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 μ, but the events follow a Poisson distribution.
The probability of n_i is then

$$\mathrm{prob}(n_i) = \frac{e^{-\mu}\mu^{n_i}}{n_i!}$$

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

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

or, more simply,

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

It seems natural to pick as an estimate of μ 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 μ, in the special case where the prior on μ is flat over the region where the likelihood is appreciable.