1 The Likelihood Formalism

This section describes the mathematical likelihood formalism used in Skyllh. First it introduces the log-likelihood approach, second the likelihood ratio test and the used test statistic and then describes the used optimizations.

1.1 The Log-Likelihood Approach

SkyLLH implements the two-component likelihood approach with a likelihood function $\mathcal{L}(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}}\ |D)$ of the form

$$\mathcal{L}(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}}\ |D) = \prod_{i=1}^{N} \left[\frac{n_{\mathrm{s}}}{N} S_i(\vec{p}_{\mathrm{s}}) + (1 - \frac{n_{\mathrm{s}}}{N}) B_i \right], \tag{1}$$

where n_{s} is the number of signal events, hence, $(N - n_{\mathrm{s}})$ the number of background events in the data sample D of N total events. The set of source model parameters is denoted as $\vec{p}_{\mathrm{s}}$. For a point-like source model, the source model parameters include the source position $\vec{x}_{\mathrm{s}}$ and the spectral index γ of the source flux. $S_i(\vec{p}_{\mathrm{s}})$ and B_i is the value of the signal and background PDF for the ith data event, respectively.

The signal and background PDFs must incorperate the detector efficiency (yield), $\mathcal{Y}_i$, which, in general, depends on the celestrial direction, the energy, and the observation time of the data event.

For computational stability reasons the logarithm of the likelihood function of equation 1 is used in SkyLLH:

$$\log \mathcal{L}(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}}\ |D) = \sum_{i=1}^{N} \log(...) \tag{2}$$

1.2 Likelihood Ratio Test

For estimating the significance of an observation, the likelihood ratio Λ with respect to a null hypothesis of no observation, i.e. equation 1 at $n_{\mathrm{s}} = 0$ is tested:

$$\log \Lambda(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}}) = \log \frac{L(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{L(n_{\mathrm{s}} = 0)} = \sum_{i=1}^{N} \log \left[1 + \frac{n_{\mathrm{s}}}{N} \left(\frac{S_i(\vec{p}_{\mathrm{s}})}{B_i} - 1 \right) \right] \tag{3}$$

By defining

$$\mathcal{X}_i(\vec{p}_{\mathrm{s}}) \equiv \frac{1}{N} \left(\mathcal{R}_i(\vec{p}_{\mathrm{s}}) - 1 \right), \tag{4}$$

with the signal over background PDF ratio value, $\mathcal{R}_i(\vec{p}_{\mathrm{s}})$, of the ith event,

$$\mathcal{R}_i(\vec{p}_{\mathrm{s}}) \equiv \frac{S_i(\vec{p}_{\mathrm{s}})}{B_i}, \tag{5}$$

this reads as:

$$\log\Lambda(n_{\mathrm{s}},\vec{p}_{\mathrm{s}}) = \sum_{i=1}^{N}\log(1+n_{\mathrm{s}}\mathcal{X}_i(\vec{p}_{\mathrm{s}})). \tag{6}$$

By defining

$$\alpha_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}}) \equiv n_{\mathrm{s}}\mathcal{X}_i(\vec{p}_{\mathrm{s}}) \tag{7}$$

the log-likelihood ratio function of the ith event can be defined as

$$\log\Lambda_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}}) \equiv \log(1+\alpha_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}})), \tag{8}$$

and the log-likelihood ratio function for all events can be written as

$$\log\Lambda(n_{\mathrm{s}},\vec{p}_{\mathrm{s}}) = \sum_{i=1}^{N}\log\Lambda_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}}). \tag{9}$$

In general, the argument of the log-function, Λ_i, might become close to zero, causing a divergence of the log-likelihood ratio function for a particular event. To circumvent this, a Taylor expension of the log-likelihood ratio function of the ith event can be performed around a pre-defined threshold value α. The event-based log-likelihood ratio function, $\log\Lambda_i$, is then approximated by a second-order Taylor expension for events with $\alpha_i \leq \alpha$:

$$\log\Lambda_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}}) \equiv \log(1+\alpha) + \frac{\alpha_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}})-\alpha}{1+\alpha} - \frac{1}{2}\left(\frac{\alpha_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}})-\alpha}{1+\alpha}\right)^2 \tag{10}$$

By defining

$$\tilde{\alpha}_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}}) \equiv \frac{\alpha_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}})-\alpha}{1+\alpha}, \tag{11}$$

the Taylor expanded log-likelihood ratio function reads more compactly:

$$\log\Lambda_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}}) = \log(1+\alpha) + \tilde{\alpha}_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}}) - \frac{1}{2}\tilde{\alpha}_i^2(n_{\mathrm{s}},\vec{p}_{\mathrm{s}}). \tag{12}$$

1.3 Test Statistic

Assuming Wilks theorem, a test statistic, TS, for the two-component log-likelihood ratio test can be formulated using the log-likelihood ratio function, $\log\Lambda$, at its maximum:

$$\mathrm{TS} = 2\,\mathrm{sgn}(\hat{n}_{\mathrm{s}})\log\Lambda(\hat{n}_{\mathrm{s}},\hat{\vec{p}}_{\mathrm{s}}), \tag{13}$$

where $\log\Lambda(\hat{n}_{\mathrm{s}},\hat{\vec{p}}_{\mathrm{s}})$ is the maximum of the log-likelihood ratio function as defined by equation (3), with separation of an over- ($\hat{n}_{\mathrm{s}} > 0$) and under-fluctuation ($\hat{n}_{\mathrm{s}} < 0$). In case the assumptions of Wilks theorem are met within the analysis, the test statistic value distribution will follow a χ^2-distribution with a degree-of-freedom equal to the number of fit parameters.

For the case $\hat{n}_{\mathrm{s}} = 0$, the log-likelihood ratio function is zero with degenerate source fit parameter values $\hat{\vec{p}}_{\mathrm{s}}$. When calculating the sensitivity of the analysis, the median test-statistic value for background-only data is required. Hence, $\hat{n}_{\mathrm{s}}$ is often zero in such cases. Having to deal with a delta-peak of the test-statistic distribution for TS $= 0$ is cumbersome. In order to resolve the delta-peak, the log-likelihood ratio function can be approximated with a second-order Taylor expansion around $n_{\mathrm{s}} = 0$, and the apex of that Taylor function defines the value of log-likelihood ratio function for $\hat{n}_{\mathrm{s}} = 0$. In this case the test-statistic function is given by

$$\mathrm{TS} = -2\frac{\left(\frac{\mathrm{d}\log\Lambda(n_{\mathrm{s}}=0,\vec{p}_{\mathrm{s}}=\vec{\hat{p}}_{\mathrm{s}})}{\mathrm{d}n_{\mathrm{s}}}\right)^2}{4\frac{\mathrm{d}^2\log\Lambda(n_{\mathrm{s}}=0,\vec{p}_{\mathrm{s}}=\vec{\hat{p}}_{\mathrm{s}})}{\mathrm{d}n_{\mathrm{s}}^2}}. \tag{14}$$

1.4 Optimizations for Spatially Restricted Sources

For spatially restricted sources, e.g. point-like sources, most of the events in the data sample will be far away from the hypothesised source, hence, the value of the signal PDF, S_i, will be zero or very close to zero. By selecting only the signal-contributing N' events from the sample, the log-likelihood ratio function, $\log\Lambda$, reads

$$\log\Lambda(n_{\mathrm{s}},\vec{p}_{\mathrm{s}}) = \sum_{i=1}^{N'}\log\Lambda_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}}) + (N-N')\log(1-\frac{n_{\mathrm{s}}}{N}), \tag{15}$$

where for $N-N'$ events $\mathcal{R}_i(\vec{p}_{\mathrm{s}})$ equals zero and hence $\alpha_i(n_{\mathrm{s}},\vec{p}_{\mathrm{s}})$ becomes $-n_{\mathrm{s}}/N$, and $\log\Lambda_i$ equals $\log(1-n_{\mathrm{s}}/N)$ for all such pure background events.

1.5 Signal PDFs

The likelihood ratio function as given in equation (3) incorperates a signal probability, S_i, for an individual event i. Without loss of generality this signal probability can be expressed as the product of a spatial-energy, $\mathcal{P}_{\mathcal{SE}}(\vec{x}_i, E_i|\vec{p}_{\mathrm{s}})$, and a time, $\mathcal{T}_{\mathrm{S}}(t_i|\vec{p}_{\mathrm{s}})$, probability, where $\vec{x}_i$, E_i, and t_i are reconstructed observables of event i. In general the spatial-energy probability is the joint probability of the spatial, $\mathcal{S}_{\mathrm{S}}$, and energy, $\mathcal{E}_{\mathrm{S}}$, signal probabilities:

$$\mathcal{P}_{\mathcal{SE}}(\vec{x}_i, E_i|\vec{p}_{\mathrm{s}}) = \mathcal{S}_{\mathrm{S}}(\vec{x}_i|\vec{p}_{\mathrm{s}})\mathcal{E}_{\mathrm{S}}(E_i|\vec{x}_i,\vec{p}_{\mathrm{s}}). \tag{16}$$

Here, we assume that the energy PDF depends on the reconstructed direction of the recorded event, but the reconstructed direction is independent of the event's energy.

The final event's signal probability can be written as

$$S_i(\vec{p}_{\mathrm{s}}) = \mathcal{S}_{\mathrm{S}}(\vec{x}_i|\vec{p}_{\mathrm{s,spatial}})\mathcal{E}_{\mathrm{S}}(E_i|\vec{x}_i,\vec{p}_{\mathrm{s,energy}})\mathcal{T}_{\mathrm{S}}(t_i|\vec{p}_{\mathrm{s,time}}), \tag{17}$$

where the signal model parameters $\vec{p}_{\mathrm{s}}$ can be divided into spatial, energy, and time parameters, i.e. $\vec{p}_{\mathrm{s}} = (\vec{p}_{\mathrm{s,spatial}},\vec{p}_{\mathrm{s,energy}},\vec{p}_{\mathrm{s,time}})$. The spatial component, $\mathcal{S}_{\mathrm{S}}$, can be identified as the convolution, $(\Psi * \mathrm{PSF})(\alpha,\delta)$[1], of the spatial

[1] The 2D convolution on the sky is defined as $(f*g)(\alpha,\delta) = \int_0^{2\pi}\mathrm{d}\alpha'\int_{-\pi}^{\pi}\mathrm{d}\delta' f(\alpha',\delta')g(\alpha-\alpha',\delta-\delta')$.

source extension, $\Psi(\alpha, \delta)$, and the point-spread-function, $\mathrm{PSF}(\alpha, \delta)$, of the detector. For a point-like spatial source extension at position $\vec{x}_{\mathrm{s}} = (\alpha_{\mathrm{s}}, \delta_{\mathrm{s}})$, that is $\Psi(\alpha, \delta) = \delta(\alpha - \alpha_{\mathrm{s}})\delta(\delta - \delta_{\mathrm{s}})$, where $\delta(\cdot)$ is the delta-distribution, this convolution collapses to a single point in the sky. With a 2D gaussian PSF $\mathcal{S}_{\mathrm{S}}(\vec{x}_i|\vec{p}_{\mathrm{s,spatial}})$ is given as

$$\mathcal{S}_{\mathrm{S}}(\vec{x}_i|\vec{p}_{\mathrm{s,spatial}}) \equiv \mathcal{S}_{\mathrm{S}}(r_i, \sigma_i|\vec{x}_{\mathrm{s}}) = \frac{1}{2\pi\sigma_i^2} \exp\left(-\frac{r_i^2}{2\sigma_i^2}\right), \tag{18}$$

where r_i is the space angle between the source position and the recorded reconstructed event direction. In equatorial coordinates, $\vec{x} = (\alpha, \delta)$, the cosine of r_i is given by

$$\cos(r_i) = \cos(\alpha_{\mathrm{s}} - \alpha_i)\cos(\delta_{\mathrm{s}})\cos(\delta_i) + \sin(\delta_{\mathrm{s}})\sin(\delta_i). \tag{19}$$

The data quantity σ_i describes the angular reconstruction uncertainty of the event, hence the PSF is narrower for well-reconstructed events, and wider for events which have a large reconstruction uncertainty.

The energy signal PDF can be constructed from monte-carlo data using the assumed source energy spectrum. When considering a power law as source flux model, the energy source parameters, $\vec{p}_{\mathrm{s,energy}}$, consists of the spectral index γ and possibly an energy cut-off parameter E_{cut}.

The source time PDF, $T_{\mathrm{S}}(t|\vec{p}_{\mathrm{s,time}})$, describes the emmission time profile of the source. Different functional forms of this time profile could be imagined. Common profiles are: A steady source profile over the entire observation (live) time, T_{obs},

$$T_{\mathrm{S}}(t) \equiv \frac{1}{T_{\mathrm{obs}}}, \tag{20}$$

with

$$T_{\mathrm{obs}} = \int T_{\mathrm{live}}(t)\mathrm{d}t, \tag{21}$$

where $T_{\mathrm{live}}(t)$ is the detector live-time function, defined as:

$$T_{\mathrm{live}}(t) = \begin{cases} 1 & \forall t \in \text{detector on-time window} \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

Other functional forms could be a box profile of length T_{W} with the box's middle time position, T_0,

$$T_{\mathrm{S}}(t|T_0, T_{\mathrm{W}}) = \begin{cases} \frac{1}{T_{\mathrm{W}}} & \forall t \in [T_0 - T_{\mathrm{W}}/2; T_0 + T_{\mathrm{W}}/2] \\ 0 & \text{otherwise} \end{cases}, \tag{23}$$

or a gaussian shaped time profile centered at T_0 with a time width of σ_T,

$$T_{\mathrm{S}}(t|T_0, \sigma_T) \equiv \frac{1}{\sqrt{2\pi}\sigma_T} \exp\left(-\frac{(t - T_0)^2}{2\sigma_T^2}\right). \tag{24}$$

For efficiency reasons, the gaussian shape source time profile is truncated at a certain distance from T_0, *e.g.* at $\pm\sigma_T$:

$$T_{\mathrm{S}}(t|T_0,\sigma_T) \equiv \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_T}\exp\left(-\frac{(t-T_0)^2}{2\sigma_T^2}\right) & |t-T_0| \leq \sigma_T \\ 0 & \text{otherwise} \end{cases}. \tag{25}$$

The final time PDF, $\mathcal{T}_{\mathrm{S}}(t|\vec{p}_{\mathrm{s,time}})$, is then the convolution of the source time profile, $T_{\mathrm{S}}(t|\vec{p}_{\mathrm{s,time}})$, and the detector live-time function, $T_{\mathrm{live}}(t)$, where the final result is normalized to unity again.

1.6 Background PDFs

In analog to the signal PDF, the background PDF can be formulated as

$$B_i \equiv \mathcal{S}_{\mathrm{B}}(\vec{x}_i)\mathcal{E}_{\mathrm{B}}(E_i|\vec{x}_i)\mathcal{T}_{\mathrm{B}}(t_i). \tag{26}$$

All the background PDF components can either be determined from the data itself or by using monte-carlo simulation.

For a background hypothesis without any time-dependence, *i.e.* a constant background flux, the background time PDF, $\mathcal{T}_{\mathrm{B}}(t)$, is given through

$$\mathcal{T}_{\mathrm{B}}(t) \equiv \frac{1}{T_{\mathrm{obs}}}, \tag{27}$$

where T_{obs} is given by equation (21).

1.7 Notes on the energy PDFs for Signal & Background

In general, the energy PDFs are detector response dependent. That means they depend on the local direction of the detected events. Hence, the spatial and energy PDFs cannot be factorized entirely in space and energy.

For IceCube the energy resolution mostly depends on the zenith angle, and hence on the declination, of the event. Thus, several energy PDFs are created for a set of (reconstructed) declination bands, both, for signal and background. At the data evaluation, the signal and background PDFs are selected corresponding to the declination band the event's declination is part of. Hence, for IceCube, the signal and background energy PDFs can be formulated as $\mathcal{E}_{\mathrm{S}}(E|\delta,\vec{p}_{\mathrm{s,energy}})$ and $\mathcal{E}_{\mathrm{B}}(E|\delta)$, respectively.

A lengthly discussion has been conducted in the past to clarify whether the true or reconstructed direction of the monte-carlo events should be used to generate the several signal energy PDFs. Since, we mainly use experimental data as background estimation it as been concluded to use the reconstructed event direction in order to be consistent in the data evaluation for signal and background PDFs.

1.8 Stacking of Sources

In general a likelihood value can be calculated for a set of K stacked sources in a weighted fassion. In this case the signal PDF expression of equation (17) becames a bit more complicated due to the relative source weighting. The sources must be weighted according to their signal detection efficiency, $\mathcal{Y}_{\mathrm{s},k}$, and a relative strength weight of the sources, W_k, with $\sum_{k=1}^{K} W_k = 1$. Hence, the combined signal PDF is given as

$$\mathcal{S}_i(\vec{p}_{\mathrm{s}}) \equiv \frac{\sum_{k=1}^{K} W_k \mathcal{Y}_{\mathrm{s}}(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k}) \mathcal{S}_i(\vec{p}_{\mathrm{s}_k})}{\sum_{k=1}^{K} W_k \mathcal{Y}_{\mathrm{s}}(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k})}. \tag{28}$$

One should note that this formalism allows for different source properties, e.g. energy spectra, for the various sources.

1.9 Gradients of the Log-Likelihood Ratio

For maximizing the log-likelihood ratio function (equation (9)), or minimizing the negative of it, the minimizer algorithm requires the derivatives of the log-likelihood ratio function w.r.t. the fit parameters, n_{s} and $\vec{p}_{\mathrm{s}}$. Hence, here we provide the expressions of these derivatives for the optimized log-likelihood ratio function as given by equation (15).

The derivative w.r.t. n_{s} is given by

$$\frac{\mathrm{d} \log \Lambda(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}}} = \sum_{i=1}^{N'} \frac{\mathrm{d} \log \Lambda_i(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}}} - \frac{N - N'}{N - n_{\mathrm{s}}}. \tag{29}$$

For the numerical stable case, where $\alpha_i > \alpha$, the derivative of the log-likelihood ratio function of the ith event w.r.t. n_{s} is given by the derivative of equation (8) w.r.t. n_{s}:

$$\frac{\mathrm{d} \log \Lambda_i(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}}} = \frac{\mathcal{X}_i(\vec{p}_{\mathrm{s}})}{1 + \alpha_i(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}. \tag{30}$$

For the numerical unstable case, where $\alpha_i \leq \alpha$, this derivative is given by the derivative of the Taylor expension of equation (12) w.r.t. n_{s}:

$$\frac{\mathrm{d} \log \Lambda_i(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}}} = \frac{1}{1 + \alpha} \left(1 - \tilde{\alpha}_i(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})\right) \mathcal{X}_i(\vec{p}_{\mathrm{s}}) \tag{31}$$

For calculating the test-statistic, *c.f.* section 1.3, the second derivative w.r.t. n_{s} become in handy for the case $n_{\mathrm{s}} = 0$. Hence, it is provided here as well:

$$\frac{\mathrm{d}^2 \log \Lambda(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}}^2} = \sum_{i=1}^{N'} \frac{\mathrm{d}^2 \log \Lambda_i(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}}^2} - \frac{N - N'}{(N - n_{\mathrm{s}})^2} \tag{32}$$

The second derivative w.r.t. n_{s} of the log-likelihood ratio function for an individual event with $\alpha_i > \alpha$ is given by the derivative of equation (30):

$$\frac{\mathrm{d}^2 \log \Lambda_i(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}}^2} = -\left(\frac{\mathcal{X}_i(\vec{p}_{\mathrm{s}})}{1 + \alpha_i(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}\right)^2 = -\left(\frac{\mathrm{d} \log \Lambda_i(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}}}\right)^2 \tag{33}$$

For the event case $\alpha_i \leq \alpha$ this second derivative would be a constant due to the second-order nature of the chosen Taylor expansion in that case. At the junction point α the second derivative would not be differentiable. Hence, equation (33) is used as well in this case, with $\mathrm{d}\log\Lambda_i(n_\mathrm{s}, \vec{p}_\mathrm{s})/\mathrm{d}n_\mathrm{s}$ given by equation (31). This provides a second derivative that is differentiable for all $(1+\alpha_i)$ values, does not diverge for $(1+\alpha_i) \to 0$, and is closer to the second derivative of $\log\Lambda_i(n_\mathrm{s}, \vec{p}_\mathrm{s})$ for the $\alpha_i > \alpha$ case.

The derivative w.r.t. an individual signal parameter, p_s, is given by

$$\frac{\mathrm{d}\log\Lambda(n_\mathrm{s}, \vec{p}_\mathrm{s})}{\mathrm{d}p_\mathrm{s}} = \sum_{i=1}^{N'} \frac{\mathrm{d}\log\Lambda_i(n_\mathrm{s}, \vec{p}_\mathrm{s})}{\mathrm{d}p_\mathrm{s}} \tag{34}$$

Again, one needs to distinguish between the numerical stable ($\alpha_i > \alpha$) and unstable ($\alpha_i \leq \alpha$) case. For the stable case the event-based derivative w.r.t. p_s is given by

$$\frac{\mathrm{d}\log\Lambda_i(n_\mathrm{s}, \vec{p}_\mathrm{s})}{\mathrm{d}p_\mathrm{s}} = \frac{n_\mathrm{s}}{1+\alpha_i(n_\mathrm{s}, \vec{p}_\mathrm{s})} \frac{\mathrm{d}\mathcal{X}_i(\vec{p}_\mathrm{s})}{\mathrm{d}p_\mathrm{s}}. \tag{35}$$

For the numerical unstable case this derivative is given by the derivative of the Taylor expension of equation (12) w.r.t. p_s:

$$\frac{\mathrm{d}\log\Lambda_i(n_\mathrm{s}, \vec{p}_\mathrm{s})}{\mathrm{d}p_\mathrm{s}} = \frac{n_\mathrm{s}}{1+\alpha} \left(1 - \tilde{\alpha}_i(n_\mathrm{s}, \vec{p}_\mathrm{s})\right) \frac{\mathrm{d}\mathcal{X}_i(\vec{p}_\mathrm{s})}{\mathrm{d}p_\mathrm{s}}. \tag{36}$$

The derivative of $\mathcal{X}_i$ can be calculated using equation 4 and the expressions for the signal and background PDFs as given in equation 17 and 26, respectively. Depending on the type of fit parameter, i.e. spatial, energy, or time, the derivative of the PDF ratio, $\mathcal{R}_i(\vec{p}_\mathrm{s}) = \mathcal{S}_i(\vec{p}_\mathrm{s})/\mathcal{B}_i$, simplifies to the derivative of the respective type of PDF ratio:

$$\frac{\mathrm{d}\mathcal{X}_i(\vec{p}_\mathrm{s})}{\mathrm{d}p_\mathrm{s}} = \frac{1}{N} \frac{\mathrm{d}\mathcal{R}_i(\vec{p}_\mathrm{s})}{\mathrm{d}p_\mathrm{s}}, \tag{37}$$

with

$$\mathcal{R}_i(\vec{p}_\mathrm{s}) = \mathcal{R}_{S,i}(\vec{p}_{s,\mathrm{spatial}})\mathcal{R}_{\mathcal{E},i}(\vec{p}_{s,\mathrm{energy}})\mathcal{R}_{\mathcal{T},i}(\vec{p}_{s,\mathrm{time}}), \tag{38}$$

and

$$\frac{\mathrm{d}\mathcal{R}_i(\vec{p}_\mathrm{s})}{\mathrm{d}p_{\mathrm{s,spatial}}} = \frac{\mathrm{d}\mathcal{R}_{S,i}(\vec{p}_{s,\mathrm{spatial}})}{\mathrm{d}p_{\mathrm{s,spatial}}}\mathcal{R}_{\mathcal{E},i}(\vec{p}_{s,\mathrm{energy}})\mathcal{R}_{\mathcal{T},i}(\vec{p}_{s,\mathrm{time}}), \tag{39}$$

$$\frac{\mathrm{d}\mathcal{R}_i(\vec{p}_\mathrm{s})}{\mathrm{d}p_{\mathrm{s,energy}}} = \mathcal{R}_{S,i}(\vec{p}_{s,\mathrm{spatial}})\frac{\mathrm{d}\mathcal{R}_{\mathcal{E},i}(\vec{p}_{s,\mathrm{energy}})}{\mathrm{d}p_{\mathrm{s,energy}}}\mathcal{R}_{\mathcal{T},i}(\vec{p}_{s,\mathrm{time}}), \tag{40}$$

$$\frac{\mathrm{d}\mathcal{R}_i(\vec{p}_\mathrm{s})}{\mathrm{d}p_{\mathrm{s,time}}} = \mathcal{R}_{S,i}(\vec{p}_{s,\mathrm{spatial}})\mathcal{R}_{\mathcal{E},i}(\vec{p}_{s,\mathrm{energy}})\frac{\mathrm{d}\mathcal{R}_{\mathcal{T},i}(\vec{p}_{s,\mathrm{time}})}{\mathrm{d}p_{\mathrm{s,time}}}. \tag{41}$$

For stacked sources the expression for $\mathcal{R}_i(\vec{p}_\mathrm{s})$ in equation (38) becomes slightly more complicated due to the source strength weighting. With equation (28) and the definitions

$$a_k(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k}) = W_k\mathcal{Y}_\mathrm{s}(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k}), \tag{42}$$

and

$$A(\vec{p}_\mathrm{s}) = \sum_{k=1}^{K} a_k(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k}), \tag{43}$$

it is given by

$$\mathcal{R}_i(\vec{p}_\mathrm{s}) = \frac{\mathcal{S}_i(\vec{p}_\mathrm{s})}{\mathcal{B}_i} = \frac{1}{A(\vec{p}_\mathrm{s})} \sum_{k=1}^{K} a_k(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k}) \frac{\mathcal{S}_i(\vec{p}_{\mathrm{s}_k})}{\mathcal{B}_i}. \tag{44}$$

The signal over background ratio $\mathcal{S}_i(\vec{p}_{\mathrm{s}_k})/\mathcal{B}_i \equiv \mathcal{R}_{k,i}(\vec{p}_{\mathrm{s}_k})$ for the single source k is then given by equation (38).

Using the same set of source fit parameters $\vec{p}_\mathrm{s}$ for all sources, i.e. called global source fit parameters, the derivative of $\mathcal{R}_i(\vec{p}_\mathrm{s})$ for all stacked sources w.r.t. the single global source fit parameter, p_s, is then given by

$$\frac{\mathrm{d}\mathcal{R}_i(\vec{p}_\mathrm{s})}{\mathrm{d}p_\mathrm{s}} = -\frac{1}{A^2}\frac{\mathrm{d}A}{\mathrm{d}p_\mathrm{s}} \sum_{k=1}^{K} a_k \mathcal{R}_{k,i}(\vec{p}_{\mathrm{s}_k}) + \frac{1}{A}\sum_{k=1}^{K}\left(\frac{\mathrm{d}a_k}{\mathrm{d}p_\mathrm{s}}\mathcal{R}_{k,i}(\vec{p}_{\mathrm{s}_k}) + a_k \frac{\mathrm{d}\mathcal{R}_{k,i}(\vec{p}_{\mathrm{s}_k})}{\mathrm{d}p_\mathrm{s}}\right). \tag{45}$$

Using $\mathcal{R}_i(\vec{p}_\mathrm{s})$ from equation (44) it simplifies to

$$\frac{\mathrm{d}\mathcal{R}_i(\vec{p}_\mathrm{s})}{\mathrm{d}p_\mathrm{s}} = \frac{1}{A(\vec{p}_\mathrm{s})}\left[-\mathcal{R}_i(\vec{p}_\mathrm{s})\frac{\mathrm{d}A}{\mathrm{d}p_\mathrm{s}} + \sum_{k=1}^{K}\left(\frac{\mathrm{d}a_k}{\mathrm{d}p_\mathrm{s}}\mathcal{R}_{k,i}(\vec{p}_{\mathrm{s}_k}) + a_k \frac{\mathrm{d}\mathcal{R}_{k,i}(\vec{p}_{\mathrm{s}_k})}{\mathrm{d}p_\mathrm{s}}\right)\right], \tag{46}$$

with the derivative of $A(\vec{p}_\mathrm{s})$ given by

$$\frac{\mathrm{d}A}{\mathrm{d}p_\mathrm{s}} = \sum_{k=1}^{K}\frac{\mathrm{d}a_k}{\mathrm{d}p_\mathrm{s}}. \tag{47}$$

In case one would fit each source individually with its own set of signal fit parameters, $\vec{p}_{\mathrm{s},k}$, $\vec{p}_\mathrm{s}$ would be a set of K sets of source fit parameters $\vec{p}_{\mathrm{s},k}$, and a derivative for each individual source fit parameter $p_{\mathrm{s},k}$ would have to be calculated. The expression for such a derivative would be similar to equation (46), but only the summand for the particular source, for which the fit parameter is for, would contribute.

1.10 Multiple Datasets

With Skyllh a set of J different data samples (datasets) D_j can be analyzed at once. Each data sample has its own detector signal efficiency $\mathcal{Y}_{\mathrm{s}_j}$.

The composite likelihood function is the product of the individual dataset likelihood functions:

$$\log\Lambda = \sum_{j=1}^{J}\log\Lambda_j \tag{48}$$

The total number of signal events n_s needs to get split-up into n_{s_j} for the individual datasets. The distribution of n_s along the different datasets is based on the detector signal efficiency, $\mathcal{Y}_{\mathrm{s}_j}$, of each dataset. For a single source it is given by:

$$n_{\mathrm{s}_j}(n_\mathrm{s}, \vec{p}_\mathrm{s}) = n_\mathrm{s} \frac{\mathcal{Y}_{\mathrm{s}_j}(\vec{x}_\mathrm{s}, \vec{p}_\mathrm{s})}{\sum_{j'=1}^{J} \mathcal{Y}_{\mathrm{s}_{j'}}(\vec{x}_\mathrm{s}, \vec{p}_\mathrm{s})}, \tag{49}$$

where $\vec{x}_\mathrm{s}$ and $\vec{p}_\mathrm{s}$ denote the source position and flux fit parameters, e.g. the spectral index γ, respectively. The detector signal efficiency can be calculated via the detector effective area and the source flux (*c.f.* section 2).

By defining the dataset weight factor

$$f_j(\vec{p}_\mathrm{s}) \equiv \frac{\mathcal{Y}_{\mathrm{s}_j}(\vec{x}_\mathrm{s}, \vec{p}_\mathrm{s})}{\sum_{j'=1}^{J} \mathcal{Y}_{\mathrm{s}_{j'}}(\vec{x}_\mathrm{s}, \vec{p}_\mathrm{s})} \tag{50}$$

with the property

$$\sum_{j=1}^{J} f_j = 1 \tag{51}$$

equation 49 reads

$$n_{\mathrm{s}_j}(n_\mathrm{s}, \vec{p}_\mathrm{s}) = n_\mathrm{s} f_j(\vec{p}_\mathrm{s}) \tag{52}$$

Using the dataset weight factor $f_j(\vec{p}_\mathrm{s})$ the likelihood ratio of equation (48) with equation (6) can now be written as

$$\log \Lambda(n_\mathrm{s}, \vec{p}_\mathrm{s}) = \sum_{j=1}^{J} \sum_{i=1}^{N} \log(1 + n_\mathrm{s} f_j(\vec{p}_\mathrm{s}) \mathcal{X}_i(\vec{p}_\mathrm{s})). \tag{53}$$

From a reuseability-of-software point of view it is advisable to be able to use the mathematical form of $\log \Lambda$ for the single dataset to calculate the combined $\log \Lambda$ value of the multiple dataset. This can be achieved by using the substitution of n_s as given by equation (52). Hence,

$$\log \Lambda(n_\mathrm{s}, \vec{p}_\mathrm{s}) = \sum_{j=1}^{J} \log \Lambda_j(n_{\mathrm{s}_j}(n_\mathrm{s}, \vec{p}_\mathrm{s}), \vec{p}_\mathrm{s}). \tag{54}$$

For multiple point sources, i.e. a stacking of K point sources with positions $\vec{x}_{\mathrm{s}_k}$, the dataset weight factor of each single source needs to be taking into account via Bayes' theorem. Thus, $f_j(\vec{p}_\mathrm{s})$ can be written as the sum of the products of the dataset weight factor $f_j(\vec{p}_{\mathrm{s}_k})$ for source k, as given by equation (50), and the relative strength, $f_k(\vec{p}_{\mathrm{s}_k})$, of the kth source in all datasets compared to all the other sources in all datasets.

$$f_j(\vec{p}_\mathrm{s}) = \sum_{k=1}^{K} f_k(\vec{p}_{\mathrm{s}_k}) f_j(\vec{p}_{\mathrm{s}_k}) \tag{55}$$

The relative strength of source k can be written as

$$f_k(\vec{p}_{\mathrm{s}_k}) = \frac{\sum_{j=1}^{J} \mathcal{Y}_{\mathrm{s}_{j,k}}(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k})}{\sum_{\kappa=1}^{K} \sum_{j=1}^{J} \mathcal{Y}_{\mathrm{s}_{j,\kappa}}(\vec{x}_{\mathrm{s}_\kappa}, \vec{p}_{\mathrm{s}_\kappa})} \tag{56}$$

By combining equation 50 with $\vec{x}_{\mathrm{s}} \equiv \vec{x}_{\mathrm{s}_k}$ and $\vec{p}_{\mathrm{s}} \equiv \vec{p}_{\mathrm{s}_k}$, and equation 56, the expression for $f_j(\vec{p}_{\mathrm{s}})$ for multiple sources is given by:

$$f_j(\vec{p}_{\mathrm{s}}) = \sum_{k=1}^{K} \frac{\left(\sum_{j'=1}^{J} \mathcal{Y}_{\mathrm{s}_{j',k}}(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k})\right) \mathcal{Y}_{\mathrm{s}_{j,k}}(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k})}{\left(\sum_{\kappa=1}^{K} \sum_{j'=1}^{J} \mathcal{Y}_{\mathrm{s}_{j',\kappa}}(\vec{x}_{\mathrm{s}_\kappa}, \vec{p}_{\mathrm{s}_\kappa})\right) \left(\sum_{j'=1}^{J} \mathcal{Y}_{\mathrm{s}_{j',k}}(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k})\right)} \tag{57}$$

The sum over the datasets of the detector signal efficiency for source k cancels out leaving the simplified equation

$$f_j(\vec{p}_{\mathrm{s}}) = \frac{\sum_{k=1}^{K} \mathcal{Y}_{\mathrm{s}_{j,k}}(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k})}{\sum_{k=1}^{K} \sum_{j'=1}^{J} \mathcal{Y}_{\mathrm{s}_{j',k}}(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k})}. \tag{58}$$

1.11 Gradients of the Multi-Dataset Log-Likelihood Ratio

By using equation (54) for the combined log-likelihood ratio, its derivative w.r.t. n_{s} is given by

$$\frac{\mathrm{d} \log \Lambda(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}}} = \sum_{j=1}^{J} \frac{\mathrm{d} \log \Lambda_j(n_{\mathrm{s}_j}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}_j}} \frac{\mathrm{d} n_{\mathrm{s}_j}}{\mathrm{d} n_{\mathrm{s}}}, \tag{59}$$

with

$$\frac{\mathrm{d} n_{\mathrm{s}_j}}{\mathrm{d} n_{\mathrm{s}}} = f_j(\vec{p}_{\mathrm{s}}). \tag{60}$$

Its second derivative w.r.t. n_{s} is given by

$$\begin{aligned} \frac{\mathrm{d}^2 \log \Lambda(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}}^2} &= \sum_{j=1}^{J} \frac{\mathrm{d}}{\mathrm{d} n_{\mathrm{s}}} \left(\frac{\mathrm{d} \log \Lambda_j(n_{\mathrm{s}_j}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}_j}} \right) \frac{\mathrm{d} n_{\mathrm{s}_j}}{\mathrm{d} n_{\mathrm{s}}} &(61)\\ &= \sum_{j=1}^{J} \frac{\mathrm{d}^2 \log \Lambda_j(n_{\mathrm{s}_j}, \vec{p}_{\mathrm{s}})}{\mathrm{d} n_{\mathrm{s}_j}^2} \left(\frac{\mathrm{d} n_{\mathrm{s}_j}}{\mathrm{d} n_{\mathrm{s}}} \right)^2. &(62) \end{aligned}$$

The derivative w.r.t. a single source fit parameter, p_{s}, consists of the partial derivatives of $\log \Lambda_j$ w.r.t. n_{s_j} and p_{s}:

$$\frac{\mathrm{d} \log \Lambda(n_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} p_{\mathrm{s}}} = \sum_{j=1}^{J} \left(\frac{\partial \log \Lambda_j(n_{\mathrm{s}_j}, \vec{p}_{\mathrm{s}})}{\partial n_{\mathrm{s}_j}} \frac{\mathrm{d} n_{\mathrm{s}_j}}{\mathrm{d} p_{\mathrm{s}}} + \frac{\partial \log \Lambda_j(n_{\mathrm{s}_j}, \vec{p}_{\mathrm{s}})}{\partial p_{\mathrm{s}}} \right), \tag{63}$$

with

$$\frac{\mathrm{d} n_{\mathrm{s}_j}}{\mathrm{d} p_{\mathrm{s}}} = n_{\mathrm{s}} \frac{\mathrm{d} f_j(\vec{p}_{\mathrm{s}})}{\mathrm{d} p_{\mathrm{s}}}. \tag{64}$$

In case of a single source, the expression for the derivative of the dataset weight factor, where $f_j(\vec{p}_{\mathrm{s}})$ is given by equation (50), reads via the quotient rule of differentation:

$$\frac{\mathrm{d} f_j(\vec{p}_{\mathrm{s}})}{\mathrm{d} p_{\mathrm{s}}} = \frac{\frac{\mathrm{d} \mathcal{Y}_{\mathrm{s}_j}(\vec{x}_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} p_{\mathrm{s}}} \sum_{j'=1}^{J} \mathcal{Y}_{\mathrm{s}_{j'}}(\vec{x}_{\mathrm{s}}, \vec{p}_{\mathrm{s}}) - \mathcal{Y}_{\mathrm{s}_j}(\vec{x}_{\mathrm{s}}, \vec{p}_{\mathrm{s}}) \sum_{j'=1}^{J} \frac{\mathrm{d} \mathcal{Y}_{\mathrm{s}_{j'}}(\vec{x}_{\mathrm{s}}, \vec{p}_{\mathrm{s}})}{\mathrm{d} p_{\mathrm{s}}}}{\left(\sum_{j'=1}^{J} \mathcal{Y}_{\mathrm{s}_{j'}}(\vec{x}_{\mathrm{s}}, \vec{p}_{\mathrm{s}})\right)^2}. \tag{65}$$

In case of multiple sources (stacking), the expression for the derivative of the dataset weight factor, where $f_j(\vec{p}_\mathrm{s})$ is given by equation (58) reads via the quotient rule of differentation:

$$\frac{\mathrm{d}f_j(\vec{p}_\mathrm{s})}{\mathrm{d}p_\mathrm{s}} = \frac{\left(\sum_{k=1}^{K} \frac{\mathrm{d}\mathcal{Y}_{\mathrm{s}_{j,k}}}{\mathrm{d}p_\mathrm{s}}\right)\left(\sum_{k=1}^{K}\sum_{j'=1}^{J}\mathcal{Y}_{\mathrm{s}_{j',k}}\right) - \left(\sum_{k=1}^{K}\mathcal{Y}_{\mathrm{s}_{j,k}}\right)\left(\sum_{k=1}^{K}\sum_{j'=1}^{J}\frac{\mathrm{d}\mathcal{Y}_{\mathrm{s}_{j',k}}}{\mathrm{d}p_\mathrm{s}}\right)}{\left(\sum_{k=1}^{K}\sum_{j'=1}^{J}\mathcal{Y}_{\mathrm{s}_{j',k}}\right)^2} \tag{66}$$

2 Detector Signal Efficiency

The detector signal efficiency $\mathcal{Y}_{\mathrm{s}_{j,k}}(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k})$ of a data sample j for a source k is defined as the integral over the energy of the product of the detector effective area and the differential flux $\frac{\mathrm{d}\Phi_\mathrm{s}}{\mathrm{d}E}$ of the source:

$$\mathcal{Y}_{\mathrm{s}_{j,k}}(\vec{x}_{\mathrm{s}_k}, \vec{p}_{\mathrm{s}_k}) \equiv \int_0^\infty \mathrm{d}E A_{\mathrm{eff}_j}(E|\vec{x}_{\mathrm{s}_k}) \frac{\mathrm{d}\Phi_\mathrm{s}}{\mathrm{d}E}(E|\vec{p}_{\mathrm{s}_k}) T_{\mathrm{live}_j} \tag{67}$$

It is the mean number of signal events per steradian expected from a source at position $\vec{x}_\mathrm{s}$ with source parameters $\vec{p}_\mathrm{s}$. In the most-general case, the source position $\vec{x}_\mathrm{s}$ consists of three quantities: right-ascention, declination, and observation time, i.e. $\vec{x}_\mathrm{s} = (\alpha_\mathrm{s}, \delta_\mathrm{s}, t_\mathrm{obs})$.

2.1 Effective Area

In Skyllh the effective area $A_{\mathrm{eff},j}$ of a data sample j is not calculated separately in order to avoid binning effects. However, the effective area can be calculated using the monte-carlo weights `mcweight`[2] of the simulation events. The monte-carlo weights have the unit GeV cm^2 sr. Using the monte-carlo weight, $w_{m,j}$, of the mth event of data sample j the effective area is given by the sum over the event weights divided by the solid angle and the energy range ΔE of the summed selected events:

$$A_{\mathrm{eff}_j}(E) = \frac{\sum_{m=1}^{M} w_{m,j}}{\Omega \Delta E} \tag{68}$$

2.2 The DetectorSignalEfficiency Class

`DetectorSignalEfficiency` provides a detector signal efficiency class to compute the integral given in equation (67). The detector signal efficiency depends on the flux model and its source parameters, which might change during the likelihood maximization process. It is also dependent on the detector effective area, hence is detector dependent. Thus, `DetectorSignalEfficiency` must

[2] In IceCube known as "OneWeight", but which already includes the number of used MC files.

Table 1: IceCube specific detector signal efficiency implementation methods.

Name of Class	Description
I3FixedFluxDetSigEff	IceCube detector signal efficiency implementation method for a fixed flux model, which might contain flux parameters, but which are not fit in the likelihood maximization process. This implementation assumes that the detector effective area depends solely on the declination of the source. This method creates a spline function of given order for the logarithmic values of the $\sin(\delta)$-dependent detector signal efficiency. The constructor of this implementation method requires a $\sin(\delta)$ binning definition for the monte-carlo events and the order of the spline function.
I3PowerLawFluxDetSigEff	IceCube detector signal efficiency implementation method for a power law flux model, implemented by the `PowerLawFlux` class. This method creates a 2D spline function of given orders for the logarithmic values of the $\sin(\delta)$-dependent detector signal efficiency for a range of γ values. This implementation method supports multi-processing.

be provided with a detector signal efficiency implementation method derived from the `DetSigEffImplMethod` class.

Detector signal efficiency values can be retrieved via the call operator `__call__(src_pos src_params)`, which takes the celestrial position of the source and the additional source parameters as arguments.

2.2.1 The DetSigEffImplMethod Class

`DetSigEffImplMethod` is an abstract base class and defines the interface between the detector signal efficiency implementation method and the `DetectorSignalEfficiency` class.

Table 1 lists all available IceCube specific detector signal efficiency implementation methods and their description.

3 The Concept of Source Hypothesis Groups

The analyses in Skyllh rely heavily on the calculation of detector signal efficiencies. As seen in section 2, the detector signal efficiency depends on the source hypothesis (spatial model and flux model) and the detector response (dataset). Hence, the analyses require detector signal efficiency instances for each combination of source and dataset. However, the sources might be of the same kind, i.e. having the same spatial model and the same flux model. For such sources detector signal efficiency instances are needed only for each dataset. Thus, we define a group of sources with the same spatial model and flux model as a *source* *hypothesis group*, G_s.

A source hypothesis group has a list of spatial source models, e.g. point-like source locations in case of point-like sources, a flux model, and a detection signal efficiency implementation method assigned.

Skyllh provides the `SourceHypoGroupManager` class to define the groups of source hypotheses.

4 Implemented Log-Likelihood Models

This section describes the implemented log-likelihood models. [1]

References

[1] Jim Braun, Mike Baker, Jon Dumm, Chad Finley, Albrecht Karle, Teresa Montaruli. Time-Dependent Point Source Search Methods in High Energy Neutrino Astronomy. *Astropart.Phys.*, 33:175–181, 2010.