Statistics and Probability Letters 85 (2014) 63–68

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Some approximations of the logistic distribution with application to the covariance matrix of logistic regression

Ronnie Pingel *

Department of Statistics, Uppsala University, SE-751 20 Uppsala, Sweden

A R T I C L E I N F O

Article history:
Received 10 July 2013
Received in revised form 15 November 2013
Accepted 15 November 2013
Available online 21 November 2013

Keywords:
Density
Gaussian
Mixture
Normal
t-distribution

A B S T R A C T

In this paper, we show that a two-component normal mixture model provides a good approximation to the logistic distribution. This model is an improvement over using the normal distribution and is comparable with using the t-distribution as approximating distributions. The result from using the mixture model is exemplified by finding an approximative analytic expression for the covariance matrix of logistic regression with normally distributed random regressors.

1. Introduction

If a logistic function, $F(\cdot)$, and its derivative, $f(\cdot)$, are functions of random variables, then it is generally not possible to find analytic expressions for the moments of these two functions unless some heavy restrictions are applied. In this paper we attempt to solve this problem by substituting the logistic function with some other function that closely resembles the logistic function. More specifically, this paper examines how well $F(\cdot)$ and $f(\cdot)$ are approximated by a normal distribution, a t-distribution and a normal mixture distribution. Using the mixture model, the resulting approximation is then applied to approximate the asymptotic covariance matrix in logistic regression having normally distributed regressors.

2. Approximation using a normal distribution or a t-distribution

Consider a standard logistic random variable X with mean zero and variance $\pi^2/3$. Its cumulative distribution function is $F(x) = [1 + \exp(-x)]^{-1}$ and its density is $f(x) = F(x)[1 - F(x)]$. The logistic distribution is a member of the location-scale family and although not belonging to the exponential family of distributions, it is well known that the logistic distribution is very similar to the normal distribution. Not only are the shapes of both distributions determined by location and scale parameters, but both distributions are also bell shaped. However, the logistic distribution has heavier tails than the normal distribution. More specifically, the excess kurtosis of the logistic distribution is 1.2. Still, because of the similarities it is appealing to approximate a logistic distribution using a normal distribution. We denote the distribution function and the density of a normal distribution having mean zero and standard deviation σ by $G(x)$ and $g(x)$ respectively.

Mudholkar and George (1978) propose a competing approximation. Because of the larger tails of Student's t-distribution compared with the normal distribution, they suggest using a t-distribution as an approximation of the logistic distribution

* Tel.: +46 739544242.
E-mail address: ronnie.pingel@statistics.uu.se.

0167-7152/$ – see front matter
http://dx.doi.org/10.1016/j.spl.2013.11.007

Table 1
Minimum approximation error when approximating a logistic distribution with a normal distribution, a t-distribution or a normal mixture model.

	Min. error	Value(s) minimizing error		
		σ		
$\|F(x) - G(x)\|_2$	2.31	1.699		
$\|F(x) - G(x)\|_\infty$	0.95	1.702		
$\|f(x) - g(x)\|_2$	2.07	1.630		
$\|f(x) - g(x)\|_\infty$	1.15	1.618		
		ν	s	
$\|F(x) - T(x)\|_2$	0.19	7.031	1.549	
$\|F(x) - T(x)\|_\infty$	0.06	6.974	1.548	
$\|f(x) - t(x)\|_2$	0.15	6.424	1.540	
$\|f(x) - t(x)\|_\infty$	0.07	6.400	1.539	
		υ	ω_1	ω_2
$\|F(x) - H(x)\|_2$	0.15	0.567	1.304	2.300
$\|F(x) - H(x)\|_\infty$	0.07	0.505	1.247	2.227
$\|f(x) - h(x)\|_2$	0.17	0.478	1.243	2.168
$\|f(x) - h(x)\|_\infty$	0.08	0.460	1.231	2.143

Note: For the univariate optimization, a mixture of golden section search and parabolic interpolation is used. For the multivariate optimization the Nelder–Mead method is applied (Nelder and Mead, 1965). Because the Nelder–Mead method may not converge at a global optimum (McKinnon, 1998), we increase robustness by restarting the optimization 1000 times with uniformly distributed random vectors as start values. All numerical calculations are done using `R version 2.15.1`.

function. However, one drawback is that the expressions of the distribution function and the density of the t-distribution are more complicated than those of the normal distribution. In the following, let $T(x)$ denote the distribution function and $t(x)$ the density of a central t-distributed random variable with ν degrees of freedom and scale parameter s.

To measure how well the approximations perform we consider two measures of accuracy: (i) the maximum absolute deviations, $\|\cdot\|_\infty$, or (ii), the square root of the average of all squared deviations, $\|\cdot\|_2$. Only a few analytic results exist in the literature, but Haley (1952) shows that $\|F(x) - G(x)\|_\infty$ is minimized at $\sigma \approx 1.702$. Other results, e.g., Mudholkar and George (1978), match the moments, which yield some degree of similarity between distributions but without actually addressing (i) or (ii). In this paper we rely on numerical optimization to minimize (i) or (ii). See the comments to Table 1 regarding the optimization routine.

According to the results in Table 1, the normal distribution provides a decent approximation of the logistic distribution. The maximum absolute error between the normal distribution and the logistic distribution is minimized to 0.0095 for the distribution function and 0.0115 for the density. Further, the minimum square root of the average absolute error is 0.0231 for the distribution function and 0.0207 for the density. However, there are two important qualifications concerning the conclusion of the approximation. First, what is good depends on the application. Second, if another function improves the approximation and is just as easily implemented, there is no reason not to use that function.

As can be seen in Table 1, using the t-distribution leads to a large decrease of the approximation errors. Depending on which error and function are studied, the errors using the t-distribution are between 12 and 16 times smaller than the errors when using the normal distribution. In fact, the maximum absolute error between the t-distribution and the logistic distribution is minimized to 0.0006 for the distribution function and 0.0007, for the density, while the minimum square root of the average absolute error are 0.0019 and 0.0017 for the distribution and density respectively. Clearly, the t-distribution is far superior in minimizing the approximation errors. On the other hand, it still suffers from having complicated expressions for the distribution and density functions.

3. The normal mixture approximation

The normal mixture model is widely used when considering occurrences of rare events (e.g., heavy-tailed probability models). Intuitively, the normal mixture distribution should therefore be able to take into account the heavier tails of the logistic distribution. For this purpose, we suggest as approximation the following two-component normal mixture model:

$$H(x) = \upsilon H_1(x) + (1 - \upsilon)H_2(x), \quad -\infty < x < \infty,\ 0 < \omega_1, \leq \omega_2 < \infty,\ 0 < \upsilon < 1, \tag{1}$$

$$h(x) = \upsilon h_1(x) + (1 - \upsilon)h_2(x), \quad -\infty < x < \infty,\ 0 < \omega_1 \leq \omega_2 < \infty,\ 0 < \upsilon < 1, \tag{2}$$

where $H_1(x), H_2(x), h_1(x)$ and $h_2(x)$ are the distribution functions and density functions of two normal distributions with zero means and standard deviations ω_1 and ω_2. Again, we seek to minimize (i) and (ii), now with respect to the three parameters that govern the shape of $H(x)$ and $h(x)$. Note that a mixture model may approximate any function arbitrarily well depending on the number of components (Sorenson and Aspach, 1971). Still, a two-component mixture model provides a balance between having a parsimonious model and providing a good approximation.

The results in Table 1 show that the normal mixture model works well as an approximation. Regarding the maximum absolute error, the approximation is roughly on par with the t-distribution, with 0.0007 and 0.0008 being the maximum