Forecasting & Futurism 4th Annual Contest Submission

Wisdom (?) Of The Crowd

Stephan Mathys, FSA

Submitted to the Society of Actuaries' Forecasting & Futurism Section

for the Actuarial Science Predictions and Forecasting 2014-2015 Contest

Abstract

This paper demonstrates the value of asking a diverse crowd of non-actuaries a specific question about the insurance industry. The crowd's responses can provide broader insight into the future than the traditional actuarial models, and can provide new tools to the actuary when making predictions.

Introduction

Actuaries, and business professionals in general, must deal with uncertainty in virtually every aspect of their work. From pricing, to reserving, to forecasting, attempting to manage uncertainty as to which events will happen in the future is the staple of the actuary's skill set.

Traditionally, one way to understand uncertainty is to assign a probability to the uncertain event. That is, if something is uncertain, but we can say it could happen with 20% probability, we can take some comfort in knowing that it's *possible*, but perhaps not *probable*.
A newer method for dealing with uncertainty is to build a model of the uncertain event, and then change input values to determine the impacts. These can be models of stock market movements, disability claims during recessions, or insurance income based on specific loss assumptions.

However, models which produce a single number are almost always wrong. In producing an expected income statement, for example, there will be assumptions regarding sales, lapses, new claims, development of existing claims, expenses, commissions, investment income, and so on. Any one of these assumptions can be wrong, from the first moment after the model is run, and will be flowing through to the final results, making those error-prone as well. As an alternative, suppose a model produced a range which has an expectation of being correct 50% of the time. The number of times the actual result fell within that range can be measured, tracked, and evaluated. As a result, the model can be tweaked, either in terms of model function or assumptions, and future expected ranges can be again evaluated. The goal with this

process should be to shrink the ranges and increase the associated probabilities, in order to allow the model users to make decisions with better information.

But how to create that range? Each model input could be considered a variable, with its own distribution, expected value, standard deviation, and so on. Now, instead of having 500 to 1,000 values to input, there could be 1,500 to 3,000. The number of potential errors and mistakes expands unmanageably. Not to mention overwhelming requirements for documentation, evaluation, review, and future tracking.

So is there an easier way to create variability around a central result? Might there also be a way to make a reasonability check against the expert's model, that could incorporate information the expert does not have access to? One way to build in this variability and reasonability at the same time is to ask another expert to build a second model, with an independent set of assumptions. This may be just as time-consuming, and may only produce one additional data point. Additionally, this model is also likely to be just as wrong immediately, due to the vast number of assumptions and features likely incorporated.

An alternative is to use the theory promoted in The Wisdom of Crowds (James Surowiecki, 2004), and ask many people, who each might have a little bit of specific knowledge that cannot be incorporated into a model, what their opinion is. The results may surprise the actuary.

Background and Question Design

The Wisdom of Crowds illustrates many examples of how crowds can be used to make better predictions than experts. For some, the knowledge has been dispersed into various front-line workers who cannot provide their expertise in a clear way back to the expert running the model. In others, the knowledge is simply too confusing to be understood by any individual, or too complex to be modeled in a reasonable, understandable way. The book also includes examples of group failures. These include not having enough independent voices to demonstrate potential errors, or having too much confidence in the expert even when the crowd is pointing to dramatically different results.

The value of crowd wisdom is that each individual on the front lines has a little bit of information. Due to bureaucracies, inefficient processes, or simplifications required in the model, the totality of the information does not make it into the prediction models. For example, in an insurance context, underwriters are reviewing applications daily. They will see leading indicators of shifts in exposure well before the actuaries can, as the actuaries are looking at sold business after the fact. Even if the underwriters can't quantify their intentions to one number of prediction, aggregating their various forecasts can give directional insight.

In order for the crowds to be valuable, Surowiecki suggests that they need four things:

1) They need to be independent. That is, there must be opportunity for each of the crowd members to come to their own conclusion. Otherwise, groupthink begins to take hold and disasters like the Challenger can occur.

2) There needs to be a way to aggregate results. If all you have is a bunch of random thoughts, it cannot make sense. The aggregation can happen in various ways. The first example of the book is from a county fair, in which an observer, Francis Galton, recorded the individual guesses of the crowd as to the weight of an animal. Galton was the aggregator. Without him, all that would have resulted was knowing one winner and a thousand losers.
3) The crowd needs to be decentralized. That is, each individual member of the crowd should have some unique knowledge that other crowd members do not have. The point of the crowd is lost if everyone is going to come to a similar conclusion anyway.
4) There should be encouragement of diversity of opinion. This is a little different from decentralization, in that opinions can vary amongst people in similar situations or job titles. Therefore, the idea should be that the crowd needs all ideas, not just a consensus number.

To these, I will add that there should also be some incentive to participate in the crowd-based activity. The alternative, to just ask for altruistic volunteers, is not likely to provide the valuable crowd wisdom sought.

So a logical question for the actuary is "Can the Crowd be wiser than the Actuary? Would the Actuary, or the Actuary's employer, find any kind of value in asking the Crowd a question, or a series of questions?" For the remainder of this paper, it will be assumed that the actuary works in the insurance industry; similarities could exist for actuaries in other industries.

Certainly, actuaries must know about all of the various elements of insurance services in order to be skilled at pricing, reserving, or forecasting the insurance products with which they work. But can an actuary know everything about claims processing changes? Can any model know all of the ins and outs of future investment income streams? Can any one individual consider all of the range of changes coming within the sales force, in new business procedural changes, in expense management, or within the technology space? Not likely. So the idea of asking the crowd for future impacts has a meaningful application to the insurance industry.

One thing to consider is that the question must be relevant to the crowd. It does no good to ask the IT group whether 65-year-old males will have higher or lower than average mortality in the next two months. It does no good to ask the claims management team whether new sales will be above, at, or behind the Sales VP's predictions next year. These people have no context for what they are being asked, and the questions are not relevant to their knowledge. So, one question that might be asked is, "How will things you are seeing impact the rest of the division?" But, as that is open-ended, and rather difficult (perhaps impossible) to quantify and aggregate the results, we must use a proxy.

The proxy question is, "What will the division's net gain be next year?" This question is relevant because it:

1) Has a quantifiable value.
2) Has some history that various crowd participants have experience with.

3) Is impacted by the tasks that the crowd participants perform day in and day out.

Clearly, since we are asking for a single number, it is quantifiable. The history is obvious; companies publish either internal documents or public documents outlining their net gain experience. And since each and every participant within the crowd is working on aspects that eventually reach the bottom line, they will have a stake in the outcome and be able to relate the things they are seeing to that question.

For example: suppose the Sales Reps are seeing sales opportunities shrink in the next year. They may conclude that profits will be lower, as there is less to go around, and adjust their guesses accordingly. The IT community may know of projects that are going to be completed in the next 6 months that will reduce processing waste by thousands of hours and tens of thousands of dollars, and adjust their guesses also. Taking the average across the crowd should give some kind of general feel for what will happen in the future.

For this question, too, asking the crowd brings another advantage. Because the crowd will have answers that are expected to be diverse, the Actuary can see just how much dispersion could be expected for future results. This has an advantage over the traditional model outlined above, in that as the model is always wrong, so is the crowd. But the crowd provides some kind of distribution of potential results, which can provide more information than the point number often given from the model. Looking at the 25^{th} and 75^{th} percentiles of the crowd's answers give a 50/50 range for future results. That is, if the crowd is going to be right half the time, this

range gives some kind of feel for how volatile the future will be. Remember – the model is likely to be wrong 100% of the time. The crowd, by virtue of a distribution of answers, has a chance to be right at least 50% of the time.

Procedure and Responses

An e-mail survey was sent to one entire division of the author's company. Additional audience members include actuaries within other divisions, general finance personnel (including accounting, investment, etc.), and employees of the company whom the author has worked with in various capacities over the years. The goal was to reach a broad spectrum of potential crowd members, each of which may have some insight as to future impacts on the division's future net gain. The question was posed as follows: "What will the Division's 2015 Quarterly GAAP Net Gain be? Please respond with a value for each quarter in thousands (000s)." Participants who responded will also be entered into a random drawing for a prize. This creates incentive to respond, and avoids the perverse incentive to attempt to influence results so that the outcome is closer to any respondent's entry.

The reason for asking for quarterly values was to get more data points to compare actual results against. After each quarter of 2015, results from the crowd can be compared to the internal model results and actual net. The measurement of how "well" the crowd has predicted, in comparison to how well the expert model predicts, will be in two ways.

First, there will be a measure of variance. The difference between actual net gain and the average of the crowd's estimate of net gain for each quarter and the year total will be tracked. This will be compared to the difference between actual net gain and the internal model estimate each quarter and year, as well. In order to provide meaningful comparisons, the square root of the sum of the squared deviations will be used.

The second measure will be whether or not the crowd's 25th and 75th percentile estimates reasonably illustrate a 50/50 range of potential results. That is, for one of the four quarters, the actual net gain should be less than the crowd's 25th percentile. One quarter's actual net gain should be greater than the 75th percentile. And the other two should be within the range. If so, then the crowd method can be seen as providing excess value over the point estimate of the expert model. If not, then the crowd method will be considered no excess value on this dimension.

Of the more than 100 e-mail recipients, 21 responded with estimates of future net gains. This crowd meets the 4 criteria outlined previously, in the following ways:

1) Independence: each survey respondent provided their own answer, not based on a committee or group, or by asking questions of the author to provide "guidance".
2) Aggregated results: the author received the responses via e-mail, and maintains a database of those responses.
3) Decentralization: responses came from a wide mix of audience members, both in an outside the department, ensuring a broad survey of information.

4) Diversity of opinion: as demonstrated in Table 1, there is a wide range of responses for each quarter and the full year.

The following table illustrates these values (values are in 000s, and have been scaled).

Table 1

Crowd and Actuary Estimates of 2015 Net Gain by Quarter

From the Crowd	Q1 Net Gain	Q2 Net Gain	Q3 Net Gain	Q4 Net Gain	Q1 – Q4 Net Gain
Minimum	-665	-332	-332	0	1,250
25th percentile	831	748	997	1,330	4,987
Median	1,330	1,330	1,579	1,912	6,483
75th percentile	2,660	1,662	2,493	2,739	8,976
Maximum	6,317	5,319	4,987	4,987	12,301
Average	**1,718**	**1,425**	**1,741**	**1,981**	**6,864**
From the Actuarial Model	2,201	2,653	2,676	2,470	10,000
Actuarial Model %ile of Crowd Responses	71.5	91.2	80.8	69.4	90.0

As demonstrated in the table, the crowd is more pessimistic than the Actuarial Model. Alternatively, it could be considered that the Crowd knows many little things that the Actuary does not know. The aggregate yearly net gain from the actuary is at the 90th percentile of the crowd estimates. This means that if the crowd is right, and the range of expected results follows the crowd's distribution, then the expert is likely to miss the mark, by a significant figure. The

Actuary is currently predicting about 150% *higher* than the crowd for the year. Decisions which use the results of the Actuary's model may need to consider the likelihood that the actuary has made a significant overstatement of future earnings.

The following histograms demonstrate this visually. The completely dark bars are single quarter Crowd values, while the striped bars are the Crowd's estimates for the year to date. The solid vertical lines indicate the Actuarial Model's quarter and year values, respectively.

Chart 1 – Quarter 1 Net Gain (000s, scaled)

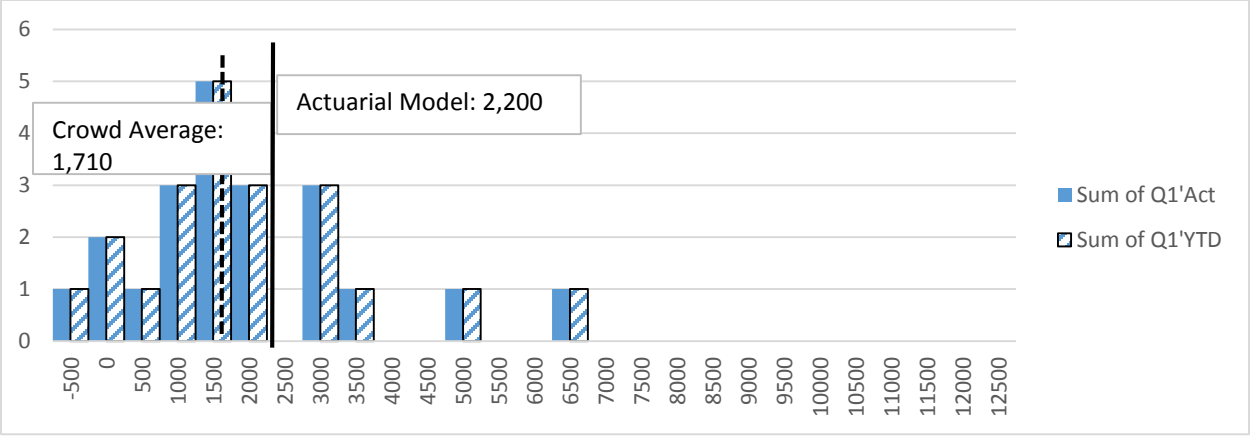
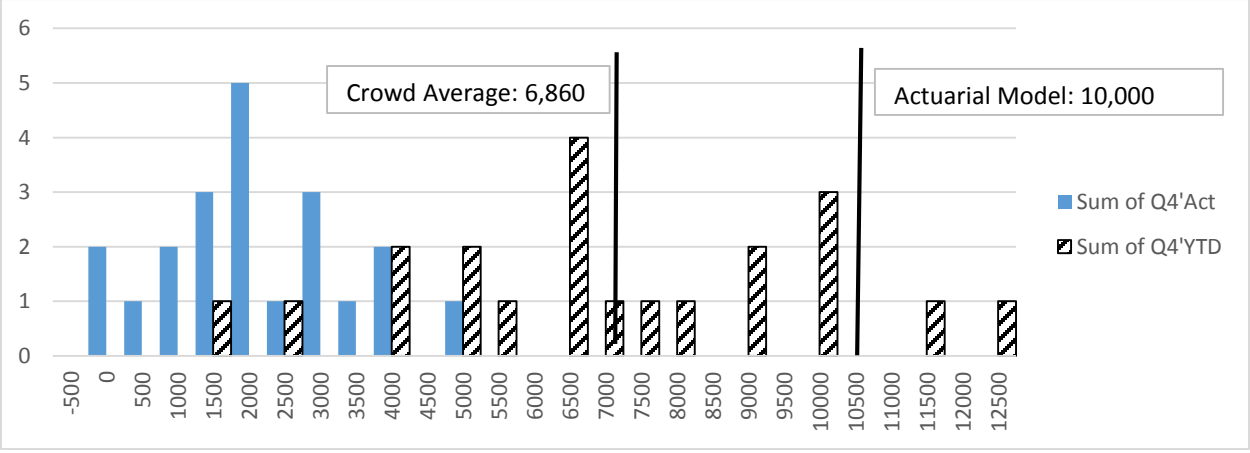


Chart 2 – Full Year Net Gain (000s, scaled)



Quarters 2 and 3 are similar to the shown quarters, so are not shown for brevity.

Additionally, demographic information was collected about the Crowd members. Of the 21 responses, 12 are from outside of the Division, 9 were Division employees. This gives a good spread of internal and external predictions. The Non-Division average prediction is $5,574 for the year. The Division expert average prediction is $8,562 for the year. This, too, is a significant difference, and is one that can indicate whether individuals internal to the division are better because they know more, or may be falling into the same traps as the actuary because they don't know enough of the external world.

Future Analysis

The next steps are to track the actual results each of the 4 quarters and measure the variance of those from the predictions of the Crowd and the Actuary. A follow-up survey is intended after the 2nd quarter of the year, to see if additional information influences the Crowd's estimates of the future. If new events have transpired throughout the first half of the year, some of the Crowd may be more optimistic or pessimistic, depending on what events they view. And on balance, the change in the Crowd's predictions for the 2nd half of the year will provide further data on whether the crowd's wisdom is diverging, converging, or staying consistent with previous estimates.

Application to Actuarial Functions

This kind of Crowd wisdom could have applications across the actuarial fields. The above example is about Financial Planning. However, a company or department may wish to ask a range of associates about Product Development. A good question might be, “Should we introduce this new product?” If the majority is overwhelming (whether for or against) then perhaps the front lines know something that the isolated experts don’t. That could be about system integrations, product risks, or other hidden aspects. If there is no clear direction (say, if it’s less than some excess majority of agreement or disagreement [60%? 70%? This value will differ by question.]), one could conclude that the people who will be actually working on the new product need more information.

Resource planning and prioritization is another area that could benefit, due to the quantifiable nature of the questions asked. “What should be done first, and then after that, and then after that?” is something that will affect everyone, and getting lots of input will help to make a decision that creates the greatest value overall.

Finally, consider pricing or underwriting. Why is it necessary to have only one underwriter review a case? Why not have three, or five? It should be clear when the answer is either yes or no. But for the questionable cases, more minds in the room at the same time, reviewing the same facts, and pulling their own individualized knowledge together can result in enhanced decision-making.

Conclusion

Using the Wisdom of the Crowd is an easy way to provide three things: a quick check of an expert's opinion, a reasonable range of results, and feedback as to whether there may be latent issues surrounding business practices which just have not yet surfaced. Companies would do well to create a forum for independent, honest feedback, with some kind of incentive to perform, in order to expand the ways they aggregate little bits of knowledge into actionable forms. Additionally, associates should be given some form of incentive to participate, in order to encourage them to make their best prediction, rather than just throwing anything and everything at the wall to see what sticks. Hopefully using the Crowd will become a valuable tool to the future Actuary as a complement to a software model, and the results will enable not only Actuaries but their employers, clients, and regulators to make informed decisions, rather than emotional reactions.