Article from
Predictive Analytics & Futurism
December 2017
Issue 16

Logistic GLM Credibility

By Matthias Kullowatz

For a recent project, our team built a logistic generalized linear model (GLM) to predict the probability of a binary outcome—in this case, whether or not the policyholder commenced lifetime withdrawals in a given quarter. We were naturally interested in determining the credibility of our probability estimates, and turned to our trusty Actuarial Standards of Practice (ASOPs) for some advice.

ASOP 25 specifically addresses credibility, and touches on extensions related to predictive modeling:

> More recent advancements in the application of credibility theory incorporate credibility estimation into generalized linear models or other multivariate modeling techniques. The most typical forms of these models are often referred to in literature as generalized linear mixed models, hierarchical models, and mixed-effects models. In such models, credibility can be estimated based on the statistical significance of parameter estimates, model performance on a holdout data set, or the consistency of either of these measures over time.[1]

It's left to us as the actuaries to develop defensible credibility methods from predictive models.

ASOP 25 comes across as purposefully open-ended as to what constitutes credible estimates from a predictive model. Because predictive modeling is relatively new to the life insurance industry, and because there exists a plethora of viable predictive modeling options, this open-endedness is essential. It's left to us as the actuaries to develop defensible credibility methods from predictive models.

It turns out there is a very familiar credibility method that GLMs are well equipped to utilize: limited fluctuation credibility. Before diving into a GLM implementation of this actuarial classic, we provide a helpful review for the reader.

LIMITED FLUCTUATION CREDIBILITY

Limited fluctuation credibility is why everyone loves the number 1,082. We'll come back to that in a moment. The method essentially revolves around calculating the probability that an estimate is within a chosen error tolerance of the true value being estimated, making it very much a frequentist approach. If that probability is high enough, then the estimate is deemed credible. Let's use a specific example that focuses on random binary outcomes.

Assume that we observe 100,000 policyholders over a defined period of time and that 1,082 of them die. Our estimate of mortality among this cohort would be approximately 0.0108. As the actuaries in charge, we decided that we want to be at least 90 percent confident that the true mortality lies within 5 percent of the estimated mortality. In probabilistic terms, that means we are requiring the following inequality to hold true to assure full credibility of the mortality estimate $\hat{q}$:

Formula 1

$$\mathrm{r}\left(0.95 \cdot \frac{1{,}082}{100{,}000} \leq \hat{q} \leq 1.05 \cdot \frac{1{,}082}{100{,}000} \mid n = 100{,}000, q = 0.0108\right) \geq 0.90$$

Note that confidence (90 percent) and proportional error tolerance (5 percent) are two parameters that we, as actuaries, selected somewhat arbitrarily. We assume that $n\hat{q}$ is a binomially distributed random variable with the aforementioned parameter values n and q. Recall that a single binary observation has a variance of q(1 – q), so we can normalize the probability statement and invoke the central limit theorem (CLT):

Formula 2

$$\Pr\left(\frac{\sqrt{100{,}000}(0.0103 - 0.0108)}{\sqrt{0.0108(0.9892)}} \leq Z \leq \frac{\sqrt{100{,}000}(0.0103 - 0.0108)}{\sqrt{0.0108(0.9892)}}\right) \geq 0.90$$

That probability in this formula is just a shade over 90 percent, and we deem this mortality estimation credible. As you may have guessed, it's no coincidence that 1,082 mortality claims led to a barely fully credible cohort. If you play with binomial distributions and the CLT long enough, you'll arrive at the following modified rule for the number of deaths required for full credibility, where k is the proportional error tolerance:

Formula 3

$$n\hat{q} \geq \left(\frac{Z_{\frac{\alpha}{2}}}{k}\right)^2 (1 - \hat{q})$$

For the 90 percent confidence and 5 percent error tolerance parameters, the required number of deaths would be 1,082.22 times $(1 - \hat{q})$. Given that mortality rates are typically lower than 1 percent, the required number of deaths for full credibility is likely to be close to 1,082 in any given cell. Thus, that number 1,082 comes directly from our choices of error tolerance and required confidence, along with some elementary probability theory.

There is one final rule to share for the credibility of binary proportion estimates, which can be derived algebraically from the one above. In words, if the margin of error on a confidence interval with a chosen probability (e.g., 90 percent) is smaller than the chosen proportional error tolerance of the estimate (e.g., $0.05\hat{q}$), then the estimate is fully credible. The general requirement is shown below:

Formula 4

$$Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{q}(1-\hat{q})}{n}} \leq k \cdot \hat{q}$$

In his featured article in *Risk Management*'s August publication, Mark Griffin stressed that the actuarial field is overdue to start thinking about limited fluctuation credibility as a hypothesis test.[2] Due to the close relationship between hypothesis testing and confidence intervals, it's a natural extension to also start thinking about limited fluctuation credibility as a comparison between confidence intervals and tolerance intervals, as described above. In fact, it's this connection to confidence intervals that paves the way to understanding the GLM credibility method outlined below.

A GLM CREDIBILITY METHOD

We presented the limited fluctuation credibility method as a comparison between a confidence interval and an error tolerance interval because it helps us to understand how the method can be applied to GLM output. Simple proportion estimates from a sample of binary outcomes and log-odds estimates (or "predictions") from a logistic GLM both have asymptotically normal distributions and calculable variances. So applying this GLM method is really just an exercise in finding the analogs between those two estimates, while navigating between the probability space [0, 1] and the log-odds space (-∞, ∞).

It is important to understand how to go mathematically between probability and log-odds because the logistic GLM explicitly models log-odds as a linear function of the selected covariates. The logit function takes us from probability (p) to log-odds (μ), and its inverse, the logistic function, takes us back. Both functions are shown below:

Formulas 5 & 6

$$\mu = logit(p) = \ln\left(\frac{p}{1-p}\right)$$

$$p = logistic(\mu) = \frac{e^{\mu}}{1+e^{\mu}} = \frac{1}{e^{-\mu}+1}$$

To create the error tolerance interval in the log-odds space, we first create the interval around the probability estimate as we did in the classical limited fluctuation credibility example. Recall that error tolerance is based on the actuary's selection of k. The two endpoints of the error tolerance interval are then translated into the log-odds space via the logit function shown above. Separately, the standard error of a GLM's log-odds estimate is constructed using the GLM's variance-covariance matrix of coefficient estimates. Standard errors of GLM estimates can be calculated and outputted very easily in most statistical software packages.

Once we've moved the error tolerance bounds into the log-odds space and calculated the standard error of each log-odds estimate, then basic normal theory takes over—that is, if the actuary desires 90 percent confidence, then she should use 1.645 standard errors, or if she desires 95 percent confidence, then she should use 1.960 standard errors, etc. If the confidence interval with chosen confidence level lives completely inside the error tolerance interval, then the GLM estimate is fully credible.

More formally, if the lower- and upper-bound conditions below are satisfied, then the GLM estimate is credible:

Formulas 7 & 8

$$logit[(1-k)\hat{p}_i] \geq \hat{\mu} - Z_{\frac{\alpha}{2}} \cdot stderr(\hat{\mu})$$

$$logit[(1+k)\hat{p}_i] \leq \hat{\mu} + Z_{\frac{\alpha}{2}} \cdot stderr(\hat{\mu})$$

Effectively, this approach uses model variance in the log-odds space as the analog for the binomial variance of a proportion estimate. Statistical theory supporting this method can be found in the article "Full Credibility with Generalized Linear and Mixed Models."[3]

CREDIBILITY CONSIDERATIONS

The limited fluctuation credibility method has one noted blind spot, described below, and now we are proposing moving into the log-odds space. The normality of GLM estimates is more fickle here than under the assumptions of the binomial distribution, and thus it's reasonable to question the utility of this GLM method. However, we found it useful for our GLM, and we think you will, too. Here are some things worth considering before applying this method to assess the credibility of GLM estimates:

1. **Defining error tolerance.** Using proportional error tolerance can be misleading when estimates range relatively close to zero, as they often do when estimating such things as mortality, lapse and withdrawal commencement rates. Using proportional error stresses how far the estimate is from zero as a driving force behind credibility, when we'd rather credibility be primarily a function of exposure and the amount at risk. Consider two cohorts, one with a 1 percent estimate and one with a 50 percent estimate. The proportional error tolerance would be 50 times greater for the 50 percent estimate, but we shouldn't expect the estimates' standard errors to vary nearly that much.

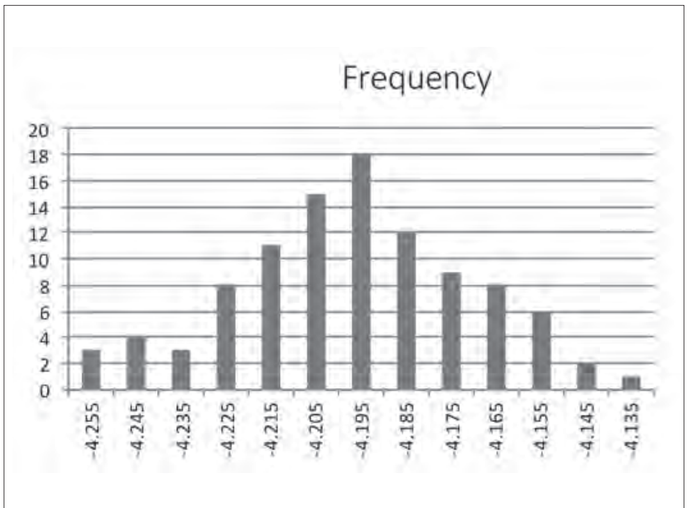
   It seems that credibility should be more closely tied to the potential bottom-line effect of estimation errors and the probability distribution of such errors, rather than to the size of the estimate itself. Those using this method should consider alternative error tolerance functions to appropriately account for such things as liabilities.

2. **Assumption of asymptotic normality in the log-odds space.** The cited paper on this GLM credibility method notes that the determination of full credibility relies on the asymptotic normality of the fitted coefficients—which in turn implies asymptotic normality of the log-odds estimates themselves. That is, the distribution of any given GLM log-odds estimate converges to normality as the training sample size increases to infinity. Thankfully, that has been proven before.[4] However, that does not guarantee normality for a

particular GLM's estimations, which are quite likely to base themselves on a **finite** data set.

To convince oneself that a GLM estimate has an approximate normal distribution, one method is to bootstrap sample the model's training data and produce a distribution of estimates. Using the training data for our GLM, we went back and randomly bootstrapped 100 samples of 2 million records each, and then refit the model to each sample to create distributions of the log-odds estimate for each policy. Figure 1 shows a sample histogram of estimates from one of our policies. Most histograms showed an approximately normal distribution with a little skewness like this one. However, each sample of 2 million records represented just 12 percent of our training sample size, so we took additional comfort

Figure 1
Distribution of Log-Odds Estimates From 100 Bootstrapped Models for a Single Policy



knowing that with increased sample size comes even closer proximity to normality.

3. **Probability space versus log-odds space.** Nonlinear link functions distort the error tolerance intervals when they are translated from the outcome space (i.e., probability space) to the link space (i.e., log-odds space). This can have a systematic effect of credibility becoming dependent on the value of the estimate itself. For the logistic model, this effect actually helps to soften the proportional error tolerance issue, discussed in the first consideration, for estimated probabilities less than 0.50. We encourage the modeler to investigate how her GLM's link function affects the relationship between the estimate's value and the estimate's credibility.
4. **Relative credibility.** Producing a credibility score is a natural extension of this GLM method. In addition to determining whether the credibility condition is met—see formula 3 or formula 4—one can back into the probability required so that the two sides of the condition are equal. That probability can be used to gauge how close the estimate is to being credible. The score can then be used in blending assumptions, such as between actuarial judgment and the GLM, or between a company's assumption and industry experience.

CONCLUDING REMARKS

While this is still an open area of research, the method presented here gives a viable option for quantifying credibility of an entire family of predictive models, presuming care is taken in defining the error tolerance desired. There are other methods of modeling and assessing credibility that each have advantages and disadvantages. For example, Bayesian analysis may allow the modeler to assess credibility directly, but Bayesian analysis is also limited by computational power.

Practitioners should expect to find that using a GLM offers greater credibility of predictions than a corresponding tabular study from the same size of data set. This is due to the fact that it absorbs information from the full domain of each predictor and that it can factor in the effect of individual predictors additively, rather than slicing the data into relatively small subsets. This GLM credibility method can help the actuary to translate the advantages of GLMs to the language of credibility. It's all about communicating what your models do and don't say to make your users comfortable with your assumptions and confident they are using them appropriately. ■

Matthias Kullowatz is an actuarial analyst and data scientist at Milliman in Seattle. He can be reached at *matthias.kullowatz@milliman.com.*

ENDNOTES

1 Actuarial Standards Board (December 2013). Actuarial Standard of Practice No. 25. Retrieved Sept. 8, 2017, from *http://www.actuarialstandardsboard.org/wp-content/uploads/2014/02/asop025_174.pdf.*

2 Griffin, M. (August 2017). Is credibility still credible? *Risk Management,* Issue 39, pp. 5-8.

3 Garrido, José, Jun Zhou. 2009. Full credibility with generalized linear and mixed models. Astin Bulletin. Retrieved Sept. 8, 2017, from *http://www.actuaries.org/LIBRARY/ASTIN/vol39no1/61.pdf.*

4 Lecture 3: Properties of MLE: Consistency, Asymptotic Normality. Fisher Information. Retrieved Sept. 8, 2017, from *https://ocw.mit.edu/courses/mathematics/18-443-statistics-for-applications-fall-2006/lecture-notes/lecture3.pdf.*