The SkyLLH framework for IceCube point-source search

The IceCube Collaboration
(a complete list of authors can be found at the end of the proceedings)
E-mail: tomas.kontrimas@icecube.wisc.edu

Hypothesis tests based on unbinned log-likelihood (LLH) functions are a common technique used in multi-messenger astronomy, including IceCube’s neutrino point-source searches. We present the general Python-based tool "SkyLLH", which provides a modular framework for implementing and executing log-likelihood functions to perform data analyses with multi-messenger astronomy data. Specific SkyLLH framework features for a new and improved time-integrated IceCube point-source analysis are highlighted, including the support for kernel density estimation (KDE) based probability density functions. In addition, the support for a variety of point-source analysis types, such as stacked and time-variable searches, will be presented.

Corresponding authors: Tomas Kontrimas\textsuperscript{1*}, Martin Wolf\textsuperscript{1}
\textsuperscript{1} Physik-department, Technische Universität München, D-85748 Garching, Germany

\textsuperscript{*} Presenter

37\textsuperscript{th} International Cosmic Ray Conference (ICRC 2021)
July 12th – 23rd, 2021
Online – Berlin, Germany
1. Introduction

The unbinned likelihood formalism is used widely in multi-messenger and neutrino astronomy as a statistical test \cite{1}. For testing two complementary hypotheses, the likelihood ratio test is expressed as a ratio of likelihoods given each hypothesis parameters $\tilde{\theta}$ from an entire parameter space $\Theta$

$$\lambda(\tilde{D}) = \frac{\sup_{\tilde{\theta} \in \Theta_0} L(\tilde{\theta}|\tilde{D})}{\sup_{\tilde{\theta} \in \Theta_1} L(\tilde{\theta}|\tilde{D})},$$

where hypotheses $H_0 : \tilde{\theta} \in \Theta_0$ and $H_1 : \tilde{\theta} \in \Theta_0^c$ are called null and alternative hypotheses, respectively. The resulting $\lambda(\tilde{D})$ value indicates whether the given data is compatible with the null hypothesis or not.

The unbinned likelihood function is generally expressed as $L = \prod_i p(\tilde{x}_i)$, with assigned probability $p(\tilde{x}_i)$ to each observed event $\tilde{x}_i$, where the total likelihood value is given by a product. In IceCube neutrino point-source searches, the probability of observed event is written as a two-component likelihood function

$$L(n_s, \tilde{p}_s|D) = \prod_{i=1}^{N} \left[ \frac{n_s}{N} S_i (\tilde{p}_s) + \left( 1 - \frac{n_s}{N} \right) B_i \right],$$

where $n_s$ is the number of signal events in the data sample $D$ of $N$ total events. The set of source model parameters is denoted as $\tilde{p}_s$ and usually contains source position $\tilde{d}_{\text{src}} = (\alpha_{\text{src}}, \delta_{\text{src}})$ and spectral index $\gamma$. $S_i$ and $B_i$ are the value of the signal and background probability density function (PDF) for the $i$th data event, respectively. Because the number of signal events in the data sample is usually small, the background and signal PDFs can be built either from scrambled experimental or Monte-Carlo simulation data.

The log likelihood ratio test statistic is expressed as

$$TS = -2 \log \lambda(n_s, \tilde{p}_s|D) = -2 \sup_{n_s, \tilde{p}_s} \left\{ \log \left[ \frac{L(n_s = 0|D)}{L(n_s, \tilde{p}_s|D)} \right] \right\},$$

where the null hypothesis is defined as an observation of zero signal events and the alternative hypothesis — observation of $n_s > 0$. The logarithm is used outside of the likelihood ratio for the numerical stability. In the limit of high statistics, the Wilk's approximation \cite{2} can be used to approximate it. The resulting $TS$ distribution follows a $k$ number of degree of freedom $\chi^2_k$ distribution when hypotheses parameters are sufficiently away from boundaries.

2. The SkyLLH Framework

The SkyLLH framework is an open-source Python3-based tool licensed under the GPLv3 license\footnote{https://www.gnu.org/licenses/gpl-3.0.txt}. It is available on the open-source IceCube Neutrino Observatory GitHub repository\footnote{https://github.com/icecube/skylhh} and provides a modular framework for implementing custom likelihood functions and executing log-likelihood ratio hypothesis tests. Because the framework is designed to be detector independent,
it is easy to extend it and perform analyses with combined multi-messenger data from IceCube, Fermi, ANTARES, KM3NeT, etc. observatories.

Such framework modularity was achieved by defining common interfaces for mathematical components of the log-likelihood function utilizing object-oriented-programming (OOP) techniques. The classes structure is tied to the mathematical objects of the likelihood (ratio) function, for example, an Analysis object can be naturally built from PDFs → PDFRatios → LLHRatio objects. This allows to define a specific analysis by choosing from already provided components or extending them with custom properties, if needed. The specific complementary classes for the IceCube Neutrino Observatory detector are provided in the i3 module and a private IceCube github repository³ containing pre-defined common analyses and datasets.

The main framework code is split into four modules. The core module holds classes defining the detector independent mathematical log-likelihood framework by utilizing Python’s abstract base class (ABC) module. As mentioned earlier, IceCube specific classes derived from the core module are in the i3 module. They extend signal and background events generators, provide detector signal yield calculation, models of PDFs and PDFRatios, and common coordinate transformations. The physics module contains definitions of source hypothesis and its flux models. The plotting module contains utility functions for plotting the generated analysis objects (PDFRatios and PDFs) and the calculated trial data.

The analysis definition process in SkyLLH is realized by creating an Analysis object with desired properties. For the users’ convenience, the Analysis object can be simply created by running a function from a list of already predefined create_analysis functions of commonly used analysis types, or they can serve as a base for a custom analysis definition. The important create_analysis function parameters are datasets, minimizer_impl, source, reference power-law flux model normalization, energy, and spectral index.

The datasets parameter contains a list of Dataset instances. SkyLLH has pre-defined functions for creating DatasetCollection of datasets, which are currently used in most of the ongoing IceCube analyses. The desired dataset collection, containing a list of the same version datasets, can simply be loaded from the provided data_samples dictionary. Predefined datasets can be easily customized by adding additional experimental, Monte-Carlo, and (or) auxiliary data definitions. In addition to the data definitions, the Dataset object can contain data preparation functions, which modify data after it is loaded, and data field renaming dictionaries for experimental and Monte-Carlo data to standardize the naming scheme between different dataset collections.

The likelihood developer can either define a custom minimizer implementation or choose it from provided Scipy⁴, Newton-Raphson, and iminuit⁵ implementations. The source is defined by e.g. a PointLikeSource instance at a given location in the sky. If the likelihood developer wants to change the source position after the Analysis object is created, the Analysis class has a change_source method, which applies necessary changes to all the source dependent objects of the Analysis. For the source flux model the PowerLawFlux is created with given flux normalization, energy, and spectral index values.

³https://github.com/icecube/i3skyllh
⁴https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html
3. Kernel Density Estimator for A New Search for Neutrino Point Sources with IceCube analysis

Using the SkyLLH framework a new point-source analysis with the IceCube data was developed. IceCube is a cubic-kilometer neutrino detector installed in the ice at the geographic South Pole [3] between depths of 1450 m and 2450 m, completed in 2010. Reconstruction of the direction, energy and flavor of the neutrinos relies on the optical detection of Cherenkov radiation emitted by charged particles produced in the interactions of neutrinos in the surrounding ice or the nearby bedrock. The new search for neutrino point-sources improves the accuracy of the statistical analysis, especially in the low energy regime [4], in comparison to previous analyses. It has a new likelihood description based on kernel density estimation (KDE) method, which allows to replace analytical expressions by non-parametrically inferring probability density functions.

The multi-dimensional KDE based PDFs are defined as

$$P_{\text{KDE}}(\vec{x}) = \frac{1}{Nh} \sum_{i=1}^{N} K \left( \frac{\vec{x} - \vec{x}_i}{h} \right),$$

(4)

where $h$ is a smoothing parameter called bandwidth, $K(x)$ is a kernel function and $\vec{x}_i$ is $i$th event in a data sample. The new point-source analysis replaces signal and background PDF terms in the likelihood expression (equation 2) with

$$f_S(\hat{E}_\mu, \hat{\delta}, \hat{\sigma}) \approx \frac{1}{2\pi \sin \psi} f_S(\hat{\psi} | \hat{\sigma}, \hat{E}_\mu, \gamma) \cdot f_S(\hat{E}_\mu | \delta_{src}, \gamma),$$

(5)

$$f_B(\hat{d}_t, \hat{\sigma}_t, \hat{E}_\mu, i) = \frac{1}{2\pi} f_B(\sin \hat{\delta}_t, \hat{\sigma}_t, \hat{E}_\mu, i),$$

(6)

respectively. It uses three observables: the estimated muon energy $\hat{E}_\mu$, the reconstructed muon direction $\hat{d}$ and its estimated uncertainty $\hat{\sigma}$, and angular distance $\hat{\psi} = ||\hat{d}_t - \vec{d}_{src}||$ between reconstructed direction and source position. The new likelihood no longer relies on the analytical Gaussian approximation of signal PDF’s spatial term, which was used in previous IceCube analyses [1].

This likelihood construction requires generation of KDE based signal PDFs on a power-law flux index $\gamma$ grid. They are generated from the Monte-Carlo data using an internally developed KDE_tool based on Meerkat [5] package, which uses cross validation technique in order to find an optimal KDE bandwidth $h$ on a coarse $\gamma$ grid. After the optimal bandwidths are found, KDEs are generated on a fine $\gamma$ grid using interpolated optimal bandwidth values. Because the KDE evaluation is slow, they are evaluated only once on a very fine parameter grid, the result is interpolated and saved as penalized B-splines using the photospline package.

In order to use the new likelihood function in SkyLLH, the framework was extended to provide linear and parabola grid manifold interpolation methods. They selectively load and interpolate penalized B-splines in-between $\gamma$ grid values while minimizing the log-likelihood function. The photosplines are defined as auxiliary Dataset object properties.

*https://github.com/icecube/photospline
4. Stacking

In general, a likelihood ratio test can be done not only for a single point-source, but also for a set of \( K \) stacked sources in a weighted fashion. This can be viewed as an extension to the single point-source search method. The sources must be weighted according to their signal detection efficiency \( Y_{s,k} \), and a relative strength weight of the \( k \)th source \( W_k \), with \( \sum_{k=1}^{K} W_k = 1 \). The \( W_k \) is a theoretical weight independent of position and spectral index of the source that accounts for different properties of the individual sources. The combined signal PDF is then given as

\[
S_i(\bar{\phi}_s) \equiv \frac{\sum_{k=1}^{K} W_k Y_s(\bar{x}_{s_k}, \bar{\phi}_{s_k}) S_i(\bar{\phi}_{s_k})}{\sum_{k=1}^{K} W_k Y_s(\bar{x}_{s_k}, \bar{\phi}_{s_k})}.
\] (7)

In SkyLLH the source hypothesis group class provides a data container to describe a group of sources that share the same flux model, detector signal yield, and signal generation implementation methods. It also supports a definition of relatively weighted multiple sources. The stacking analysis can be constructed by setting up a \texttt{TimeIntegratedMultiDatasetMultiSourceAnalysis} object.

5. Summary and Outlook

The SkyLLH framework is being developed within the IceCube collaboration as a standard tool to search for neutrino emitting sources in the Universe. The implementation of generalized concepts in terms of source hypothesis and hypothesis parameter definition makes it easy to use the SkyLLH framework also for searches of other messenger particles in other experiments. Whenever a likelihood ratio test as given in equation 1 with celestial data has to be performed, SkyLLH is a suitable tool. Possible future applications of SkyLLH could be combined analyses of same-kind messenger particle data, for instance from different neutrino telescopes like IceCube and ANTARES/KM3Net, or of different messenger particle data of neutrinos and gamma-rays, for instance from IceCube and Fermi/LAT.

References


The SkyLLH framework for IceCube point-source search

16 Dept. of Physics and Institute for Global Prominent Research, Chiba University, Chiba 263-8522, Japan
17 Department of Physics, Loyola University Chicago, Chicago, IL 60660, USA
18 Dept. of Physics and Astronomy, University of Canterbury, Private Bag 4800, Christchurch, New Zealand
19 Dept. of Physics, University of Maryland, College Park, MD 20742, USA
20 Dept. of Astronomy, Ohio State University, Columbus, OH 43210, USA
21 Dept. of Physics and Center for Cosmology and Astro-Particle Physics, Ohio State University, Columbus, OH 43210, USA
22 Niels Bohr Institute, University of Copenhagen, DK-2100 Copenhagen, Denmark
23 Dept. of Physics, TU Dortmund University, D-44221 Dortmund, Germany
24 Dept. of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA
25 Dept. of Physics, University of Alberta, Edmonton, Alberta, Canada T6G 2E1
26 Erlangen Centre for Astroparticle Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg, D-91058 Erlangen, Germany
27 Physik-department, Technische Universität München, D-85748 Garching, Germany
28 Département de physique nucléaire et corpusculaire, Université de Genève, CH-1211 Genève, Switzerland
29 Dept. of Physics and Astronomy, University of Gent, B-9000 Gent, Belgium
30 Dept. of Physics and Astronomy, University of California, Irvine, CA 92697, USA
31 Karlsruhe Institute of Technology, Institute for Astroparticle Physics, D-76021 Karlsruhe, Germany
32 Karlsruhe Institute of Technology, Institute of Experimental Particle Physics, D-76021 Karlsruhe, Germany
33 Dept. of Physics, Engineering Physics, and Astronomy, Queen’s University, Kingston, ON K7L 3N6, Canada
34 Dept. of Physics and Astronomy, University of Kansas, Lawrence, KS 66045, USA
35 Department of Physics and Astronomy, UCLA, Los Angeles, CA 90095, USA
36 Dept. of Physics, Mercer University, Macon, GA 31207-0001, USA
37 Dept. of Astronomy, University of Wisconsin–Madison, Madison, WI 53706, USA
38 Dept. of Physics and Wisconsin IceCube Particle Astrophysics Center, University of Wisconsin–Madison, Madison, WI 53706, USA
39 Institute of Physics, University of Mainz, Staudinger Weg 7, D-55099 Mainz, Germany
40 Department of Physics, Marquette University, Milwaukee, WI 53201, USA
41 Institut für Kernphysik, Westfälische Wilhelms-Universität Münster, D-48149 Münster, Germany
42 Bartol Research Institute and Dept. of Physics and Astronomy, University of Delaware, Newark, DE 19716, USA
43 Dept. of Physics, Yale University, New Haven, CT 06520, USA
44 Dept. of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK
45 Dept. of Physics, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA
46 Physics Department, South Dakota School of Mines and Technology, Rapid City, SD 57701, USA
47 Dept. of Physics, University of Wisconsin, River Falls, WI 54022, USA
48 Dept. of Physics and Astronomy, University of Rochester, Rochester, NY 14627, USA
49 Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA
50 Oskar Klein Centre and Dept. of Physics, Stockholm University, SE-10691 Stockholm, Sweden
51 Dept. of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794-3800, USA
52 Dept. of Physics, Sungkyunkwan University, Suwon 16419, Korea
53 Institute of Basic Science, Sungkyunkwan University, Suwon 16419, Korea
54 Dept. of Physics and Astronomy, University of Alabama, Tuscaloosa, AL 35487, USA
55 Dept. of Astronomy and Astrophysics, Pennsylvania State University, University Park, PA 16802, USA
56 Dept. of Physics, Pennsylvania State University, University Park, PA 16802, USA
57 Dept. of Physics and Astronomy, Uppsala University, Box 516, S-75120 Uppsala, Sweden
58 Dept. of Physics, University of Wuppertal, D-42119 Wuppertal, Germany
59 DESY, D-15738 Zeuthen, Germany
60 Università di Padova, I-35131 Padova, Italy
61 National Research Nuclear University, Moscow Engineering Physics Institute (MEPhI), Moscow 115409, Russia
62 Earthquake Research Institute, University of Tokyo, Bunkyo, Tokyo 113-0032, Japan

Acknowledgements

USA – U.S. National Science Foundation-Office of Polar Programs, U.S. National Science Foundation-Physics Division, U.S. National Science Foundation-EPSCoR, Wisconsin Alumni Research Foundation, Center for High Throughput Computing (CHTC) at the University of Wisconsin–Madison, Open Science Grid (OSG), Extreme Science and Engineering Discovery Environment (XSEDE), Frontera computing project at the Texas Advanced Computing Center, U.S. Department of Energy-National Energy Research Scientific Computing Center, Particle astrophysics research computing center at the University of Maryland, Institute for Cyber-Enabled Research at Michigan State University, and Astroparticle physics computational facility at Marquette University; Belgium – Funds for Scientific Research (FRS-FNRS and FWO), FWO Odysseus and Big Science programmes, and Belgian Federal Science Policy Office (Belspo); Germany – Bundesministerium für Bildung und Forschung (BMBF), Deutsche Forschungsgemeinschaft (DFG), Helmholtz Alliance for Astroparticle Physics (HAP), Initiative and Networking Fund of the Helmholtz Association, Deutsches Elektronen Synchrotron (DESY), and High Performance Computing cluster of the RWTH Aachen; Sweden – Swedish Research Council, Swedish Polar Research Secretariat, Swedish National Infrastructure for Computing (SNIC), and Knut and Alice Wallenberg Foundation; Australia – Australian
The SkyLLH framework for IceCube point-source search

Research Council; Canada – Natural Sciences and Engineering Research Council of Canada, Calcul Québec, Compute Ontario, Canada Foundation for Innovation, WestGrid, and Compute Canada; Denmark – Villum Fonden and Carlsberg Foundation; New Zealand – Marsden Fund; Japan – Japan Society for Promotion of Science (JSPS) and Institute for Global Prominent Research (IGPR) of Chiba University; Korea – National Research Foundation of Korea (NRF); Switzerland – Swiss National Science Foundation (SNSF); United Kingdom – Department of Physics, University of Oxford.