Track Reconstruction and Point Source Searches with Antares
Heijboer, A.J.

Citation for published version (APA):
Heijboer, A. J. (2004). Track Reconstruction and Point Source Searches with Antares

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Unbinned likelihood

In this appendix the formula for the unbinned likelihood is derived by taking the expression for the binned likelihood and letting the bin-size go to zero. A similar derivation can be found in e.g. [88].

Consider the case that the data consists of uncorrelated events and that each event is characterised by $k$ observed parameters $x^1, ..., x^k$. Then the events could be binned (into an N-dimensional histogram) and, for each hypothesis (or theory), the expectation value of the number of entries in bin $i$ is given by the $k$-dimensional integral

$$\mu_i(H) = \int_{\text{bin}_i} \frac{dN(x^1, \ldots, x^k|H)}{dx^1 \ldots dx^k} dx^1 \ldots dx^k,$$  \hspace{1cm} (A.1)

where $\frac{dN(x^1, \ldots, x^k|H)}{dx^1 \ldots dx^k}$ is the number density of expected events with observed parameters $x^1, ..., x^k$ for the hypothesis $H$ and where the integration boundaries are the boundaries of bin $i$. For small bins, the integral is proportional to the value of the PDF for the observed parameters of event $i$:

$$\mu_i(H) \propto \frac{dN(x^1, \ldots, x^k|H)}{dx^1 \ldots dx^k}.$$  \hspace{1cm} (A.2)

The observed number of events $r_i$ in bin $i$ is distributed according to a Poisson distribution:

$$P(r_i|\mu_i) = \frac{e^{-\mu_i} \mu_i^{r_i}}{r_i!}.$$  \hspace{1cm} (A.3)

The total log likelihood is given by the sum of the log likelihood of the individual bins:

$$\log P(\text{data}|H) = \sum_i \log P(r_i|\mu_i(H)).$$  \hspace{1cm} (A.4)

If the size of the bins is chosen sufficiently small, all bins will contain either zero or one entries; equation A.4 can then be written as

$$\log P(\text{data}|H) = \sum_{i \in B_1} \log(\mu_i e^{-\mu_i}) + \sum_{i \in B_0} \log(e^{-\mu_i}),$$  \hspace{1cm} (A.5)

where $B_m$ indicates the collection of all bins with exactly $m$ entries. This can be rewritten as:

$$\log P(\text{data}|H) = \sum_{i \in B_1} \log(\mu_i) - \sum_{i \in \text{all bins}} \mu_i.$$  \hspace{1cm} (A.6)
The second term is the total number of predicted events \( \langle N_{\text{tot}} \rangle \). The first term can be expressed as a sum over all events:

\[
\log P(\text{data}|H) = \sum_{\text{events}} \log \left( \frac{dN(x_1, \ldots, x_N|H)}{dx_1 \ldots dx_N} \right) - \langle N_{\text{tot}} \rangle + C, \tag{A.7}
\]

where we have used equation A.2. The constant \( C \) does not depend on the hypothesis \( H \) and therefore it plays no role when calculating ML estimates or likelihood ratios.

For brevity, we introduce the following definition:

\[
\mathcal{N}(x_1, \ldots, x_N|H) \equiv \frac{dN(x_1, \ldots, x_N|H)}{dx_1 \ldots dx_N}, \tag{A.8}
\]

which is the ‘event density’. \( \mathcal{N}(x_1, \ldots, x_N|H) \) may be thought of as the number of events we expect within a certain interval around the measured values \( x_1, \ldots, x_N \) for the hypothesis \( H \).

**Example**

As a simple example, consider the case were \( \mathcal{N} \) depends linearly on one of the model parameters, \( \varphi \), i.e. \( \mathcal{N}(x_1, \ldots, x_N|H(\varphi)) \propto \varphi \). The ML estimate of \( \varphi \) can be calculated by setting \( \frac{d}{d\varphi} \log P(\text{data}|H(\varphi)) = 0 \), which yields

\[
\hat{\varphi} = MA, \tag{A.9}
\]

where \( M \) is the number of observed events, and the constant \( A \equiv \frac{\langle N_{\text{tot}} \rangle}{\hat{\varphi}} \). Thus, the value of \( \hat{\varphi} \) is such that the expected number of events precisely equals the actual number of observed events: \( \langle N_{\text{tot}} \rangle = M \), irrespective of the observed parameters of the events and irrespective of the other model parameters. One may note the role of the term \( -\langle N_{\text{tot}} \rangle \) in equation A.7: if this term were omitted, the likelihood would have no maximum (\( \hat{\varphi} = \infty \)).