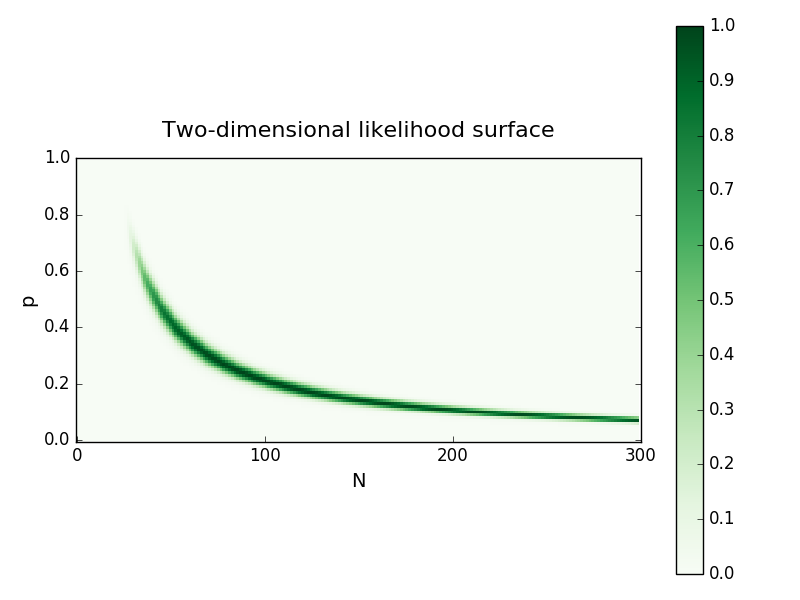
A complete quantitative description of experimental data not only requires a model of the intrinsic processes of interest, but also of the noise.

The process of evaluating the performance of a statistical on experimental observations frequently boils down to the construction of a likelihood function, which, loosely put, gives the probability of the observed data given the parameters of the model.

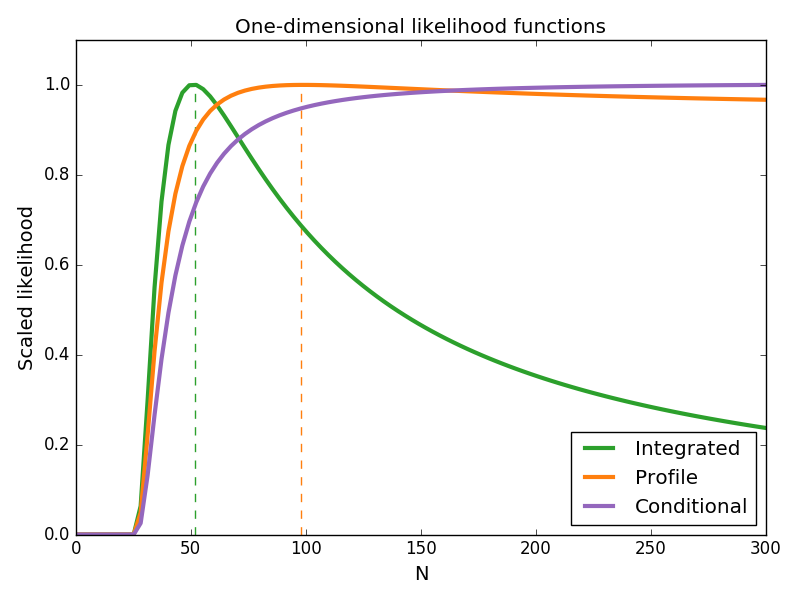
The likelihood function forms the basis for great deal of statistical inference techniques. For example, one could estimate the parameters of the model by maximizing the likelihood. The problem is, likelihood functions often contain more parameters than we care about. The inexplicable or uninteresting noise that buffets processes of interest has to be parameterized and accounted for in the likelihood function. As Berger et al. discuss in this paper, the existence of so called ‘nuisance parameters’ severely hampers inference in many cases. They review a few of the common frequentist techniques for dealing with nuisance parameters in likelihood functions, but fall strongly in favor of integrating the likelihood function over the nuisance parameters. Although this method has a very Bayesian flavor to it, the authors emphasize the pragmatic benefits of integrated likelihoods.

Berger et al. present a number of scenarios where an integrated likelihood function forms a more robust basis for inference than the frequentist candidates. While some of the examples seem a little contrived, the benefit of using an integrated likelihood is exemplified in Example 4 of the paper, which involves estimating a parameter from samples drawn from a binomial distribution. This example is explored in this Jupyter notebook [hyperlink].

To help understand this example, imagine that we’ve been challenged by a friend to estimate the number of times a coin has been flipped in *k* sessions, but we’re only told how many times the coin landed on heads. In each session, the coin is flipped *N* times, and heads appears with a probability *p.* We’re told the number of times heads appeared for each session and want to estimate *N,* but don’t care what *p* is. The binomial distribution is the natural probabilistic model for this situation, and we can use it to construct a likelihood function for the number of times heads is drawn. The example in the paper of k=5, with number of heads 16, 18, 22, 25, 27 is gone through in detail here [hyperlink]. For this case, the full two-dimensional likelihood surface looks like this:



One way to estimate *N* is to maximize the two-dimensional likelihood and record the value of N and discard the value of p. This is equivalent to maximizing the ‘profile likelihood’ of N, the one-dimensional projection of the likelihood surface over the values of p that maximize the likelihood for a give value of N. An alternative way to reduce the dimensionality of the full likelihood is to condition it on the sufficient statistic of the data, creating the so called ‘conditional likelihood’. Instead, Berger et al. recommend *integrating* the likelihood surface over p, which is approximately proportional to the probability of observing N irrespective of the value of p. These (dimensionality reduced) likelihoods are shown below:



These three methods for eliminating the nuisance parameter *p* produce strikingly different likelihoods. As the conditional likelihood increases monotonically with *N*--- without limit---it provides no basis for estimating N. Both the integrated and profile likelihoods allow maximum likelihood estimation of N. However, the integrated likelihood has is more highly peaked than the profile likelihood, and has a maximum roughly half that of profile likelihood (shown in dotted lines).

with the outputs of the inherent process we’re trying to model are buffeted by either inexplicable or uninteresting noise.

If the model is parametric, it’ Often this boils down to construction of a likelihood function, which is best summarized as the probability of the data given

Scientists have inexorable to model observations. Models

Much of statistical inference boils down to the construction of the likelihood function. For a given model, the likelihood function can be thought of loosely as the probability of the observation given the parameters of the model.