

# A Common Subtle Error: Using Maximum Likelihood Tests to Choose between Different Distributions

By Gyasi Dapaa<sup>1</sup>

Predictive Modeling Group, Allstate Insurance, Northbrook, IL

The Maximum Likelihood Estimation (MLE) is one of the most popular methodologies used to fit a parametric distribution to an observed set of data. MLE's popularity stems from its desirable asymptotic properties. Maximum Likelihood (ML) estimators are consistent, which means that as the sample size increases the researcher becomes increasingly confident of obtaining an estimate that is sufficiently closer to the true value of the parameter; they are asymptotically normal with the lowest possible variance (achieve the Cramer-Rao lower bound on variance), which makes inference tests relatively easy and statistically more powerful. In addition, they are translation-invariant, which means that all functions of the ML estimates are by default the MLE predictors of the respective functions. For instance, if a pricing analyst computes pure premium from relativities estimated by MLE, the predicted pure premium is also an ML estimate and, hence, satisfies all the desirable aforementioned properties.

Because of the known asymptotic distribution of ML estimators, there are numerous asymptotic tests to help researchers make statistical inferences about their ML estimates: examples include, among others, the Likelihood Ratio Test, the Lagrange Multiplier Test, and the Schwarz Bayesian Criterion (SBC). In general, all of these aforementioned tests are used to determine if the measured signals (the ML estimates) are statistically different from some pre-specified values. For instance, suppose a researcher believes that frequency follows a Poisson distribution with mean  $\lambda$ , and computes the sample mean as the MLE for  $\lambda$ . To test whether or not the measured signal is noise, the researcher may use one of these tests to check whether the ML estimate is statistically different from zero. In addition, the researcher may use one of these tests to check whether the ML estimate(s) is (are) statistically different from some pre-conceived or historic values. However, the aforementioned tests **may not** be used to make inferences about the **functional form of the distribution** of the data. In other words, as an example, one may not compare the ML values (as is implicitly done by these tests) to choose between a Poisson and a Negative Binomial distribution. This article argues why such a comparison is incorrect and would be no better than an apple to orange comparison.

A critical assumption underlying MLE is that the researcher knows everything but a finite number of parameters of the specified distribution. (The functional form of a distribution has an infinite dimension). An implication of this is that this estimation technique could only be used after **the functional form of the distribution (hence forth, simply referred to as the distribution) has been pre-specified**. That is, a researcher needs to first specify whether the data is Poisson, Negative Binomial, Exponential, Lognormal, etc. before she could use the MLE technique to estimate the unknown parameters of the pre-specified distribution. Hence, the reader should easily see that the

---

<sup>1</sup> The author holds a Master's degree in Economics with a concentration in Econometrics and is currently an actuarial candidate.

MLE technique doesn't have the capability to determine the distribution of an observed data; otherwise, such a pre-specification of distribution would be unnecessary.

There is even a subtle contradiction invoked by comparing ML values obtained under different distribution assumptions as these tests implicitly do. For instance, if we assume data follows a Normal distribution and hence use the sample mean as the MLE of the shape parameter, it is easy to see that the sample mean would no longer be an MLE upon discovery that our data actually follows a Pareto distribution. **In other words, since a given data could only follow one distribution and since the ML estimator is valid only when the assumed distribution is right, comparing MLEs obtained under different distributions is self-contradictory!**

The reader should note, however, that distributions with different names do not necessarily have different functional forms. For instance, the Exponential and the Gamma distributions have the same functional form but differ only in the value of the shape parameter. (In other words, they differ in a finite number of parameters.) In fact, the Exponential distribution is a special form of the Gamma distribution. Hence, the ML tests are valid and can be used to make inferences about whether or not the shape parameter is one (and hence Exponential). **However, when the two distributions are rather distinct in functional form, but not in parameter values (such as the Weibull and Lognormal distributions), the ML tests are invalid!**<sup>2</sup>

In light of the above argument, all MLE inference tests such as the Likelihood Ratio Test, the Lagrange Multiplier Test, and the Schwarz Bayesian Criterion (SBC) are not appropriate under different distributions. Unfortunately, many researchers unknowingly misapply these tests to choose between distributions, e.g., Poisson vs. Negative Binomial). Even in much of the exam oriented actuarial literature such as *Manuals for Actuarial Exam 4/C*, as well as some past exams, have questions that mistakenly ask candidates to use one of these ML tests to make inferences about different distributions. It is also worthy to point out that, under such scenarios, the inference statistics such as the Likelihood Ratio Statistic and the SBC are not only meaningless, but do not even follow a Chi-square distribution (as they traditionally do); hence, using the Chi-square critical regions to accept or reject the null hypothesis is erroneous.

An important question, therefore, is what tests can a researcher use to choose between different distributions. There are numerous statistical tests of distribution fit: Kolmogorov-Smirnov tests and Chi-square Goodness of Fit tests are examples of such tests. These tests tell the modeler whether or not there is good reason to trust the fitted distribution. Unfortunately, each of these tests could accept multiple distributions as good fits. When this happens, the modeler could choose the distribution with the maximum<sup>3</sup> p-value.

---

<sup>2</sup> There is, however, a hot debate about the validity of MLE when functional forms are parameterized so that they differ by a finite number of parameters. For instance, the Tweedie distribution could be parameterized by a p-parameter so that, by changing the p-parameter, the assumed functional form of the distribution changes.

<sup>3</sup> Notice that for most inference tests about distributions, a high p-value is support for the null hypothesis (the distribution being tested).