2.7 Maximum likelihood and the Poisson distribution

Our assumption here is that we have \( N \) independent trials, and the result of each is \( n_i \) events (counts, say, in a particle detector). We also assume that each trial has the same population mean \( \mu \), but the events follow a Poisson distribution.

The probability of \( n_i \) is then

\[
\text{prob}(n_i) = \frac{e^{-\mu} \mu^{n_i}}{n_i!}
\]

and so the likelihood for the whole set \( n_1, n_2 \ldots \) is

\[
\mathcal{L}(\text{data}|\mu) = \frac{e^{-\mu} \mu^{n_1}}{n_1!} \times \frac{e^{-\mu} \mu^{n_2}}{n_2!} \ldots
\]

or, more simply,

\[
\log \mathcal{L}(\text{data}|\mu) = -N\mu + \left( \sum_i n_i \right) \log \mu + \text{constant}.
\]

It seems natural to pick as an estimate of \( \mu \) the value that maximizes the likelihood (or its logarithm); the differentiation is easy and we get

\[
\hat{\mu} = \frac{1}{N} \sum_i n_i
\]

which is intuitive.

Maximizing the likelihood makes sense to a Bayesian because we are finding the peak of the posterior probability of \( \mu \), in the special case where the prior on \( \mu \) is flat over the region where the likelihood is appreciable.