The problem(s) with hypothesis tests

Daniel Mortlock

Imperial Centre for Inference and Cosmology
Astrophysics Group
Department of Physics

Imperial College London

Hypothesis tests

- If data (absolutely) contradict a hypothesis then it is proven to be incorrect, regardless of whether there is a viable alternative.
- The statistical (i.e., less absolute) equivalent is the hypothesis test.
- Qualitatively: if the observed data were "unlikely" under the null hypothesis, it can be rejected "at some level".
- Completely standard part of human reasoning (both scientific and "everyday").

Mathematical formulation

Single-tail p-value:

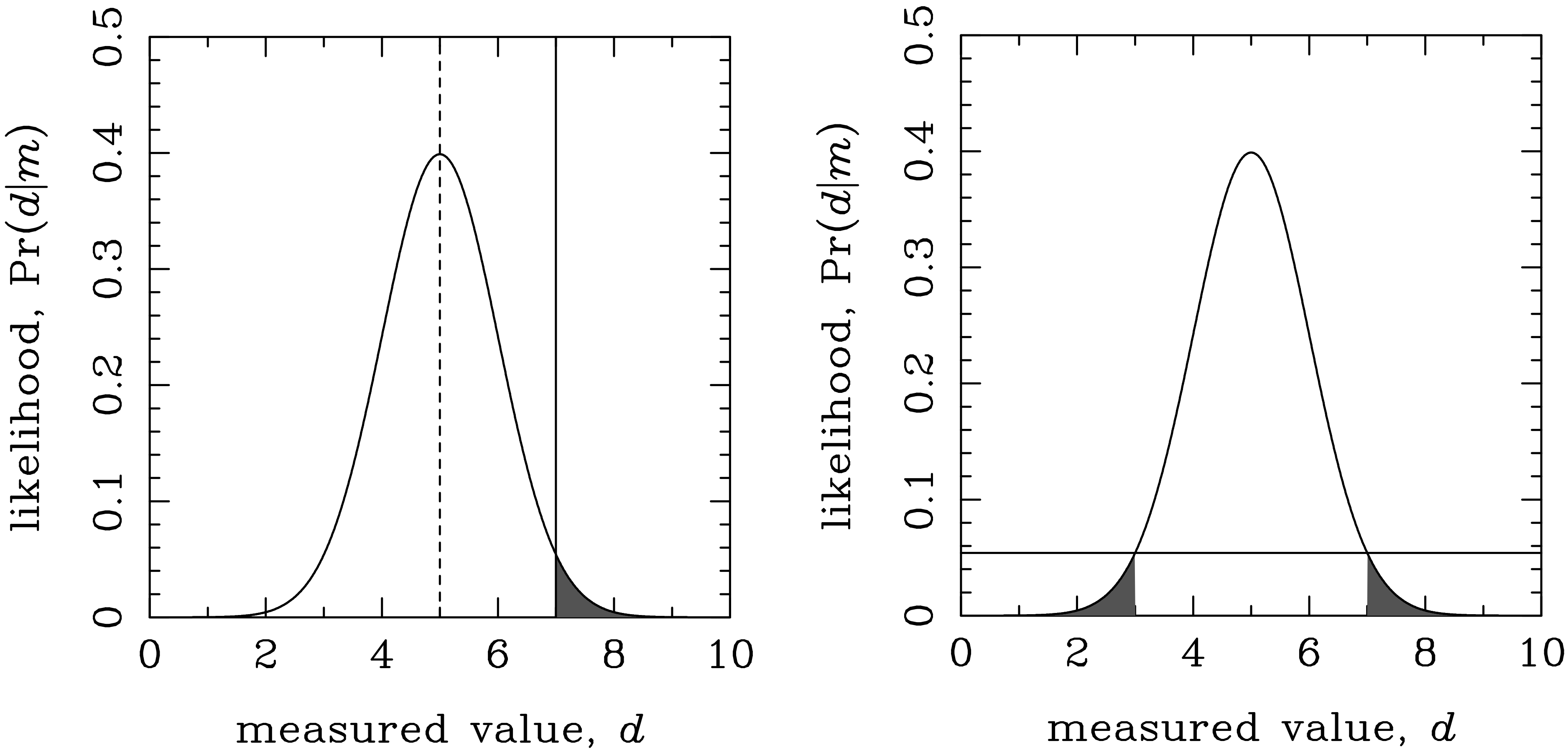
$$p = \int \mathrm{d}d' \Theta(d' - d)\, P(d'|H_0)$$

Likelihood threshold p-value:

$$p = \int \mathrm{d}d' \,\Theta[P(d|H_0) - P(d'|H_0)]\, P(d'|H_0)$$

(Other, related formulations possible.)

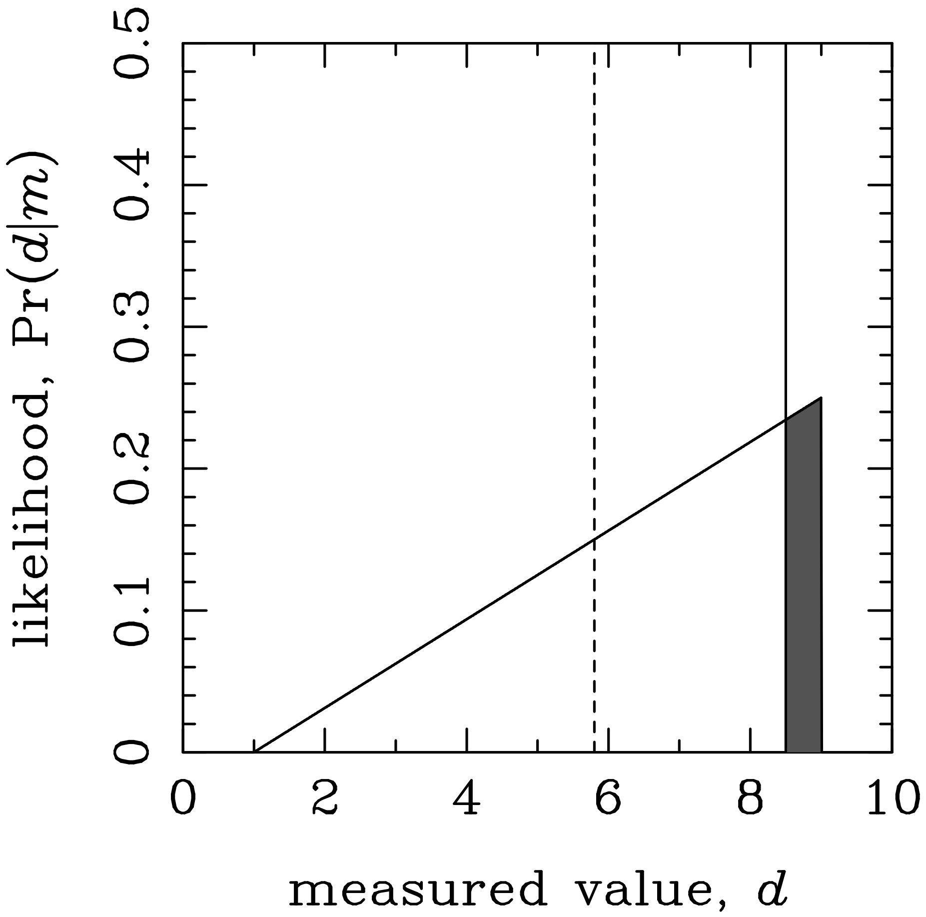
Centrally-peaked distribution



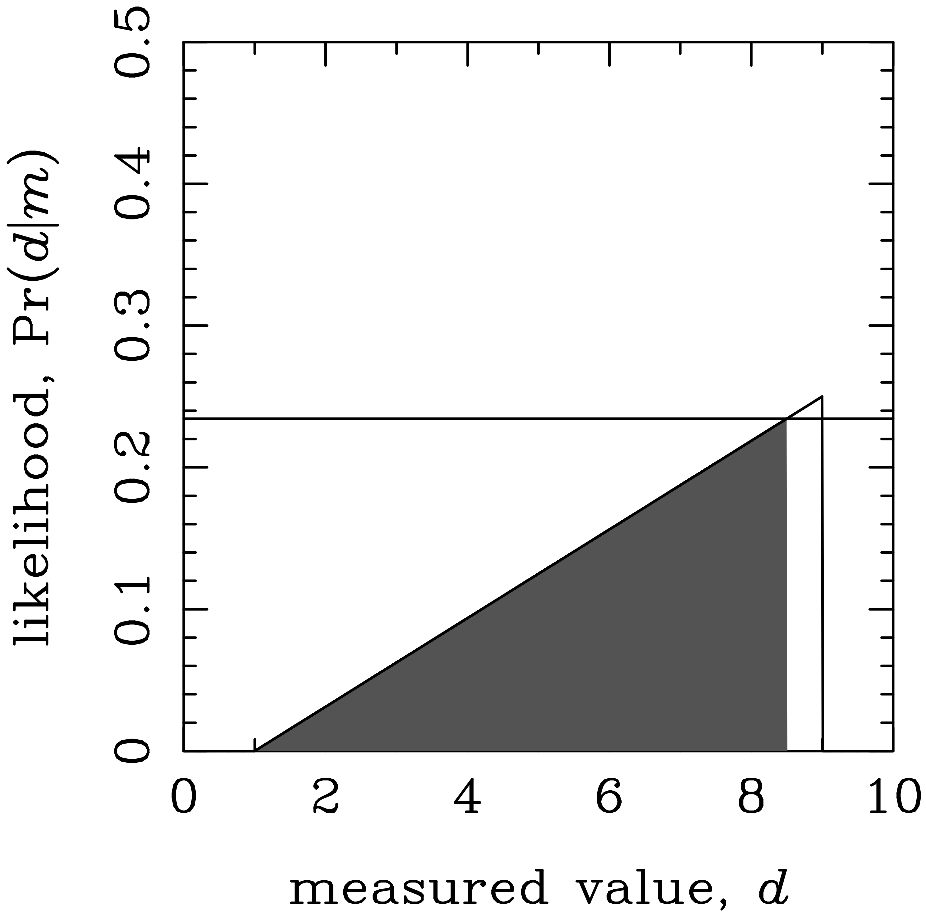
single-tail p-value

likelihood-based p-value
(same as two-tail p-value)

Asymmetric distribution

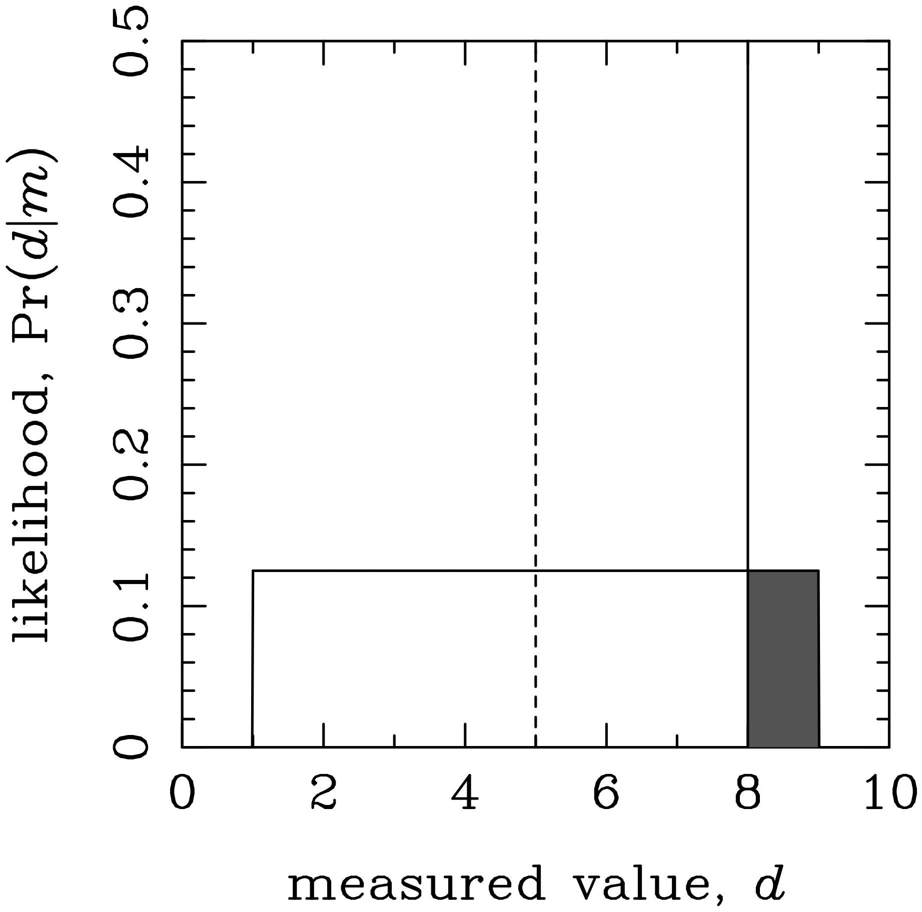


single-tail p-value

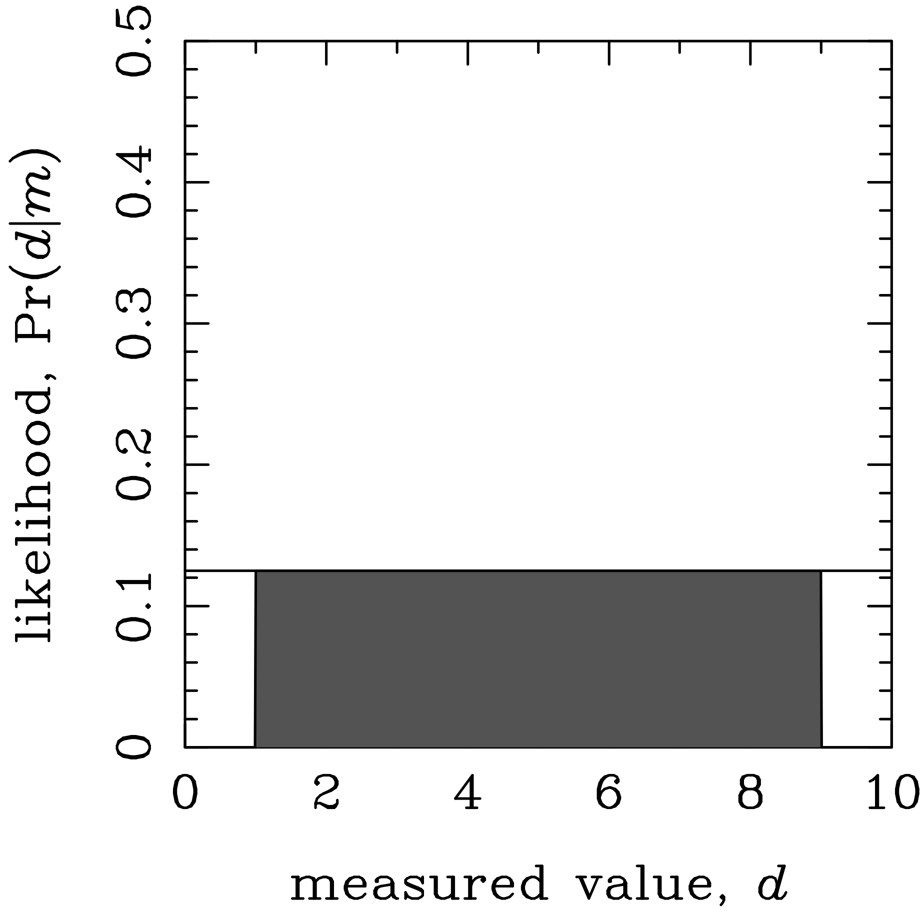


likelihood-based p-value

Uniform distribution

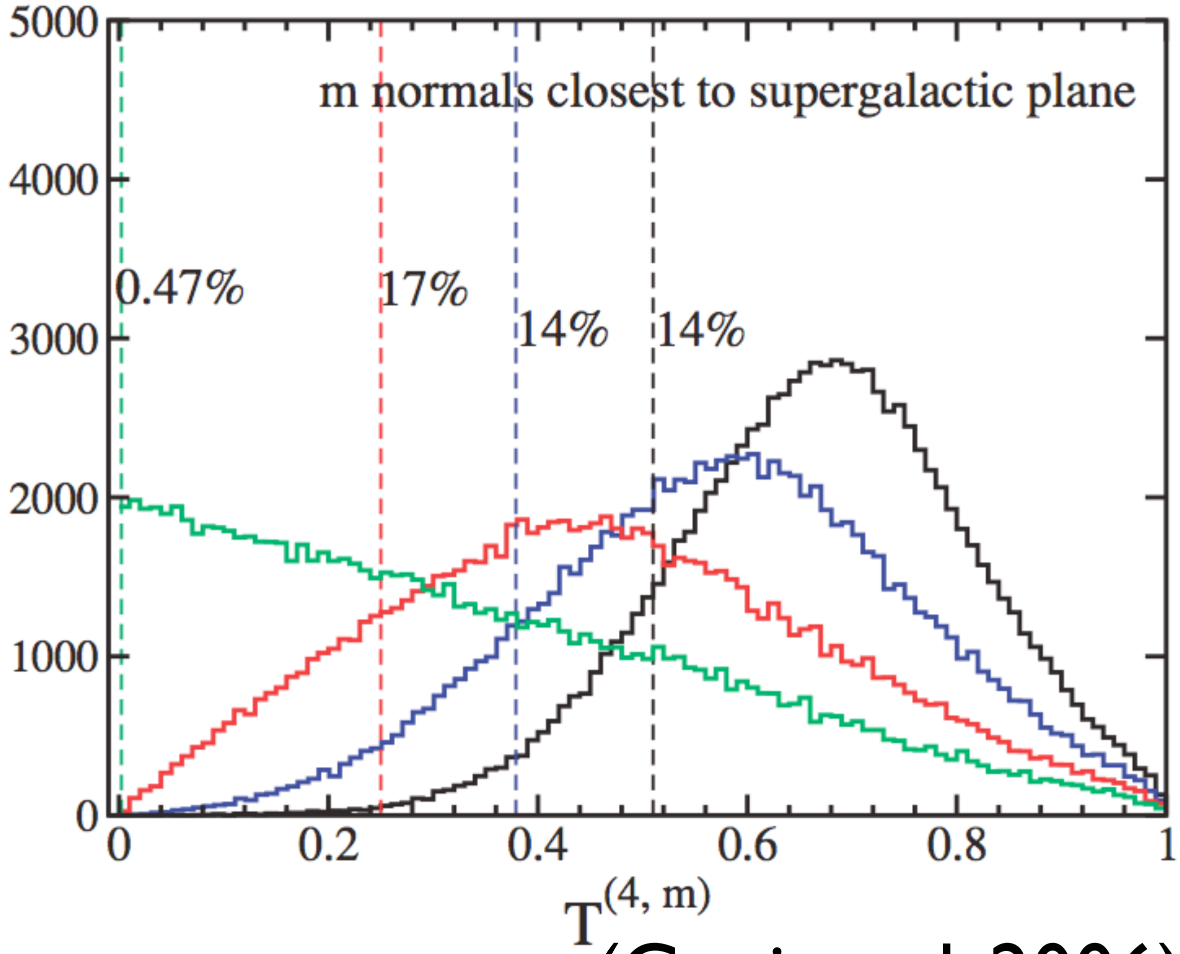


single-tail p-value



likelihood-based p-value

Cosmological example



(Copi et al. 2006)

(Test used to assess whether the CMB is isotropic)

Bayesian inference

- I (and many others!) want to ascribe probabilities to models/hypotheses.
- Cox (1946) proved that the only valid method for manipulating such probabilities is Bayesian inference (or mathematically equivalent formulations).
- Prescriptive methodologically (so no freedom about one- or two-tails, etc.)
- Hence I would like to be able to base all my reasoning on (posterior) probabilities obtained by applying Bayes's theorem.

Bayesian inference vs. hypothesis tests

- Bayesian inference requires all possible models to be included in the calculation.
- Hypothesis tests are only concerned with one model, which could be rejected without an alternative.
- Apparent implication: single-model hypothesis tests cannot satisfy the Cox (1946) self-consistency criteria.

Bayesian model comparison

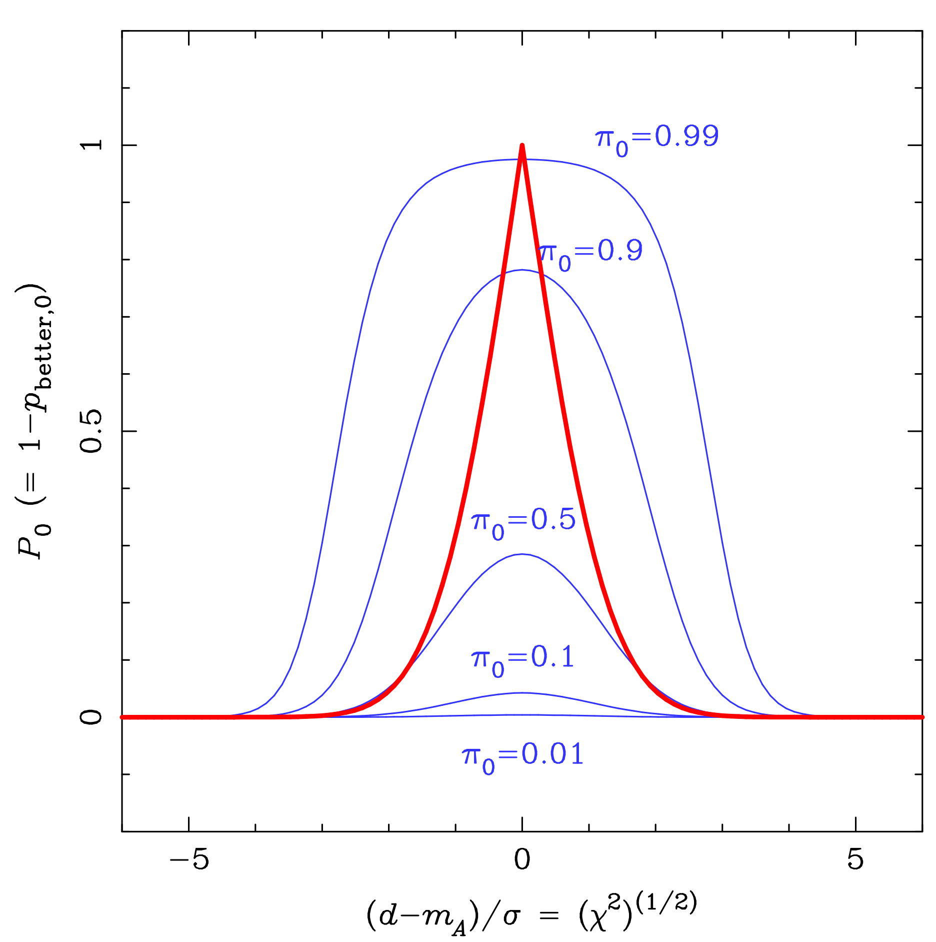
$$P(M_1|d) = \frac{P(M_1)\,P(d|M_1)}{P(M_1)\,P(d|M_1) + [1 - P(M_1)]P(d|M_2)}$$

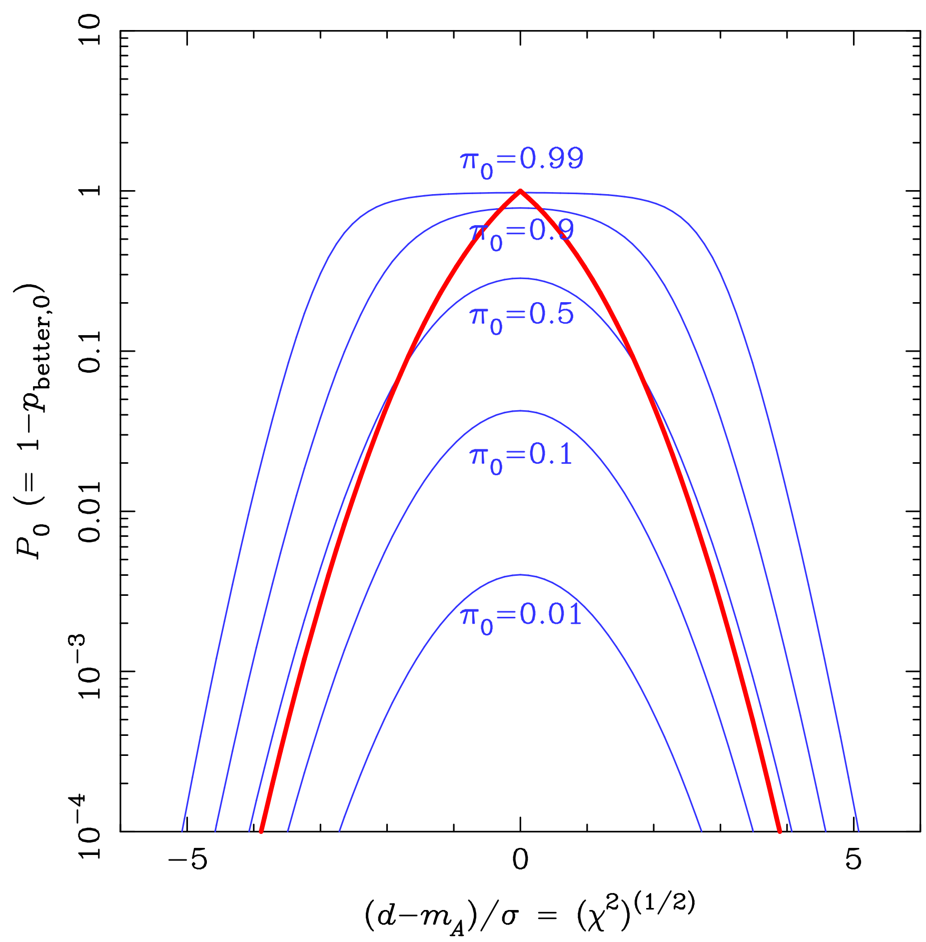
Suggested hypothesis test structure:

$$P(H_0|d) = \frac{P(H_0)\,P(d|H_0)}{P(H_0)\,P(d|H_0) + [1 - P(H_0)]\boxed{P(d|???)}}$$

If denominator is maximised (e.g., "just so" model) then posterior is minimised, giving a lower bound …

Single, normally-distributed measurement:





Standard p-value based hypothesis test

Bayesian model comparison against "just so" model

(cf Lindley's paradox)