SOCIETY OF ACTUARIES

Article from

Risk & Rewards

August 2015
Issue 66

Dealing With Difficult Data

By Joshua Boehme

Life, so they say, brings with it the two certainties of death and taxes. Actuaries could perhaps make a convincing case that life also offers a third certainty: imperfect data.

Actuaries often encounter difficult data. Traditionally, we deal with problematic data in a number of ways. We check for data quality issues—such as incorrectly mapped codes—and correct them where possible. We question our preconceptions about what the data "should" look like and revise our working assumptions as needed. Mindful of how Simpson's paradox can lead us to incorrect conclusions, we check for additional, confounding variables. When we can, we gather additional data to reduce the influence of random noise. These techniques definitely have their worth and often solve the problem. Sometimes, though, they don't. When the standard methods fail, actuaries can turn to statistical methods to make the most of uncooperative data.

AN ILLUSTRATIVE PROBLEMATIC DATA SET

To illustrate the issues and some of the potential techniques to deal with them, consider the follow data set, which shows the annual default rates for financial institutions in 2012 by S&P rating category.

Table 1

One-year default rates for financial institutions (2012)[1]

S&P Rating	Defaults	Exposure	Default Rate
AAA	0	10	0.00%
AA+	0	25	0.00%
AA	0	12	0.00%
AA-	0	72	0.00%
A+	0	119	0.00%
A	0	130	0.00%
A-	0	107	0.00%
BBB+	0	108	0.00%
BBB	0	134	0.00%
BBB-	0	92	0.00%
BB+	0	48	0.00%
BB	0	50	0.00%
BB-	1	47	2.13%
B+	1	56	1.79%
B	2	65	3.08%
CCC/C	2	13	15.38%

Note: the summary here excludes withdrawn ratings.

As the basis for setting assumptions, this data has many shortcomings. The 12 highest rating levels all have the same observed default rate. In addition, the observed default rates do not increase monotonically, since BB- has a higher default rate than B+ and likewise for B and B-.

This does not necessarily mean that this data refutes the assumption that worse ratings have higher default rates. As a hypothetical example, suppose BB- has a true (unobservable) default rate of 1.80 percent and B+ has a true rate of 2.10 percent. Given the sample sizes we have available in the exposure amounts above, we would have a roughly 43 percent chance of observing a higher default rate for BB- than for B+.

Since we cannot "re-run" 2012, we cannot gather more data.[2] Since the data covers a single type of company and a single year, we have no obvious variables to divide the data into smaller cohorts. If we want an assumption where companies with lower ratings have higher default rates, we will have to work with this data further. In particular, we want to produce an assumption that satisfies this constraint:

Monotonicity: the default rate must strictly increase as ratings get worse.

THREE POTENTIAL TECHNIQUES

To work with this data, we move from looking at single point estimates to looking at distributions of estimates. This shift in perspective makes it easier to adjust the data to reflect any constraints we want to impose (such as monotonicity). With point estimates, we may know that we need to make an adjustment, but the individual values do not provide us with enough information to determine the size of the adjustment to make. With distributions, though, we can eliminate any regions that fall outside of our constraints; in effect, we take the distributions and make them conditional on our constraints.

The resulting distributions depend on the model we apply.

NORMAL DISTRIBUTION APPROACH

One possible approach, which many actuaries may already know, uses the normal approximation for the maximum likelihood estimator. Given d defaults and exposure n, this normal distribution has mean $\frac{d}{n}$ and variance

$$\frac{\left(\frac{d}{n}\right)\cdot\left(\frac{n-d}{n}\right)}{n} = \frac{d(n-d)}{n^3}.$$

Since the normal distribution can take any real values, we will reject any iterations that produce values outside of the [0, 1] interval. This will truncate the normal distribution so that it produces estimates of default probability between 0 percent and 100 percent.

BOOTSTRAPPING APPROACH WITH THE BINOMIAL DISTRIBUTION

Another approach makes direct use of the empirical distribution to draw new samples of the same size and with replacement from the observed results. In other words, we draw from a binomial distribution with *n* trials and *d/n* event probability. This technique—known as bootstrapping—offers a quick way to estimate parameters or variances in situations where we observe a process with an unknown distribution function and do not have a closed formula available.

BAYESIAN APPROACH WITH BETA PRIOR DISTRIBUTION

A third technique would take a Bayesian approach, which combines an assumed prior distribution and the observed data to produce a posterior distribution for our parameters. In this approach, the prior distribution and the data each are assigned a weight based on the credibility of the data; the more observed data points, the greater the relative weight assigned to the data. The example below uses the beta distribution as the prior distribution. The beta distribution is a convenient conjugate prior distribution for binomial data.

The table below presents the results of these techniques, each applied in several different ways, over 5,000 iterations using the same set of random numbers in each case. The estimated default rates in the table represent the means of samples drawn from the respective distributions (and in the case of the rejection method—discussed in more detail below—filtered to exclude observations outside the acceptance region). Note in particular that all of the techniques produce results that satisfy the monotonicity constraint.

The "Rejection" and "Gibbs" approaches are alternate sampling methods used to implement each technique. The "Prior/Adjustment" input is used in revising the initial result. These are explained more fully below.

Table 2

Results of estimation techniques

		Default Rates				
Technique	Prior/ Adjustment	BB-	B+	B	B-	Rejection %
Observed Data	Not applicable	2.13%	1.79%	3.08%	2.94%	Not applicable
Normal/ Rejection	0	1.21%	2.26%	3.61%	6.04%	95.3%
Normal/Gibbs	0	1.24%	2.33%	3.77%	5.87%	0%
Bootstrapping	0	0.32%	2.24%	4.20%	7.68%	97.1%
Bayesian/ Rejection	0	0.63%	1.47%	3.00%	6.10%	92.1%
Bayesian/ Gibbs	0	0.62%	1.47%	2.95%	5.84%	0%
Normal/ Rejection	0.5	1.62%	2.96%	4.62%	7.47%	94.3%
Normal/Gibbs	0.5	1.69%	3.07%	4.72%	7.39%	0%
Bootstrapping	0.5	0.83%	2.91%	4.92%	8.71%	95.5%
Bayesian/ Rejection	0.5	1.14%	2.27%	4.01%	7.58%	91.9%
Bayesian/ Gibbs	0.5	1.15%	2.27%	3.96%	7.32%	0%

For clarity, the results focus on just the ratings from BB- through B- (inclusive). The same techniques could apply to the entire table; the results in that case, though, would differ slightly.[3]

Although the specific values vary from technique to technique, some high-level similarities emerge across all the results. If we believe in our monotonicity constraint, then we might assume a significantly higher default for B- than we actually observed. Each of our models suggests this. The size of the increase in default rate from B to B- resulting from each model above may look surprising. However there is an intuitive explanation. If we were to evaluate only the B- default rate in isolation, we would have a range of plausible default probabilities with the observed rate (2.94 percent) most likely somewhere in the middle of the range. Applying the constraint, though, chops off any rates lower than B's default rate (observed at 3.08 percent), which, as a very rough approximation, removes the left half of the distribution.[4] The remaining portion has a mean value out in the right tail of our original unconstrained distribution.

Similarly, we see the results indicate we should assume a lower rate from BB- than we observed. The results show some disagreement about B+ and B, but most suggest somewhat higher default rates than we observed.

MULTIVARIATE SAMPLING

METHODS - REJECTION AND GIBBS SAMPLING

For the normal and Bayesian approaches, Table 2 shows two different methods of drawing from the respective constrained

CONTINUED ON PAGE 22

distributions. The rejection method draws each parameter independently from its distribution (normal or beta), ignoring our monotonicity constraint. In each random draw, if the resulting four sample default probabilities do not meet the constraints (i.e., among [0,1] and monotonic), they are discarded. We can envision this as drawing samples from a larger (but easier to simulate) distribution than the one we actually want. We then retain only the observations within the desired region—also known as the acceptance region—which leaves us with a sample from the desired (but more difficult to simulate) distribution.

Gibbs sampling, on the other hand, allows us to sample directly from complex distributions. The key to Gibbs sampling comes from observing that each variable, taken one at a time instead of collectively, has a distribution with a more tractable form. For example, in the case of the normal distribution we have four normally distributed variables, which depend on each other via the monotonicity constraint. Thus, the overall multivariate distribution has the same joint distribution as four independent normal variables, except truncated down to the region that satisfies our constraints (and scaled proportionally so the integral of the density still equals 1). That truncation makes it difficult to directly sample from it. If we look at just a single variable at a time, though, we have a univariate normal distribution subject to upper and lower bounds. As long as we know the other three variables, we know the bounds

> ASOP 23 should be consulted for additional guidance on data gathering and disclosure.

for the distribution of the fourth and thus can easily draw a sample from it. It turns out that the approach suggested by this observation—namely, sample from each variable individually based on the most recent value of the others—converges to the correct multivariate distribution.

Specifically, starting with an arbitrarily picked initial observation, Gibbs sampling generates additional observations through the following algorithm:

Loop through each variable, one at a time:[5]

1. Determine the distribution for the current variable, conditional on the current values of all other variables; and
2. Set the current variable equal to the value generated from a single random draw from that conditional distribution.

Each time we loop through all variables once, we produce one observation. We repeat this process as many times as needed to produce the desired number of observations.[6]

In our case, those conditional distributions take the form of a truncated univariate normal distribution, with the upper and lower bounds coming from the largest and smallest values that still satisfy the monotonicity constraint. Similarly, in the Bayesian case, we end up with truncated univariate beta distributions.

When we apply the rejection method, we treat the distributions as independent, making it easier to draw samples; however, many of the samples get rejected. Gibbs sampling, though less intuitively straightforward and requiring more up-front setup, draws precisely from the region of interest, making it more efficient.

In the end, both methods give us samples from the same distribution; the specific application dictates which works better. As the number of constraints increases, the rejection method rejects a greater proportion of the samples.[7] The rejection method, though, takes less time to program and to explain to non-technical stakeholders.

THE BAYESIAN PRIOR AND PSEUDO-OBSERVATIONS

The "Prior/Adjustment" column in Table 2 represents the number of pseudo-observations added to the observed number of defaulting and of non-defaulting companies. This provides a mathematical adjustment that reduces the credibility of the observed number of defaults. This is done as an adjustment for the relatively small number of exposures in each rating category. For the Bayesian techniques, it represents the value of the alpha and beta parameters for the assumed beta prior distribution.[8]

The concept of an adjustment factor may seem strange to some actuaries at first. To understand why one would add pseudo-observations to our actual data, consider the implications of not using an adjustment factor. In the extreme, this would mean, for example, that if a rating category had only one exposure and it represented a default we would assume a default probability of 100 percent in that rating category. This would clearly be a rather extreme approach.

A factor of 0.5, on the other hand, means that if we observe one default event after one trial, we would estimate the default probability for that rating category as $(1 + 0.5) / (1 + 0.5 \cdot 2) = 3/4$—not an outlandish place to start given a sample size of one.[9]

SELECTING A SINGLE TECHNIQUE

So, with a plethora of techniques to choose from, how do we narrow things down? The choice depends on the specific situation and must reflect non-technical factors, such as stakeholder buy-in. In this example, though, the technical factors favor one approach over the others.

First, given the small sample sizes and the even smaller number of defaults observed, the normal approximation seems dubious; in addition, the fact that the normal distribution ends up putting a significant proportion of the distribution on negative values gives us another reason to question it.

On the other hand, the bootstrapping technique can only produce certain discrete values (0/n, 1/n, 2/n, etc.). Since the monotonicity constraint leads us to eliminate certain overlapping regions of the parameters' distributions, the discreteness of the values leads to questions about the accuracy of the final distribution. (Consider, for example, the difference between a particular probability mass falling just inside the constrained region versus falling just outside the region.)

This leaves the Bayesian technique as the strongest approach. Further, as discussed above, using a non-zero parameter for the prior distribution produces better results.

In the end, though, these only reflect the technical considerations. The context of the work must guide the selection of the final technique. Actuarial Standard of Practice 23 should be consulted for additional guidance on data gathering and disclosure requirements.

In life we often encounter imperfect data. By using these and other statistical techniques, though, actuaries can prevent uncooperative data from causing as much unpleasantness as life's other certainties. ■

ENDNOTES

1 Table 52 from *http://www.nact.org/resources/NACT_2012_Global_Corporate_Default.pdf*, accessed 2/14/2015

2 Ignoring the challenge of incorporating data from prior years consistently

3 For example, including the BB rating means that our assumption for BB- now has a nonzero lower bound. This would cause us to exclude some of its smallest potential values and would thus increase its estimate. Similarly, including CCC/CC puts a ceiling on our assumed default rate for B-.

4 The extent of the adjustment depends on the likelihood of B's default rate at the given point. If we consider a point much lower than B's observed default rate (i.e., in B's left tail), it is highly likely that B's true value is greater than or equal to that point. Thus, given our monotonicity constraint it is highly unlikely B- can have a rate that small since it must exceed B's rate. In the absence of the constraint we would have assigned some probability to B-'s rate being that small; with the constraint, though, there is an even smaller chance. Conversely, for values much larger than B's observed rate (i.e., in B's right tail), there is a relatively small chance that B has a default that large; therefore, applying the constraint has only a minor impact on the likelihood of B- having a rate that high.

5 Or multiple variables at a time, in which case we draw from the joint distribution conditional on all the other variables.

6 In practice, we often make some adjustments to the resulting series of observations. Because of the iterative nature of the process, consecutive observations exhibit correlation – we do not get independent samples. In addition, depending on the initial starting point it may take some number of iterations to converge to the desired distribution. We can correct for auto-correlation by thinning the observations and for non-convergence by dropping observations from an initial burn-in period. Since this paper only considers means, we do not need to correct for auto-correlations. In addition, for simplicity the results do not discard any initial burn-in period (based on a visual inspection, the results quickly converge to the stationary distribution). Readers interested in further details can consult the extensive literature available on Gibbs sampling.

7 The inefficiency of the rejection method can reach rather extreme levels. The author encountered one situation involving a two-dimensional ratings transition matrix where the rejection method produced less than one valid result per million samples.

8 For 0, view this as the limit of the posterior distribution as alpha and beta go to zero.

9 The exact choice of a factor (or a prior distribution in general for Bayesian approaches) can present some problems beyond the scope of this article, but the illustrative 0.5 factor in this case has three desirable properties. First, in the extreme case of n=1 it produces defensible results. Second, for Bernoulli trials, a Beta distribution with parameters $\alpha = ½$ and $\beta = ½$ is the Jeffreys prior. As the Jeffreys prior, it has a certain invariance under re-parameterization. Third, it still results in a whole number of total observations, since 0.5 gets added to both the number of defaults and the number of non-defaults.

Joshua Boehme, FSA, MAAA works in the Asset-Liability Management department at Jackson National Life.