

# **An Algorithm for Approximating Integrated Likelihood Functions with Applications in Meta-Analysis**

**A Ph.D. Dissertation Prospectus**

Timothy Ruel

## **Introduction**

The research for my dissertation involves developing a novel algorithm for numerically integrating the likelihood function of a statistical model with respect to a nuisance parameter. This prospectus aims to demonstrate how the algorithm works and explain the appeal of using an integrated likelihood function over other types of pseudolikelihood functions to make inferences about the parameter of interest in a model.

The motivation behind my research developed out of the observation that the expression for the integral of a likelihood function follows a form similar to that of the calculation of the marginalizing constant of a posterior distribution.

## **Background**

Consider a random sample  $\mathbf{x} = (x_1, \dots, x_n)$  drawn from a population. What can we say about the population based on  $\mathbf{x}$ ? Where is its point of central tendency located? Are its values clustered tightly around this point, or are they more diffuse? Are they distributed symmetrically or skewed to one side or the other? Questions like these were the original motivation behind the field of statistical inference, and many of the techniques devised to answer them are still used by statisticians today.

It is important to remember, however, that the real world is messy and no mathematical function will ever perfectly capture the complexities of a population or random process whose properties we wish to infer. To overcome this difficulty, statisticians sacrifice a small amount of accuracy for (hopefully) a large reduction in complexity by imposing additional assumptions on the population of interest. These assumptions are essentially never true in the sense that they are not a flawless representation of reality, but they may nevertheless serve as convenient approximations capable of producing sufficiently accurate answers in their own right. As George Box famously put it, “All models are wrong, but some are useful.”

And indeed, in the aggregate these assumptions create what is known as a statistical model. In its most general framework, a statistical model can be formulated as a tuple  $(\mathcal{S}, \mathcal{P})$  where  $\mathcal{S}$  is the set of all possible observations (i.e. the population), and  $\mathcal{P}$  is a set of probability distributions on  $\mathcal{S}$ . The first and (in this author’s opinion) the most fundamental assumption we make when defining our models is that there exists some unknown mechanism in the population that generates the data we observe from  $\mathcal{S}$ . This mechanism is what induces the “true” probability distribution on  $\mathcal{S}$  though  $\mathcal{P}$  need not contain this distribution, and in practice it seldom does.

Another assumption found in almost every model is that the set  $\mathcal{P}$  is considered to be *parameterized*. That is, we assume the probability distributions contained in  $\mathcal{P}$  are indexed by a *parameter* that controls their features. This parameter acts like a tuning dial for the population - rotate the dial and certain behaviors of the population (e.g. its location, scale, or shape) will change. Much of statistical inference can be boiled down to figuring out the particular value to which a population parameter<sup>1</sup> has been set. Mathematically, we represent this assumption as  $\mathcal{P} = \{\mathcal{P}_\theta | \theta \in \Theta\}$ , where  $\theta$ <sup>2</sup> denotes the parameter, and  $\Theta$ , the parameter space, represents the set of all possible values  $\theta$  can take on.

---

<sup>1</sup>The phrases “population parameter” and “model parameter” have the same meaning and may be used interchangeably.

<sup>2</sup>Note that  $\theta$  can be multi-dimensional, and in fact this is usually the case. Despite this, a common convention is to refer to a model as having a single parameter  $\theta$  whose individual components represent .

Statisticians also like to assume the parameters in their models can be uniquely identified based on the data they observe. A model is considered *identifiable* if having perfect knowledge of the population<sup>3</sup> enables us to determine the true value of its parameter with absolute certainty. More formally, for any two parameters  $\theta_1$  and  $\theta_2$  in  $\Theta$ , if  $\mathcal{P}_{\theta_1} = \mathcal{P}_{\theta_2}$ , then it must follow that  $\theta_1 = \theta_2$ . Models that are not identifiable possess the undesirable property of having two or more distinct parameter values that give rise to the same probability distribution. Since we have already assumed  $\mathcal{P}$  is the mechanism generating the data we have observed in the first place, this would make it impossible to determine which value is the “correct” one on the basis of the data alone.

## What is a Likelihood Function?

Consider a model

## What is a Pseudolikelihood Function?

Nuisance parameters in a statistical model are a serious hindrance to making inferences about the parameter of interest. Consequently, statisticians have dedicated much time to developing techniques that will eliminate them from their models.

If we let

$$\Theta(\psi) = \{\theta \in \Theta \mid \varphi(\theta) = \psi\},$$

then corresponding to  $\psi \in \Psi$  is the set of likelihoods

$$\mathcal{L}_\psi = \{L(\theta) \mid \theta \in \Theta(\psi)\}.$$

Statisticians will often attempt to find a summary of the values in  $\mathcal{L}_\psi$  that does not depend on  $\lambda$ . This summary is called a **pseudolikelihood function** and can take several

---

<sup>3</sup>This is of course almost always impossible in practice, but in theory it could be accomplished by obtaining an infinite number of observations from  $\mathcal{S}$  or simply all of its observations if  $|\mathcal{S}|$  is finite.

different forms: - Maximizing - Conditioning - Marginalizing\* - **Integrating**

**Profile likelihood**

**Conditional Likelihood**

**Marginal Likelihood**

**Integrated Likelihood**

**Why Use an Integrated Likelihood?**

The appeal of the integrated likelihood function as a means of eliminating nuisance parameters from the model is that it incorporates

**Approximating the Integrated Likelihood**

**The Zero-Score Expectation Parameter**

**Markov Chain Monte Carlo**

**The IL Algorithm**

**Applications**

**Multinomial Distribution**

**Standardized Mean Difference**

**References**