Halo-independence with quantified maximum entropy at DAMA/LIBRA

A. Fowlie, (2017), arXiv:1708.00181 [hep-ph]

Andrew Fowlie

September 5, 2017

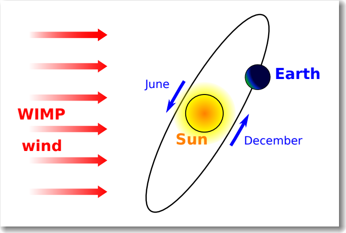
Monash University

Table of contents

DAMA/LIBRA

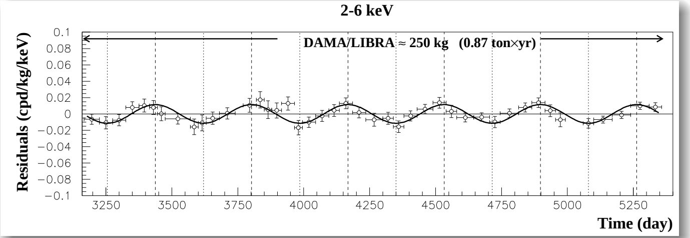
Direct detection

We can search for dark matter (DM) in direct detection experiments. DM elastic scatters with nucleons in a detector on Earth.



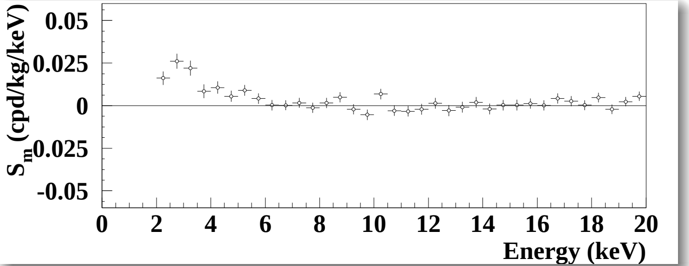
The flux of DM undergoes annual modulation because of the Earth's motion around the Sun and Sun's motion through Milky Way.

For over a decade, the DAMA/LIBRA experiment observed annual modulation of events [2–4].



The signal is $9.3\,\sigma$ in 14 annual cycles. Phase and period agree with Earth's orbit around the Sun and solar system's orbit around the Milky Way.

For over a decade, the DAMA/LIBRA experiment observed annual modulation of events [2–4].



The signal appears only at $E \lesssim 6$ keV. The high-energy bins agree with background.

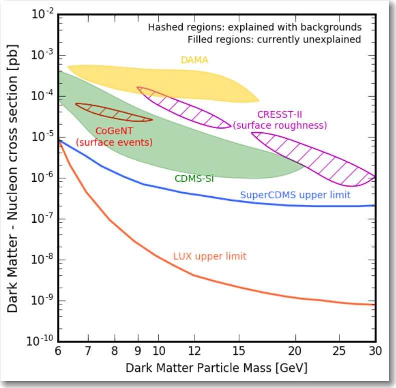
No one believes that it is dark matter. Why not?

- Conditions outside (and possibly inside) the laboratory — temperature, humidity and light — vary seasonally!

Skeptiscism

No one believes that it is dark matter. Why not?

- Competing direct detection (e.g., XENON and LUX) experiments observe no signal! [5]



Belief

DAMA/LIBRA responded with poetry — *If you can bear to hear the truth you've spoken twisted by knaves to make a trap for fools* (Kipling) — and science

- Conditions inside the laboratory are controlled and monitored
- No diurnal modulation of signal so far
- No annual modulation of mutliple-hit events or in higher energy bins
- Comparison with competing experiments requires assumptions about e.g., velocity profile, $f(v)$

No known systematic effect or background explains DAMA/LIBRA. Anomalous events are signal-like.

Why not repeat DAMA/LIBRA in the southern hemisphere?

Everyone would like such an experiment to be built, just by someone else (Bertone [6])

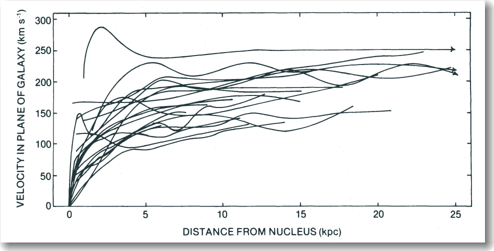
Bertone presents it as lose-lose: confirm anomaly and DAMA/LIBRA receive Nobel prize and you nothing or reject anomaly and everyone says we knew already.

Fortunately SABRE ignored Bertone's no-win theorem.

Dark matter velocity profile

The flux of DM and amplitude of scattering in detector depend on DM velocity.

We don't know the identity of DM, but we know something about its density and velocity from e.g., rotation curves.



Velocity curves. Rubin et al [7].

Dark matter velocity profile

From $F_G = mv^2/r$, we find $\rho \propto 1/r^2$.

By the collisionless Boltzmann equation, this density corresponds to Maxwell-Boltzmann

$$f(v) \propto \begin{cases} v^2 e^{-\left(\frac{v}{v_0}\right)^2} & v < v_{\text{esc}} \\ 0 & v \geq v_{\text{esc}} \end{cases}$$

We truncate it at the escape velocity of our galaxy (though don't use $\rho \propto 1/r^2$ as v_{esc} and mass would be infinite).

This neglects non steady-state effects: clumps, streams and a possible dark disk.

Modelling uncertainty in profile

What if we want to reflect our uncertainty in the DM profile?

- Parametric approach: permit variation in v_0 and v_{esc} parameters in Maxwellian profile or shape parameters in another distribution.

What if we want uncertainty about distribution not just shape parameters?

- Non-parametric approach: permit all possible profiles.

How should we handle an infinite set of profiles?

Frequentist approach

1. Throw away all prior knowledge about DM in galaxy.
2. Profile infinite set of profiles by e.g., minimising chi-squared or maximising likelihood. Provable that the ansatz

   $$f(v) = \sum_i \kappa_i \delta(v - v_i)$$

   is sufficient for minimising chi-squared and finding confidence intervals for signal rates [8, 9].
3. "Best-fit" profile not unique. One found from above procedure is an unphysical sum of delta-functions.

What if we could combine, in a coherent manner, experimental data and our background knowledge about the profile?

Maybe it isn't exactly Maxwellian, but perhaps it's something similar?

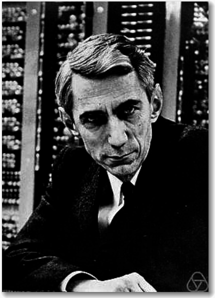
What can we do?

Quantified MaxEnt

Ask the Bayesian/information theory wizards!

Artwork Viktor Beekman and concepts Eric-Jan Wagenmakers [10].

Ask the Bayesian/information theory wizards!



Shannon's [11] information theory — Jaynes' [12] principle of maximum entropy — Skilling's [12] quantified maximum entropy.

Shannon mathematician. Jaynes/Skilling physicists. No statisticians.

Shannon entropy

Construct a measure of information learnt by receiving a message m_i that you expected with belief p_i. Requirements

- Anti-monotonic — more learnt from unexpected message
- $I \geq 0$ — information positive
- $I[p = 1] = 0$ — no new information if already certain about message
- $I[pq] = I[p] + I[q]$ — information additive for independent messages

imply that $I = -\ln p$.

Shannon entropy for discrete distributions is the expected information in a message [13]:

$$H = E[I] = -\sum_i p_i \ln p_i$$


$C=B\log_2(1+\frac{S}{N})$
$H(X)=-\sum_x p(x)\log p(x)$
$C(X,Y)=\max_{p(x)}(H(X)-H(X|Y))$
IEEE MILESTONE IN ELECTRICAL ENGINEERING AND COMPUTING
Development of Information Theory, 1939-1967
The mathematical principles of Information Theory, laid down by Claude Elwood Shannon during the period 1939-1967, set in motion a revolution in communication systems engineering. They quantified the concept of information, established fundamental limits for the representation and reliable transmission of information, and revealed the architecture of systems for approaching them. Today, Information Theory continues to provide the foundation for advances in information collection, storage, distribution, and processing.
April 2016
IEEE
Claude E. Shannon
Father of Information Theory

Who learns the information?!

In the early days, this was a point of confusion:

- The sender: but he knows what he sent!
- The "pipe"/communication channel: this is strange, but was Shannon's thought and lead to ideas about channel capacity.
- The receiver: with this interpretation, the Shannon entropy measures ignorance of receiver/how much he expects to learn from message.

Shannon's naive generalisation to continuous distributions

$$H = -\int p(x) \ln p(x) dx$$

violates Shannon's axioms and has other undesirable properties, e.g., it is not invariant under reparameterisations.

Relative Shannon entropy

Correct expression found by Jaynes by limiting density of discrete points [15]. A simple derivation: first rewrite Shannon entropy as

$$H = \ln N - \sum_i p_i \ln \frac{p_i}{\frac{1}{N}} = \ln N - \sum_i p_i \ln \frac{p_i}{u_i}$$

where N is the number of possible outcomes and $u_i = \frac{1}{N}$ is a uniform distribution upon them. We now take $N \to \infty$,

$$H = \lim_{N\to\infty} \ln N - \int p(x) \ln \frac{p(x)}{u(x)} dx$$

It is customary to omit the divergent term and make an arbitrary change of variables $y = f(x)$ such that

$$H = - \int p(y) \ln \frac{p(y)}{m(y)} dy$$

Meaning of $m(x)$

In the discrete case, the distribution with maximum Shannon information is uniform. This represents maximum ignorance.

In the continuous case, there is no such unique distribution because of covariance under changes of variable. It is the age old question, which distribution represents ignorance? This is not solved; you must pick one, $m(x)$.

Principle of Maximum Entropy (MaxEnt)

Famous problem in Bayesian statistics: which prior represents ignorance? Jaynes' used Shannon entropy to make his famous MaxEnt principle [16]:

> *The prior that represents ignorance, subject to constraints, is the maximum entropy one.*

For example, if you know only $\langle x \rangle$, the MaxEnt distribution is the exponential.

If you know $\langle x \rangle$ and $\langle x^2 \rangle$, the MaxEnt distribution is the Gaussian.

This lead to Jaynes' view that statistical mechanics was itself merely an application of his maximum entropy principle [16].

Best description of system is maximum entropy, subject to constraints upon known macroscopic variables. This reproduces predictions of statistical mechanics.

Gibbs (maybe) favoured this epistemic view of statistical mechanics.

Skilling's quantified MaxEnt

Jaynes' MaxEnt principle worked for constraints on moments, but not not noisy data, and failed to provide a measure of reliablility.

We in fact want $p(f)$ — a distribution upon possible choices of f.

Skilling [17] demonstrated that if a general rule of assigning $p(f)$ exists, it must depend on the Shannon entropy by

$$p(f) \propto e^{\beta H[f,m]}$$

where β represents the strength of our prior conviction that $f = m$.

By Bayes theorem, we can combine our data and prior into a posterior:

$$p(f|\text{data}) \propto p(\text{data}|f) \cdot p(f)$$

In our case,

$$p(f|\text{data}) \propto e^{-\frac{1}{2}\chi^2+\beta H[f,m]}$$

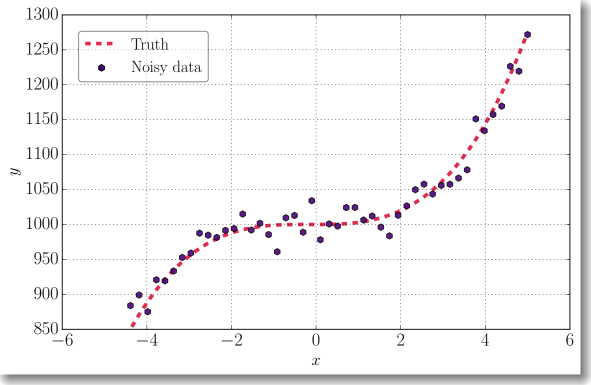
where χ^2 contains Gaussian measurements from DAMA/LIBRA.

Role of β and relation to machine learning

The regularisation parameter β mediates a competition between fitting the data and our background knowledge about the profile.

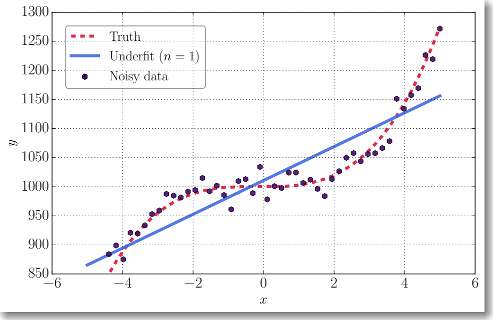
Related to idea of overfitting in machine learning

Role of β and relation to machine learning



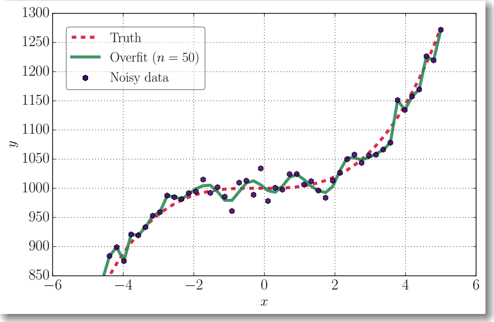
Prediction from cubic + Gaussian noise!

Role of β and relation to machine learning



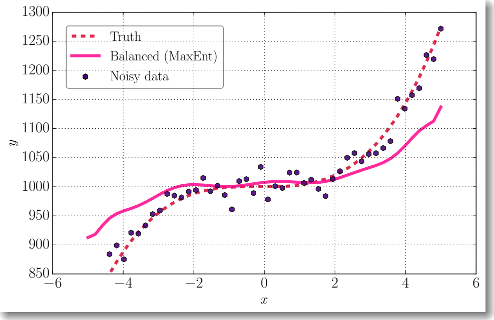
"Underfit" by linear model! $\chi^2 \approx 193$ and $-H \approx -50\,432$ such that $\chi^2 - \beta H \approx -50\,3933$.

Role of β and relation to machine learning



"Overfit" by $n = 50$ order polynomial model! $\chi^2 \approx 14$ and $-H \approx -50\,391$ such that $\chi^2 - \beta H \approx -50\,3900$.

Role of β and relation to machine learning



"Balanced fit" by MaxEnt! $\chi^2 \approx 142$ and $-H \approx -50\,424$ such that $\chi^2 - \beta H \approx -50\,4105$.

Role of β and relation to machine learning

In machine learning, the "regularization function" is ad hoc, but always a trade-off between goodness-of-fit and regularization (bias and variance).

Just as in quantified maximum entropy, there is a trade-off between goodness-of-it and prior knowledge in the entropy.

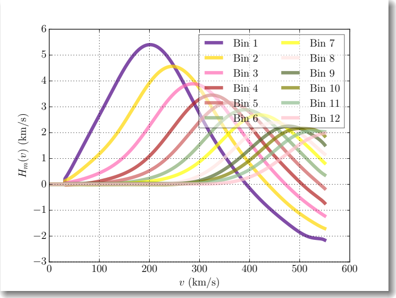
Above results were for $\beta = 10$ and $m(x) = \text{const}$ — which is why in balanced fit, non-constant tails are poorly modelled

Quantified MaxEnt at DAMA/LIBRA

We now have the ingredients that we need:

- Modulated signal in 12 energy bins from DAMA/LIBRA. Described by response function H_m^i [18] and profile for each bin measured by DAMA/LIBRA

$$\mu_i \pm \sigma_i \propto \int H_m^i(v) f(v) dv$$



Quantified MaxEnt at DAMA/LIBRA

We now have the ingredients that we need:

- Modulated signal in 12 energy bins from DAMA/LIBRA.
- Background knowledge about the velocity profile. The prior depends upon a parameter β describing our conviction that the profile is Maxwellian.

Quantified MaxEnt at DAMA/LIBRA

We now have the ingredients that we need:

- Modulated signal in 12 energy bins from DAMA/LIBRA.
- Background knowledge about the velocity profile. The prior depends upon a parameter β describing our conviction that the profile is Maxwellian.
- A formalism — quantified MaxEnt — for combining them to infer the plausibility of DM and most plausible velocity profile.

Halo-independent and halo-dependent analysis

The parameter β in the posterior

$$p(f|\text{data}) \propto e^{-\frac{1}{2}\chi^2+\beta H[f,m]}$$

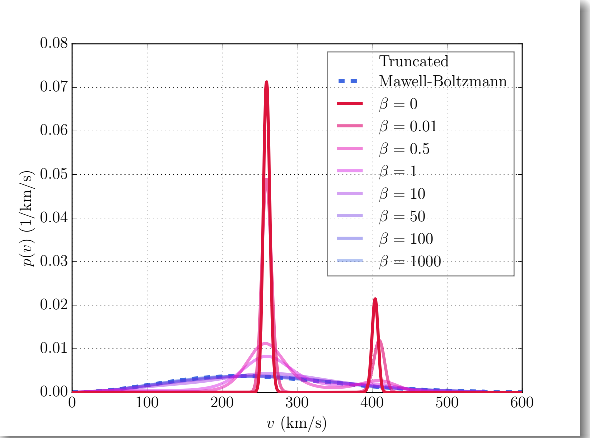
permits us to interpolate between halo-independent and dependent analysis.

- $\beta \to \infty \leftrightarrow$ halo-dependent: means that profile $\to$ Maxwellian regardless of our data as $H[f,m]$ minimum at $f = m$.
- $\beta \to 0 \leftrightarrow$ halo-independent: discards all prior information about profile.

The behaviour with β known as maximum entropy trajectory.

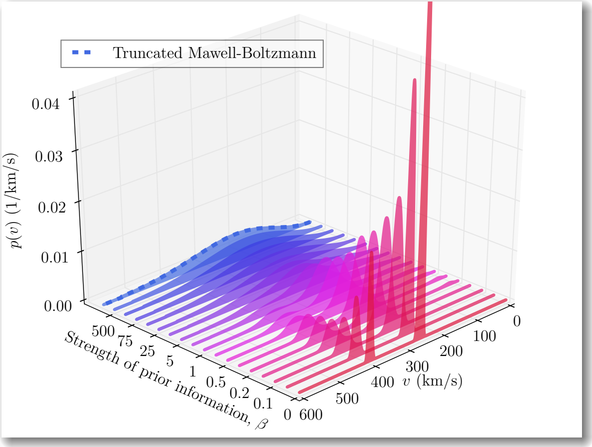
Results

By varying β, we move between a bimodal spiky velocity profile and a Maxwellian profile. The profiles are the modes of $p(f|\text{data})$.



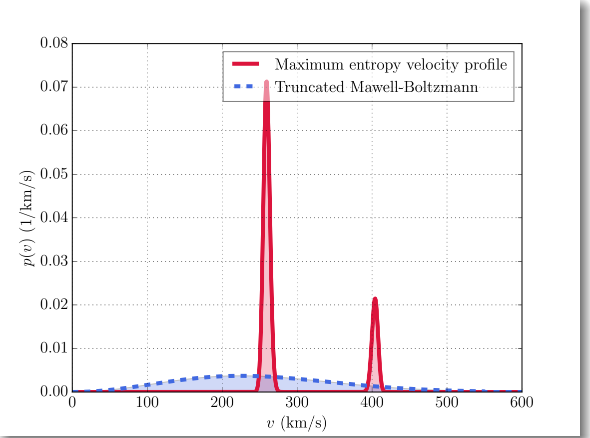
Maximum entropy trajectory for profile

By varying β, we move between a bimodal spiky velocity profile and a Maxwellian profile. The profiles are the modes of $p(f|\text{data})$.



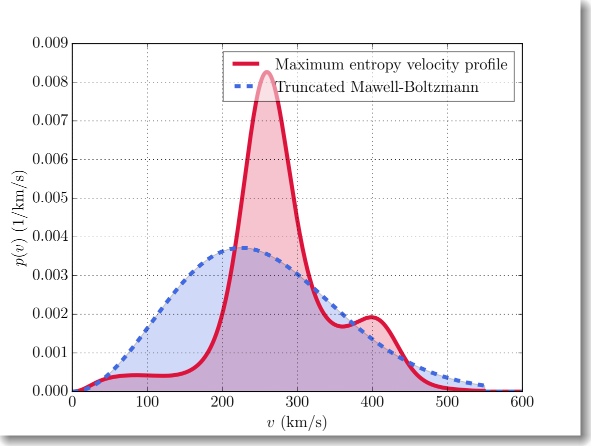
Particular profiles

$\beta = 0$ — halo-independent



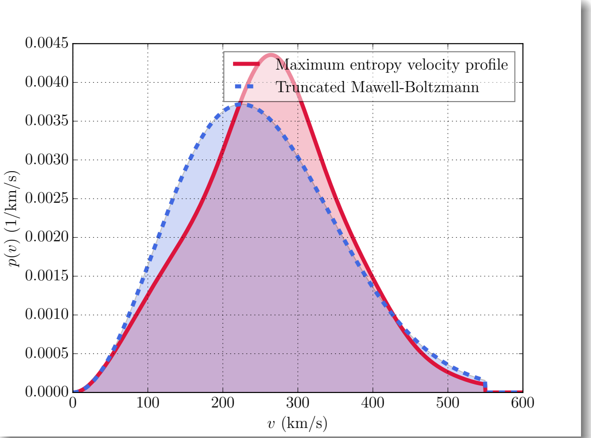
Particular profiles

$\beta = 1$



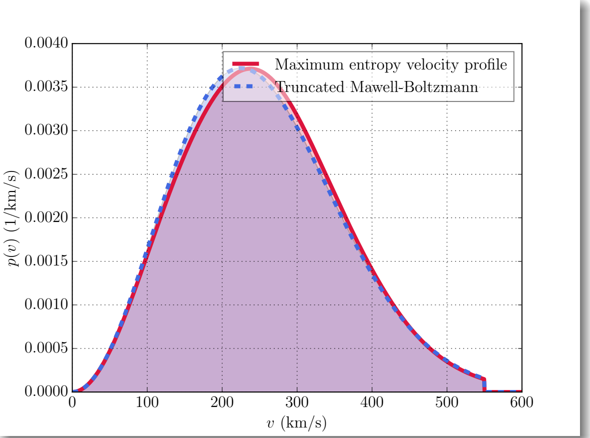
Particular profiles

$\beta = 10$

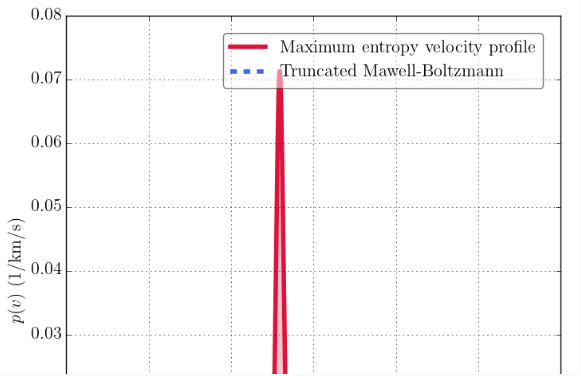


Particular profiles

$\beta = 100$ — approximately halo-dependent

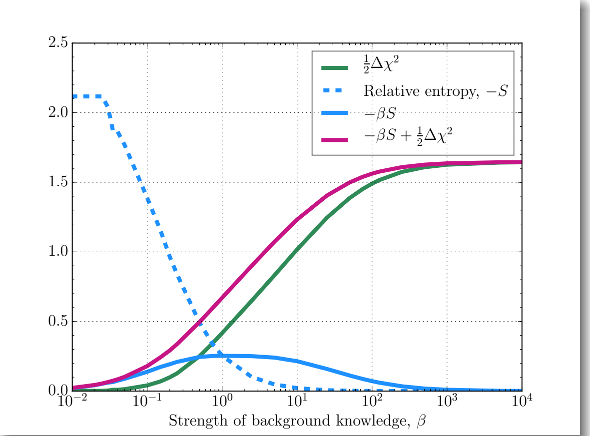


MaxEnt trajectory



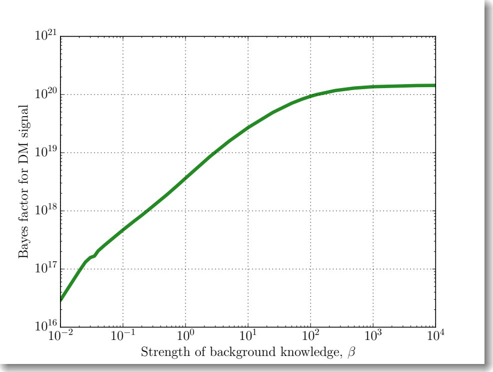
What's going on?

There's a competition between our prior knowledge, in the entropy, and data, in the chi-squared.



Plausibility of DM

Bayes factor for DM versus background only with DAMA/LIBRA data increases as β increases!



This is because Maxwellian profile is alright and increasing β concentrates prior about it/prior less diffuse.

Predictions for unmodulated moments

DM signal should look like

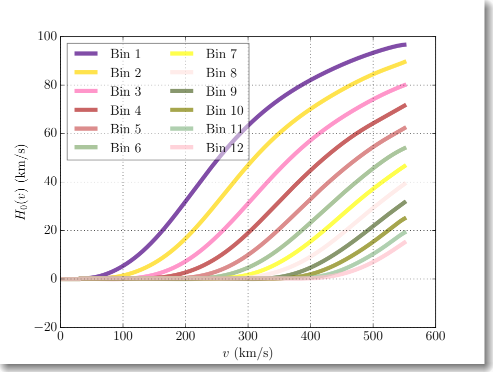
$$S = S_0 + S_m \cos(\omega t + \phi)$$

The unmodulated S_0 component poorly constrained though must be less than backgrounds.

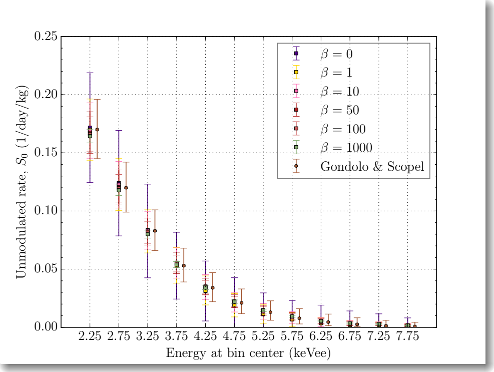
Predictions for unmodulated moments

The response functions are different but dependence on profile the same,

$$S_0 \propto \int H_0(v) f(v) dv$$



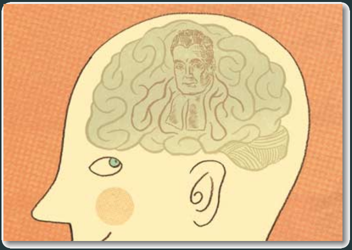
We agree with frequentist analysis that $S_0 \sim 10\%$ background



Conclusions

- Quantified MaxEnt is a technique for combining data with prior knowledge
- Strength of conviction that profile Maxwellian parameterised by β
- Interpolated between halo-independent and halo-dependent analysis
- Found that no Bayes factor in fact increase with β! No tension between DAMA/LIBRA and background knowledge
- What happens when we include conflicting data from XENON etc?
- How to model plausibility of systematic effect at DAMA/LIBRA?

Questions?



References

1. A. Fowlie, "Halo-independence with quantified maximum entropy at DAMA/LIBRA," (2017), arXiv:1708.00181 [hep-ph].
2. R. Bernabei et al., "First results from DAMA/LIBRA and the combined results with DAMA/NaI," Eur. Phys. J. C56, 333–355 (2008), arXiv:0804.2741 [astro-ph].
3. R. Bernabei et al., "New results from DAMA/LIBRA," Eur. Phys. J. C67, 39–49 (2010), arXiv:1002.1028 [astro-ph.GA].

[4] R. Bernabei et al., "Final model independent result of DAMA/LIBRA-phase1," Eur. Phys. J. C73, 2648 (2013), arXiv:1308.5109 [astro-ph.GA].

[5] https://commons.wikimedia.org/wiki/File:Direct_Detection_Constraints.png.

[6] G. Bertone, *Behind the Scenes of the Universe: From the Higgs to Dark Matter*, (Oxford University Press, 2013).

[7] V. C. Rubin, W. K. Ford, Jr., and N. Thonnard, "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/," Ap.J. 238, 471–487 (1980).

[8] A. Ibarra and A. Rappelt, "Optimized velocity distributions for direct dark matter detection," (2017), arXiv:1703.09168 [hep-ph].

[9] G. B. Gelmini, J.-H. Huh, and S. J. Witte, "Unified Halo-Independent Formalism Derived From Convex Hulls," (2017), arXiv:1707.07019 [hep-ph].

[10] *Artwork by Viktor Beekman and concepts by Eric-Jan Wagenmakers*, https://www.bayesianspectacles.org/library/.

[11] https://commons.wikimedia.org/wiki/File:ClaudeShannon_MFO3807.jpg.

[12] http://bayes.wustl.edu/.

[13] C. E. Shannon, "A Mathematical Theory of Communication," Bell System Technical Journal 27, 379–423 (1948).

[14] http://spectrum.ieee.org/tech-talk/telecom/internet/bell-labs-looks-at-claude-shannon-legacy-future-of-information-age.

[15] E. T. Jaynes, "Information Theory and Statistical Mechanics (Notes by the lecturer)," in Statistical physics 3 (1963), p. 181.

[16] E. T. Jaynes, "Information Theory and Statistical Mechanics," Phys. Rev. 106, 620–630 (1957).

[17] J. Skilling, “The Axioms of Maximum Entropy,” in *Maximum-Entropy and Bayesian Methods in Science and Engineering: Foundations*, edited by G. J. Erickson and C. R. Smith (Springer Netherlands, Dordrecht, 1988), pp. 173–187.

[18] P. Gondolo and S. Scopel, “Halo-independent determination of the unmodulated WIMP signal in DAMA: the isotropic case,” (2017), arXiv:1703.08942 [hep-ph].