Signal estimation in On/Off measurements including event-by-event variables

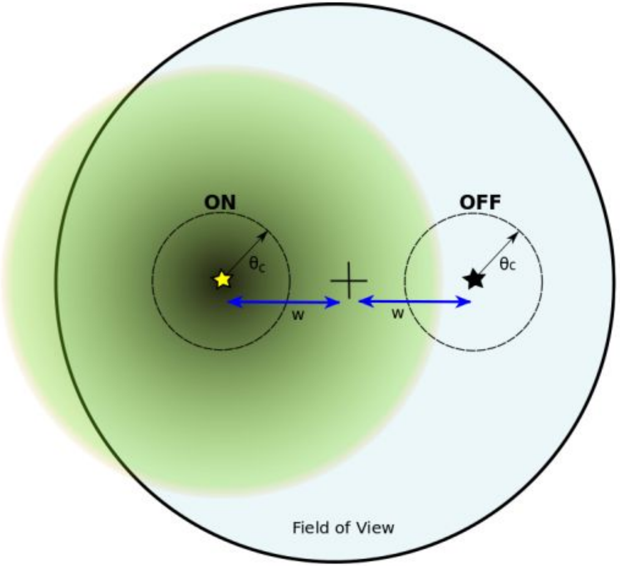
Statistic exam presentation
Professor: Paolo Francavilla
PhD student: Franjo Podobnik
University of Siena, PhD cycle XXXVII.
18.05.2023.

Outline

- Introduction
- Comparison between frequentist and bayesian approach
- PDF of the signal rate including single-event observables
- The case of IACT’s
- Conclusion

Introduction

- What are the On/Off measurements?



Introduction

- If we know the signal flux s, the number of signal events N_s in the On measurement would be a random variable following a Poisson distribution:

$$p(N_s \mid s) = \frac{(\xi \cdot s)^{N_s} e^{-\xi \cdot s}}{N_s!}, \quad s \geq 0.$$

$$\boxed{\xi = t_{eff} \cdot A_{eff}} = 1$$

variable	description	property
N_{on}	number of events in the On region	measured
N_{off}	number of events in the Off region	measured
α	exposure in the On region over the one in the Off regions	measured
b	expected rate of occurrences of background events in the Off regions	unknown
s	expected rate of occurrences of signal events in the On region	unknown
N_s	number of signal events in the On region	unknown

Introduction

- Assuming flat priors p(s) and p(b) and applying Bayes theorem we get PDF:

Posterior likelihood Priors

$$\boxed{p(s \mid N_{on}, N_{off}; \alpha)} = \frac{\int db\, \boxed{p(N_{on}, N_{off} \mid s, b; \alpha)}\boxed{p(b)\, p(s)}}{\int ds\, db\, p(N_{on}, N_{off}, s, b; \alpha)} \propto \int db\, p(N_{on}, N_{off} \mid s, b; \alpha).$$

- The likelihood function can be expressed as:

$$p(N_{on}, N_{off} \mid s, b; \alpha) = p(N_{on} \mid s, \alpha b) \cdot p(N_{off} \mid b) = \frac{(s+\alpha b)^{N_{on}}}{N_{on}!} e^{-(s+\alpha b)} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b},$$

- Using binomial identity we factorize likelihood in two Poisson distributions:

$$p(N_{on}, N_{off} \mid s, b; \alpha) \propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on}+N_{off}-N_s)!}{(1+1/\alpha)^{-N_s}(N_{on}-N_s)!} \cdot \frac{s^{N_s}}{N_s!} e^{-s} \times \frac{(b(1+\alpha))^{N_{on}+N_{off}-N_s}}{(N_{on}+N_{off}-N_s)!} e^{-b(1+\alpha)}.$$

- The first integral now is as follows:

$$p(s \mid N_{on}, N_{off}; \alpha) \propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on}+N_{off}-N_s)!}{(1+1/\alpha)^{-N_s}(N_{on}-N_s)!} \cdot \frac{s^{N_s}}{N_s!} e^{-s}.$$

- Using identities (marginalization)

$$p(s \mid N_{on}, N_{off}; \alpha) = \sum_{N_s=0}^{N_{on}} p(N_s \mid N_{on}, N_{off}; \alpha) \cdot p(s \mid N_s) \qquad p(s \mid N_s) = \xi \frac{(s\cdot\xi)^{N_s}}{N_s!} e^{-\xi \cdot s}.$$

- We can compare last two equations in order to obtain PMF of the variable N_s:

$$\boxed{p(N_s \mid N_{on}, N_{off}; \alpha) \propto \frac{(N_{on}+N_{off}-N_s)!}{(1+1/\alpha)^{-N_s}(N_{on}-N_s)!}}$$

Allows us to define the most probable value (the mode) as an estimation of the number of signal events

Check references for complete procedure (2)

Comparison between frequentist and Bayesian approach

- On/Off measurement problem is often solved using frequentist approach. In this approach we usually define following test statistic:

$$\lambda(s) \equiv \frac{p(N_{on}, N_{off} \mid s, b = \hat{b}\ ;\ \alpha)}{p(N_{on}, N_{off} \mid s = N_{on} - \alpha N_{off}\ ,\ b = N_{off}\ ;\ \alpha)}$$

Maximization of the likelihood

$$\boxed{\hat{b}} = \frac{N + \sqrt{N^2 + 4(1+\alpha)sN_{off}}}{2(1+\alpha)}$$

- Advantage of this equation is that according to Wilks theorem, the function $-2\log\lambda(s)$ has an approximate chi square distribution with 1 degrees of freedom, which can be used to extract confidence intervals.

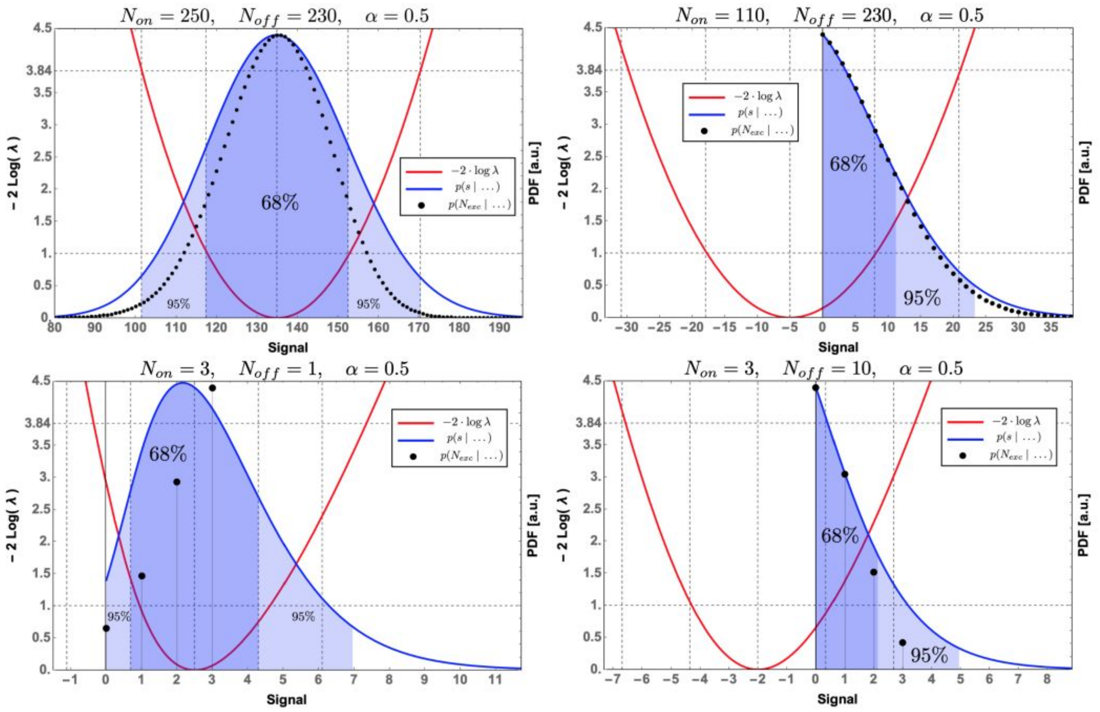
$$s = (N_{on} - \alpha N_{off}) \pm \sqrt{N_{on} + \alpha^2 N_{off}}$$

- The problems with ML approach defined are:
  - It only works well with counts number large enough. Bayesian approach has no limitations on that
  - Only information about confidence intervals can be extracted. What is probability of having N_s signal events in a sample of N_{on} events can't be answered.
  - The frequentist result does not exclude negative rate but Poisson process conflicts with the negative.
- Another advantage of Bayesian approach is that, once we have the PDF of the signal rate. All information are encoded in $p(s \mid N_{on}, N_{off}; \alpha)$

$$\int_{s_{left}}^{s_{right}} p(s \mid N_{on}, N_{off}; \alpha) ds = 0.68, \quad \text{with}$$

$$p(s_{left} \mid N_{on}, N_{off}; \alpha) = p(s_{right} \mid N_{on}, N_{off}; \alpha)$$

Comparison between frequentist and Bayesian approach



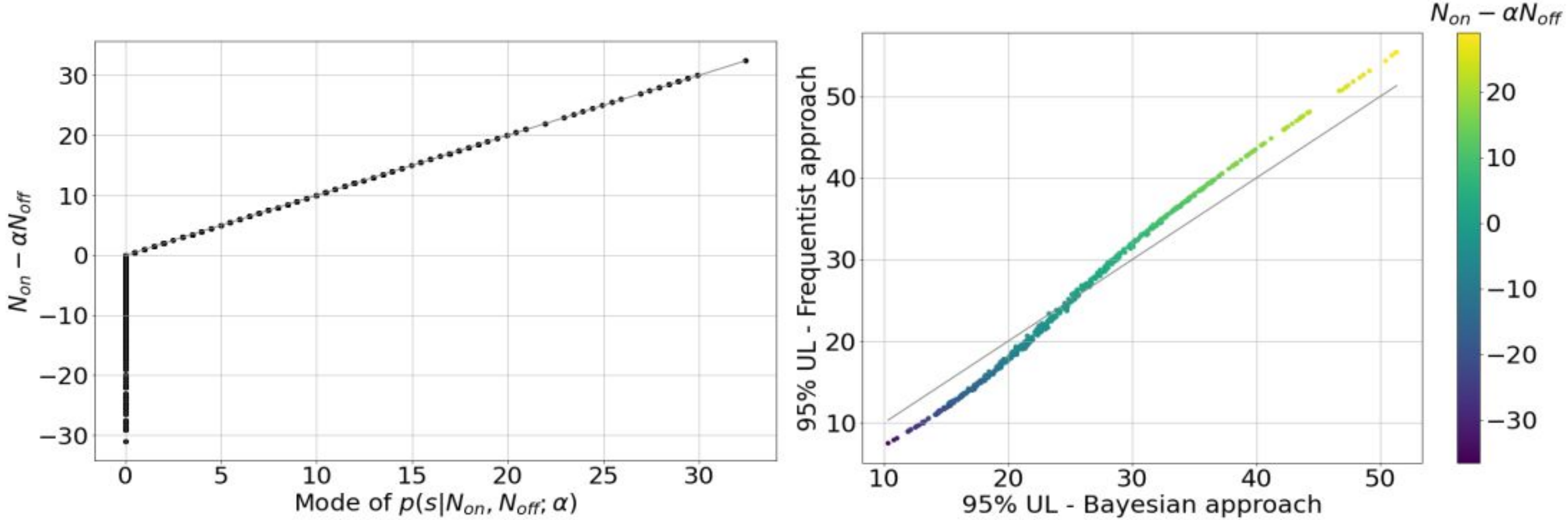
Comparison between Bayesian and frequentist approaches in estimating the signal rate.

Black dots - PMF of the N_s

Blue line - PDF of signal rate

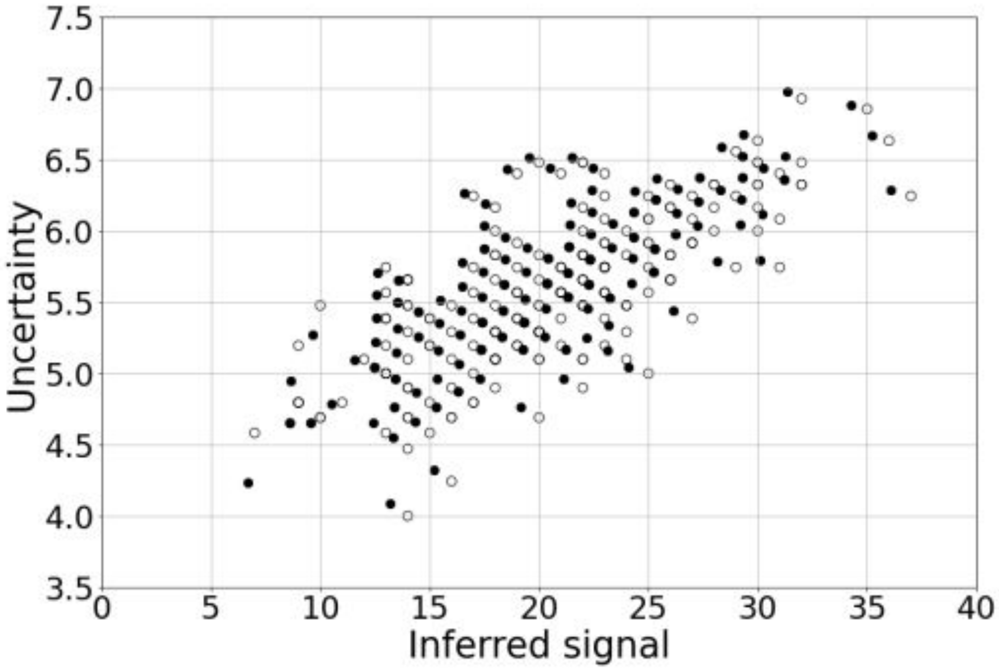
Red line - function $-2\log\lambda(s)$

Comparison between frequentist and Bayesian approach



Inferred signal from Bayesian and frequentist approach using MC simulations. (Left)
95% UL comparison between both approaches (Right)
Simulations made with expected signal s=0, background b=200 and alpha 0.5

Comparison between frequentist and Bayesian approach



- Inferred signal and its uncertainty (68% confidence/credibility band) from MC simulations in which s=20, b=5 and α=1. Filled circles mark Bayesian approach and empty circles mark frequentist approach.
- Good agreement between both approaches

PDF of the signal rate including single-event observations

- Defined PMF of the number of signal events is based on the number of events in the On and Off regions.
- Common example of this in astronomy is a cut performed in a region around the source, so that all events outside such region are ignored.
- Disadvantage of cutting data is that also a fraction of the signal events are excluded
- Point of this topic is to show how by replacing a fixed signal extraction cut with a statistical weighting of the events according to specific information in order to obtain a more precise signal estimation.
- That method is called Bayesian Analysis including Single-event Likelihoods BASiL

PDF of the signal rate including single-event observations

- We start by including the information about the variable **x**

$$p(s\,|\,\vec{\mathbf{x}}, N_{on}, N_{off}; \alpha) \propto \int_0^{\infty} db\; p(\vec{\mathbf{x}}, N_{on}, N_{off}\,|\,s, b; \alpha)$$

$$p(\vec{\mathbf{x}}, N_{on}, N_{off} \mid s, b; \alpha)$$
$$= \boxed{p(\vec{\mathbf{x}} \mid N_{on}, s, \alpha b)} \cdot p(N_{on} \mid s, \alpha b) \cdot p(N_{off} \mid b)$$

- Given that all measured events are independent from each other, the probability is:

$$p(\vec{\mathbf{x}} \mid N_{on}, s, \alpha b)$$
$$= \prod_{i=1}^{N_{on}} \left[\, p(\mathbf{x}_i\,|\,\gamma) \cdot p(\gamma\,|\,s,\, \alpha b) + p(\mathbf{x}_i\,|\,\bar{\gamma}) \cdot p(\bar{\gamma}\,|\,s,\, \alpha b) \,\right]$$

$$p(\gamma\,|\,s,\, \alpha b) = 1 - p(\bar{\gamma}\,|\,s,\, \alpha b) = \frac{s}{s + \alpha b}$$

- This term is the prior probability that one event is a signal event gamma.

$$C(\vec{\mathbf{x}}, N_s) = \sum_{A \in F_{N_s}} \prod_{i \in A} p(\mathbf{x}_i|\gamma) \cdot \prod_{j \in A^c} p(\mathbf{x}_j|\bar{\gamma})$$

$$p(\vec{\mathbf{x}}, N_{on}, N_{off} \mid s, b; \alpha)$$
$$\propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on} + N_{off} - N_s)!}{(N_{on} - N_s)!(1 + 1/\alpha)^{-N_s}} \boxed{\frac{C(\vec{x}, N_s)}{\binom{N_{on}}{N_s}}}$$
$$\times \frac{s^{N_s}}{N_s!} e^{-s} \cdot \frac{(b(1+\alpha))^{N_{on}+N_{off}-N_s}}{(N_{on} + N_{off} - N_s)!} e^{-b(1+\alpha)}$$

Nuisance parameter

- Similar to introduction part the parameter $\boxed{b}$ can be marginalized to obtain PDF of the signal rate s:

$$p(s \mid \vec{\mathbf{x}}, N_{on}, N_{off}; \alpha) \propto \int db\; p(\vec{\mathbf{x}}, N_{on}, N_{off} \mid s, b; \alpha)$$

$$\propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on} + N_{off} - N_s)!}{(N_{on} - N_s)!(1 + 1/\alpha)^{-N_s}} \frac{C(\vec{\mathbf{x}}, N_s)}{\binom{N_{on}}{N_s}} \times \frac{s^{N_s}}{N_s!} e^{-s}$$

- Now we can identify PMF as a:

$$p(N_s \mid \vec{\mathbf{x}}, N_{on}, N_{off}; \alpha)$$
$$\propto \frac{(N_{on} + N_{off} - N_s)!}{(N_{on} - N_s)!(1 + 1/\alpha)^{-N_s}} \frac{C(\vec{\mathbf{x}}, N_s)}{\binom{N_{on}}{N_s}}$$

PDF of the signal rate including single-event observations

- If all events are l time more likely of being a signal event ($p(\mathbf{x}_i|\gamma) = l \cdot p(\mathbf{x}_i|\bar{\gamma}) \quad \forall i \in \{1, \ldots, N_{on}\}$) then using

$$C(\vec{\mathbf{x}}, N_s) \propto \binom{N_{on}}{N_s} l^{N_s} \Longrightarrow \boxed{p(N_s \mid \vec{\mathbf{x}}, N_{on}, N_{off}; \alpha) \propto \frac{(N_{on} + N_{off} - N_s)!}{(N_{on} - N_s)!(1 + 1/\alpha)^{-N_s}} \cdot l^{N_s}}$$

- By taking into account the information that all events are l times more likely of being a signal event we have updated our PMF of the number of signal events introducing last factor.
- The **power** of this method depends on the specifics of the datasets in which it is applied and:
  - the event parameters that are used
  - how they distribute for the signal and background population
  - how performing is the signal extraction method

The case of IACT's

- We will assume case of MAGIC telescopes (theta^2, Hadroness). We consider individual event variables:

$$\mathbf{x} = (\theta^2, h, E)$$

- The likelihoods of being a signal or a background event can be factorized into three terms where I stands for the conditions under which the observation has been performed:

$$p(\mathbf{x}\,|\,\gamma) = p(h\,|\,E, I, \gamma) \cdot p(\theta^2\,|\,E, I, \gamma) \cdot p(E\,|\,I, \gamma),$$
$$p(\mathbf{x}\,|\,\bar{\gamma}) = p(h\,|\,E, I, \bar{\gamma}) \cdot p(\theta^2\,|\,E, I, \bar{\gamma}) \cdot p(E\,|\,I, \bar{\gamma}).$$

- Theta^2 and Hadroness have a small correlation so we can consider these variables as independent
- To get likelihood for each event of being s or b event, computation of distribution of th^2 and h is needed
- In the case of IACT's we have to rely on the MC simulations - How do we simulate it?

The case of IACT's

- Generate N_s and N_{bkg} from Poisson distribution with expected value respectively s and alpha times b and define the number of counts in On region as:

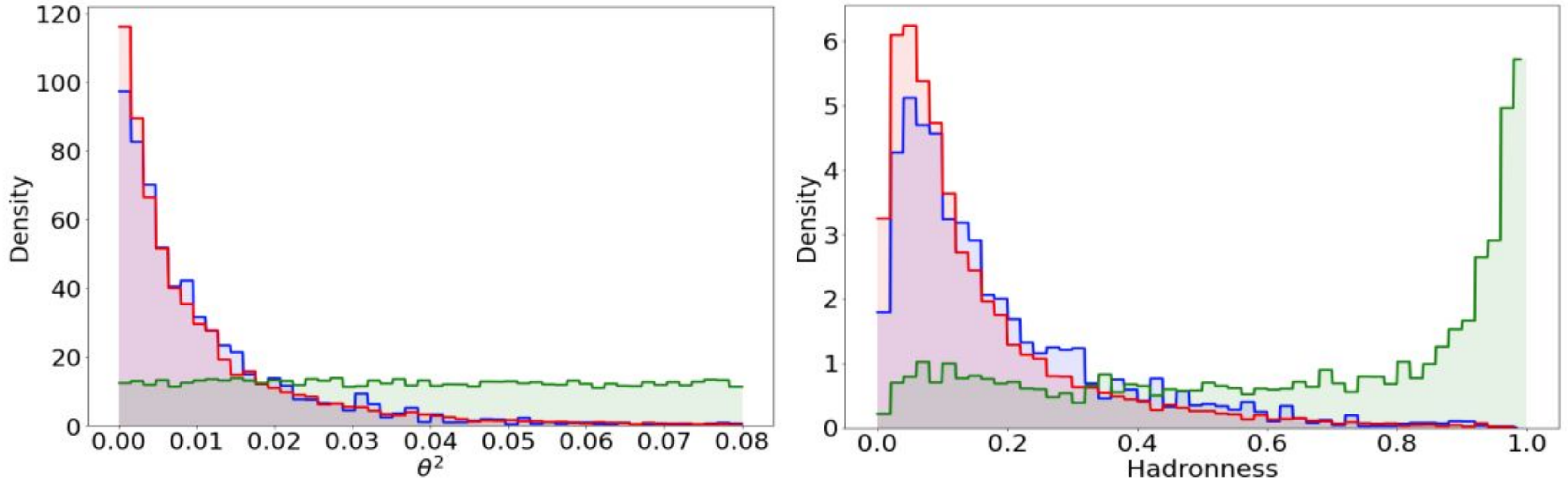
$$N_{on} = N_s + N_{bkg}$$

- Generate N_{off}, the total number of events in the Off region, from a Poisson distribution with expected value b

- Generate th^2 values for the events in the On region by randomly picking up N_s values from the signal distributions and N_{bkg} values from the background distributions

- Finally, the same is done for generating Hadronness values for the On and Off measurements

The case of IACT's



- Simulated (red) and observed (blue) distribution of th^2 and h together with background events (green) normalized to 1

  The observed signal is from observation of **Crab Nebula**

The case of IACT's

- Having an On and Off measurements we get an estimation $\hat{s}$ of the signal rate using only the information about total counts N_{on} and N_{off} and the single event variables $\mathbf{x} = (h, \theta^2)$

1. The estimated signal rate is obtained from

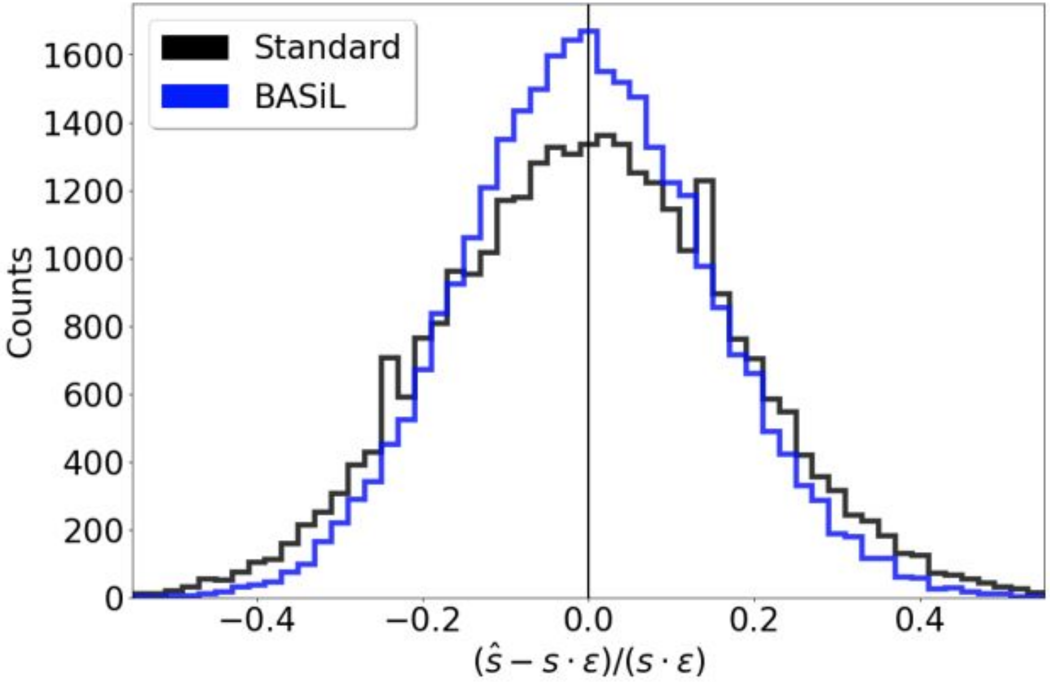
$$\hat{s} = N_{on} - \alpha N_{off}$$

   where N_{on} and N_{off} are the number of events surviving the cut in th^2 and/or hadroness for the On and Off measurement. Being the most common way of estimating s, we can refer to this as "standard" method

2. In the BASiL method $\hat{s}$ is instead defined from the mode of the PMF where **x** can be either th^2 and h or only one of them. The combinatorial term can be obtained using signal and background likelihood values from the signal and background distributions.

$$\boxed{\begin{aligned} prec. &= \sigma\left(\frac{\hat{s} - s \cdot \epsilon}{s \cdot \epsilon}\right), \\ bias &= < \frac{\hat{s} - s \cdot \epsilon}{s \cdot \epsilon} > . \end{aligned}}$$

The case of IACT's

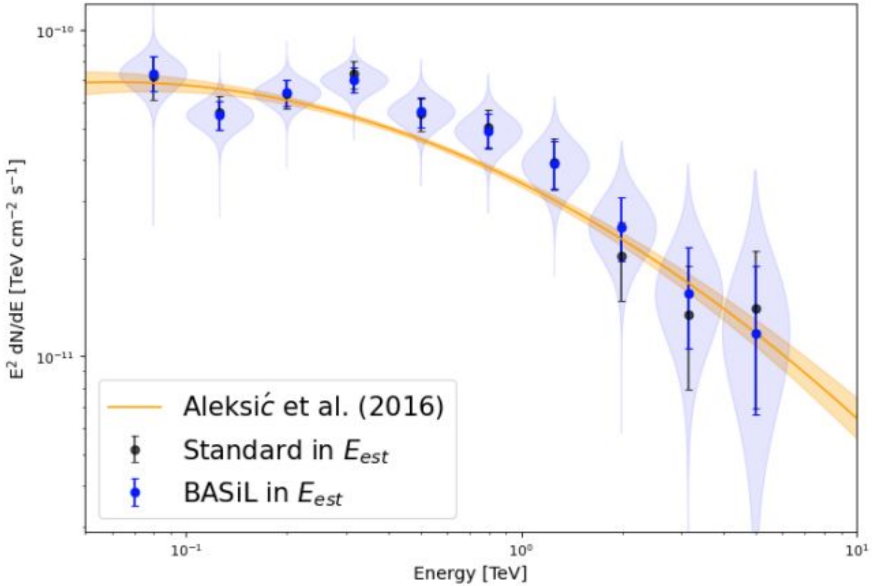


Two simulations of distributions, one for the standard approach in black and one in the BASiL approach in blue.

In standard approach we have efficiency for the h of 95% and for th^2 75% which in total translates to the 67.6%

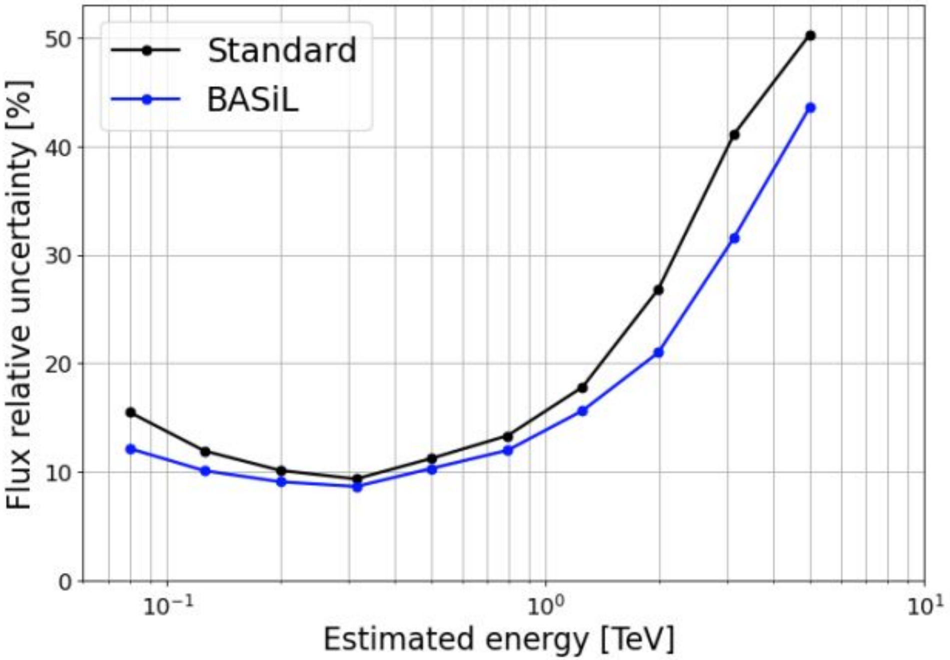
In the BASiL approach efficiency is **1**

The case of IACT's



E [GeV]	N_{on}	N_{off}	Signal (freq.)	Signal (BASiL)
63-100	1714	4494	216 ± 47	204 ± 23
100-158	933	2349	150 ± 35	187 ± 18
158-251	622	1327	180 ± 28	185 ± 16
251-398	439	846	157 ± 23	174 ± 14
398-631	335	593	137 ± 20	114 ± 11
631-1000	215	435	70 ± 16	$77.7^{+9.6}_{-8.9}$
1000-1585	132	256	47 ± 13	$41.8^{+7.0}_{-6.3}$
1585-2512	95	203	27 ± 11	$21.6^{+5.1}_{-4.5}$
2512-3981	56	140	9.3 ± 8.5	$8.6^{+3.4}_{-2.7}$
3981-6310	30	83	2.3 ± 6.3	$3.9^{+2.3}_{-1.6}$

The case of IACT's



$$\boxed{\frac{\sqrt{N_{on} + \alpha^2 N_{off}}}{N_{on} - \alpha N_{off}}}$$

$$\boxed{\frac{(s_{right} - s_{left})/2}{s^*}}$$

Conclusion

- Main feature of Bayesian approach (BASiL) that it weights events according to their individual likelihood of being a signal or background considering all available information
- BASiL method avoids cutting data according to some variable to suppress the background which discards a part of the signal.
- This method, while yielding results consistent with the standard data analysis method, it improves the precision of the signal estimation (last slide)
- The biggest improvement has been seen in cases of small signal rates, due to this feature, BASiL would be useful for analysis of data from measurements of short transients or weak signal.

Thank you for your attention!

References:

1. https://arxiv.org/pdf/2105.01019.pdf
2. T. J. Loredo, in Statistical Challenges in Modern Astronomy (1992), pp. 275–297
3. T. P. Li and Y. Q. Ma, Astrophys. J. 272, 317 (1983)
4. S. Wilks, Annals Math. Statist. 9, 60 (1938)

BACKUP SLIDES

$$
\begin{aligned}
p(\vec{\mathbf{x}}, N_{on}, N_{off} \mid s, b; \alpha) &= p(\vec{\mathbf{x}} \mid N_{on}, s, \alpha b) \cdot p(N_{on} \mid s, \alpha b) \cdot p(N_{off} \mid b) \\
&= \prod_{i=1}^{N_{on}} \left(p(\mathbf{x}_i \mid \gamma) \frac{s}{s + \alpha b} + p(\mathbf{x}_i \mid \bar{\gamma}) \frac{\alpha b}{s + \alpha b} \right) \cdot \frac{(s + \alpha b)^{N_{on}}}{N_{on}!} e^{-s-\alpha b} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b} \\
&= \sum_{N_s=0}^{N_{on}} \sum_{A \in F_{N_s}} \prod_{i \in A} p(\mathbf{x}_i | \gamma) \cdot \prod_{j \in A^c} p(\mathbf{x}_j | \bar{\gamma}) \cdot \frac{s^{N_s} (\alpha b)^{N_{on} - N_s}}{(s + \alpha b)^{N_{on}}} \cdot \frac{(s + \alpha b)^{N_{on}}}{N_{on}!} e^{-s-\alpha b} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b} \\
&= \sum_{N_s=0}^{N_{on}} C(\vec{\mathbf{x}}, N_s) \cdot \frac{s^{N_s} (\alpha b)^{N_{on} - N_s}}{N_{on}!} e^{-s-\alpha b} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b} \\
&= \frac{\alpha^{N_{on}} / N_{off}!}{(1 + \alpha)^{N_{on} + N_{off}}} \sum_{N_s=0}^{N_{on}} \frac{(N_{on} + N_{off} - N_s)!}{(1 + 1/\alpha)^{-N_s}} \frac{C(\vec{\mathbf{x}}, N_s)}{N_{on}!/N_s!} \cdot \frac{s^{N_s}}{N_s!} e^{-s} \cdot \frac{(b(1 + \alpha))^{N_{on} + N_{off} - N_s}}{(N_{on} + N_{off} - N_s)!} e^{-b(1+\alpha)} \\
&\propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on} + N_{off} - N_s)!}{(N_{on} - N_s)!(1 + 1/\alpha)^{-N_s}} \frac{C(\vec{\mathbf{x}}, N_s)}{\binom{N_{on}}{N_s}} \cdot \frac{s^{N_s}}{N_s!} e^{-s} \cdot \frac{(b(1 + \alpha))^{N_{on} + N_{off} - N_s}}{(N_{on} + N_{off} - N_s)!} e^{-b(1+\alpha)} \, .
\end{aligned}
$$

BACKUP SLIDES

$$
\begin{aligned}
C(\vec{x},0) =& p(x_1|\bar{\gamma}) \cdot p(x_2|\bar{\gamma}) \cdot p(x_3|\bar{\gamma}), \\
C(\vec{x},1) =& p(x_1|\gamma) \cdot p(x_2|\bar{\gamma}) \cdot p(x_3|\bar{\gamma}) \\
& + p(x_1|\bar{\gamma}) \cdot p(x_2|\gamma) \cdot p(x_3|\bar{\gamma}) \\
& + p(x_1|\bar{\gamma}) \cdot p(x_2|\bar{\gamma}) \cdot p(x_3|\gamma), \\
C(\vec{x},2) =& p(x_1|\gamma) \cdot p(x_2|\gamma) \cdot p(x_3|\bar{\gamma}) \\
& + p(x_1|\gamma) \cdot p(x_2|\bar{\gamma}) \cdot p(x_3|\gamma) \\
& + p(x_1|\bar{\gamma}) \cdot p(x_2|\gamma) \cdot p(x_3|\gamma), \\
C(\vec{x},3) =& p(x_1|\gamma) \cdot p(x_2|\gamma) \cdot p(x_3|\gamma).
\end{aligned}
$$