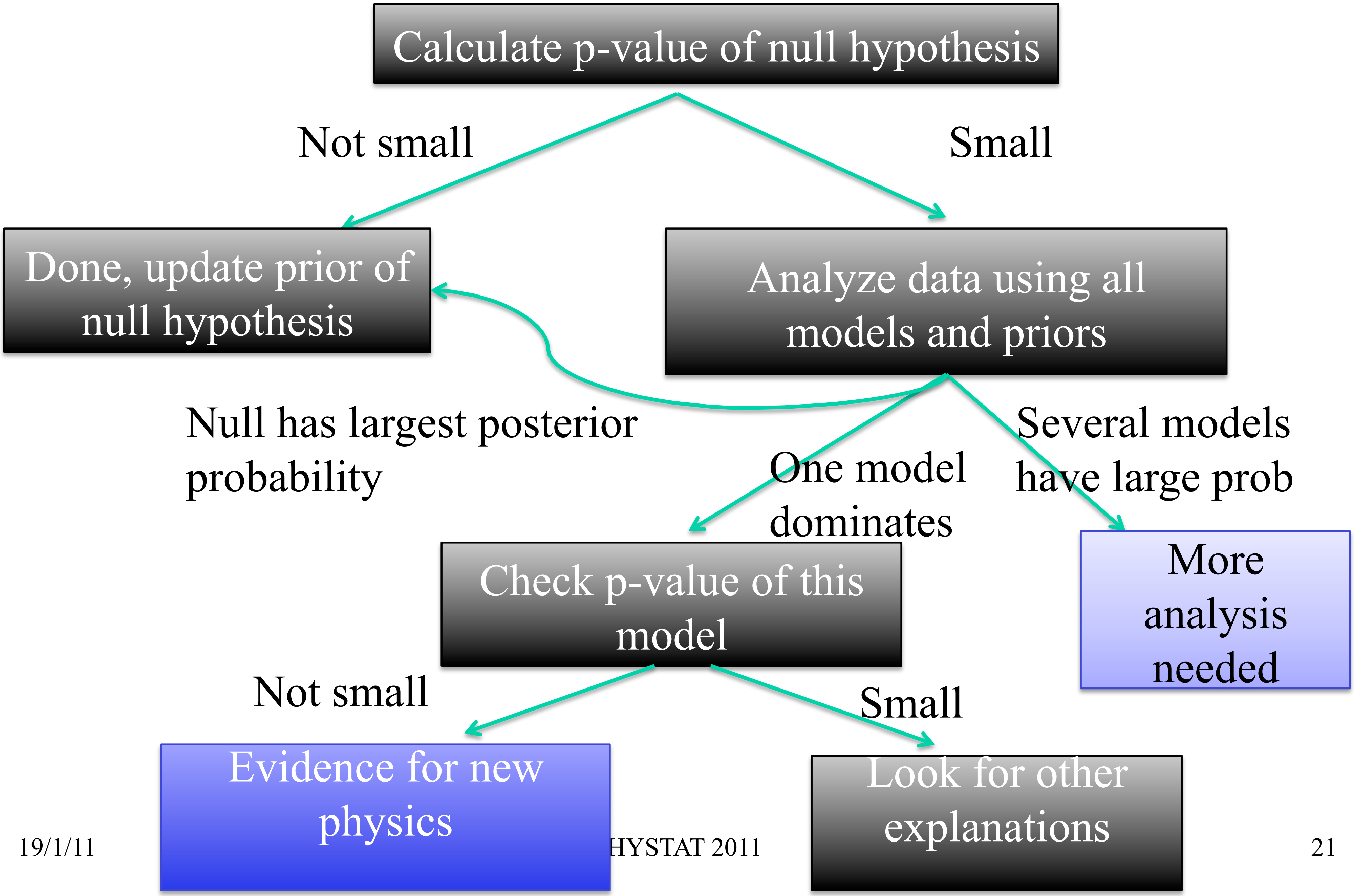
Phase II: 100 kg-yr, new segmented ^{enr}Ge crystals

Discussion

The exampled chosen was special in the sense that we could define a complete set of models, and therefore produce a coherent probability analysis. Prior beliefs could be defined, and the knowledge update from an experiment was clear. There will always be some discussion on which prior beliefs to choose.

What if we don't have a complete set of models ? No real probability analysis possible, but we still update our degree-of-belief.

Incomplete set of models



Consensus Priors

Should define informative priors whenever possible. In cases where range important, can define 2 or more consensus priors to show us what we learn from the data.

Different new physics obviously have different priors (e.g., compare Higgs at 120 GeV vs new extra dimensions, or, historically, top quark at 175 GeV vs leptoquarks at HERA).

Suggestion: form of a particle physics committee which defines consensus priors. Could be more that one prior per process (pessimistic, optimistic). Could cover all particle physics topics (dark matter, neutrino mass, Majorana, Higgs, SUSY, …)

Having consensus priors would allow for a transparent and coherent knowledge update. Consensus priors should be updated as new data becomes available.

Look Elsewhere Effect

In the double beta example, we know where to look in energy. If we scanned the energy spectrum and looked for an excess, would we get an enhancement of false discoveries (Look Elsewhere Effect) ?

the look elsewhere effect is suppressed in Bayesian analysis by a factor:

$$\sigma / \Delta$$

Where σ is the width of the expected signal and Δ is the range over which we search.

LEE

Example: 1D spectrum, search for signal anywhere in spectrum. Assume amplitude and width of new signal fixed.

$$p(H_2|D) = \frac{\int P(D|H_2,\mu)P_0(H_2,\mu)d\mu}{\int P(D|H_2,\mu)P_0(H_2,\mu)d\mu + P(D|H_1)P_0(H_1)}$$

Assume we can take:

$$P_0(H_2,\mu) = P_0(H_2)P_0(\mu|H_2)$$

$$P_0(H_2) = P_0(H_1) = 1/2$$

$$P_0(\mu|H_2) = \frac{1}{L_\mu}$$

$$p(H_2|D) = \frac{\int P(D|H_2,\mu)d\mu}{\int P(D|H_2,\mu)d\mu + L_\mu P(D|H_1)}$$

LEEE

Can write: $$\int P(D|H_2,\mu)d\mu = P(D|H_2,\mu^*)\delta_\mu$$

Mode value
for posterior

so $$p(H_2|D) = \frac{P(D|H_2,\mu^*)\delta_\mu}{P(D|H_2,\mu^*)\delta_\mu + L_\mu P(D|H_1)}$$

Degree of belief in new physics hypothesis limited by $\frac{\delta_\mu}{L_\mu} \propto \frac{\sigma}{L_\mu}$

Summary

1. Bayesian framework is natural for quantifying scientific knowledge

2. In some cases, a complete analysis possible (e.g., double beta decay). Should aim for this whenever possible.

3. Often, cannot propose a complete set of models is not available, and a hierarchical analysis is needed. Still need to compare new model to null to say you have found something better.