technische universität dortmund

astroparticle physics

Project F5

Asimov Datasets for gamma-ray Astronomy

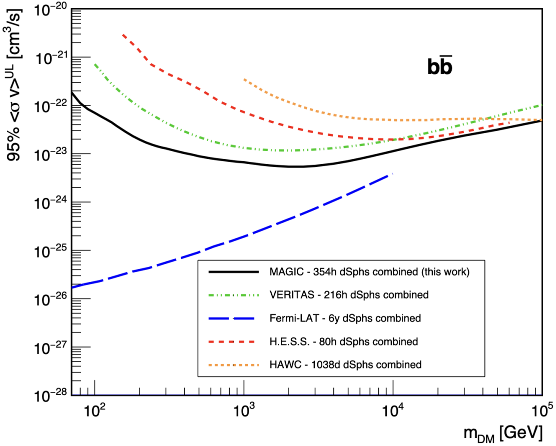
Stefan Fröse

7. November 2023

Astroparticle Physics
TU Dortmund

Motivation: Hunt for Dark Matter

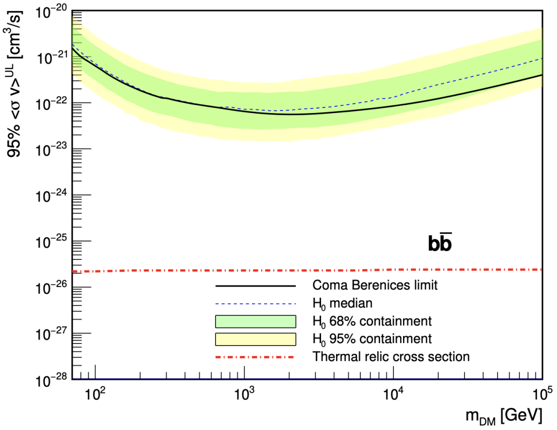
- Suitable targets: Dwarf spheroidal galaxies (dSphs)
- Observations by MAGIC, Fermi-LAT, HAWC, H.E.S.S., VERITAS, ...
  → Combination of all measurements
- **Problem:** Different analysis methods, different analysis chains → **Not** open source and **not** standardized
- **Goal:** Upper limits on the annihilation cross section in a **reproducable** and **accessible** way



arxiv:2111.15009

Motivation: Hunt for Dark Matter

- Classical upper limit plot consists of:
  - Upper limits for the observed signal
  - Median, 1σ-region, 2σ-region for the expected signal
  - Problem: Computation of the error band is computational expensive
    → Requires lots of toy simulations
- Goal: Make it faster and more sustainable



arxiv:2111.15009

Reproducable, Accessible, Faster & Sustainable

Open Source Gamma-ray astronomy

- All experiments produce so-called **EventList's**
- Gamma Astro Data Formats (GADF) define a common format for these EventList's
  → Data Level 3 (DL3) doi:10.3390/universe7100374
- Open source software for the analysis of gamma-ray data: **Gammapy**
  → Can read DL3 files and produce all sorts of science products
  → Flux Maps, SEDs, Lightcuves, Source Catalogues, ...

⇒ Accessible & Reproducable: Build upper limit analysis on top of Gammapy (GitHub:TITRATE)

A **Python** package for **gamma-ray** astronomy

doi:10.1051/0004-6361/202346488

Energy	RA	DEC
12 186.64 MeV	260.4594°	-33.553 337°
25 496.60 MeV	261.3751°	-34.395 004°
15 621.50 MeV	259.5697°	-33.409 416°
12 816.32 MeV	273.9588°	-25.340 391°
18 988.39 MeV	260.8568°	-36.355 804°

Poisson Statistics

$$L(\mu, b) = \frac{(\mu s + b)^{N_{\mathrm{ON}}}}{N_{\mathrm{ON}}!} e^{-(\mu s + b)} + \frac{(\tau b)^{N_{\mathrm{OFF}}}}{N_{\mathrm{OFF}}!} e^{-\tau b}$$

μ = strength parameter

s = signal counts from model

N_{ON} = signal counts from data

b = background counts from model

N_{OFF} = background counts from data

xkcd:892

Test Statistic

Profiled log-likelihood ratio (pLLR):

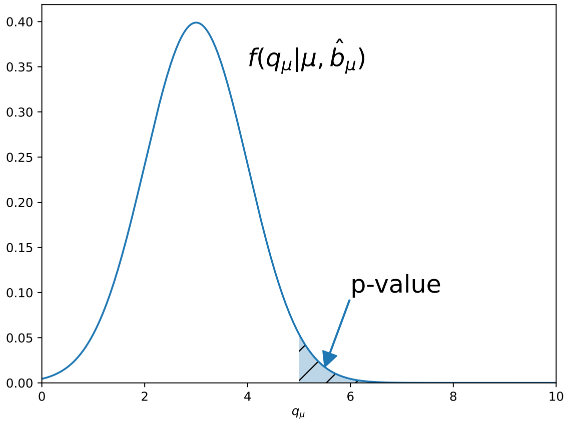
→ Null hypothesis

$$q_\mu = -2\log\frac{L(data|\mu,\hat{\hat{b}}_\mu)}{L(data|\hat{\mu},\hat{b})}$$

p-value: Probability that result at least as extreme as observed result under true null hypothesis

→ small p-value means very unlikely under null hypothesis

$$p = \int_{q_\mu^{obs}}^{\infty} f(q_\mu|\mu,\hat{\hat{b}}_\mu)\,dq_\mu$$

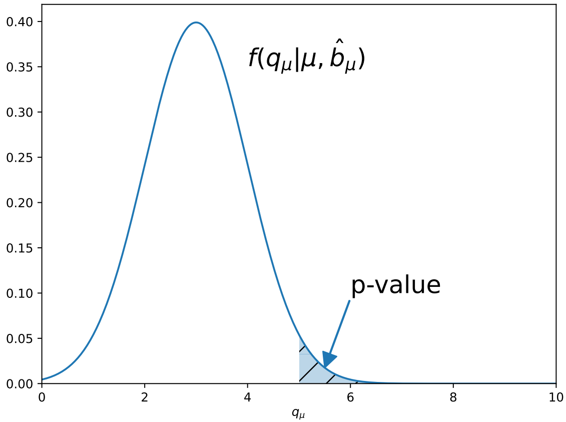


Significance - No Signal

$$\mu = 0$$

$$p = \int_{q_0^{\text{obs}}}^{\infty} f(q_0|0, \hat{b}_0)\, dq_0$$

→ small p-value means: background only is very unlikely



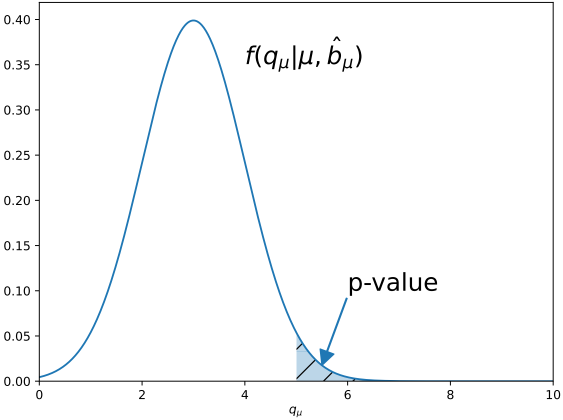
⇒ Calculate significane from p-value

Upper Limit

$$p = \int_{q_{\mu_{UL}}^{obs}}^{\infty} f(q_{\mu_{UL}} | \mu_{UL}, \hat{b}_{\mu_{UL}})\, dq_{\mu_{UL}}$$

→ Find UL for Confidence Level (CL) of **95 %**

$\Rightarrow p = 1 - \mathrm{CL}$



Problem

We need to know $f(q_\mu | \mu, \hat{b}_\mu)$

Asimov Dataset

- **Solution 1:** Generate a lot of toy datasets to compute $f(q_\mu|\mu,\hat{b}_\mu)$
- **Solution 2:** Use Wald's take on the approximation $\rightarrow \chi^2$ distributed

$$-2\log\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N})$$

$\Rightarrow$ Gaussian distributed! But how do we find $\boldsymbol{\sigma}$? $\Rightarrow$ Asimov dataset

Asimov Dataset

- Generate dataset where the expected counts are equal to the measured counts
- Calculating the estimators of the dataset obtains the true parameters
- $n_i = E[n_i] = \mu' s_i + b_i$
- Asimov likelihood:

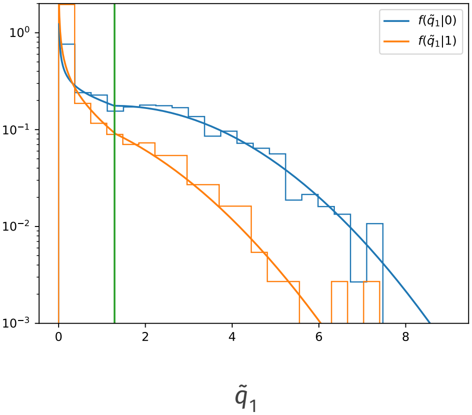
$$\lambda_A = \frac{L(\mu, \hat{b}_\mu)}{L(\hat{\mu}, \hat{b})} = \frac{L(\mu, \hat{b}_\mu)}{L(\mu', b)}$$

- Find σ:

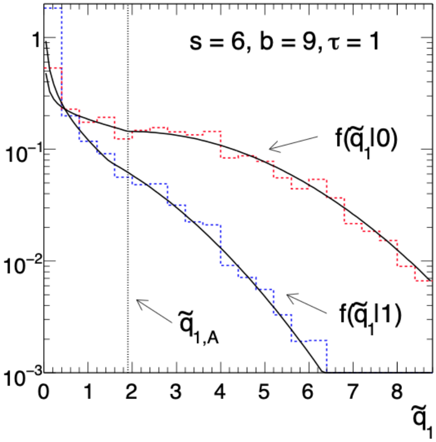
$$V_{jk}^{-1} = -E\left[\frac{\partial^2 lnL}{\partial\theta_j \partial\theta_i}\right] = -\frac{\partial^2 lnL_A}{\partial\theta_j \partial\theta_i}$$

Distribution

Gammapy + TITRATE



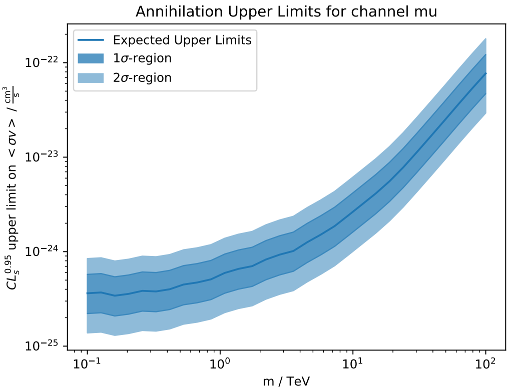
Original Paper



arxiv:1007.1727

Finally

- UL data: $\mu_{\mathrm{up}} = \hat{\mu} + \sigma\Phi^{-1}(1-\alpha)$
- Median: $med[\mu_{\mathrm{up}}|\mu'] = \mu' + \sigma\Phi^{-1}(1-\alpha)$
- Band: $band_{N\sigma} = \mu' + \sigma(\Phi^{-1}(1-\alpha) \pm N)$



⇒ No use of toy MCs → Faster and more sustainable

Conclusion & Outlook

- Goals achieved
- Asimov dataset is a good approximation
- Test in real analysis → MAGIC Coma dwarf
- **Next:** Proposal for observation time with MAGIC+LST1 for dSphs

Reproducable

Accessible

Faster

Sustainable