RECORD, Volume 27, No. 2*

Toronto Spring Meeting
June 20–22, 2001

Session 60CS
Using Case Studies in Actuarial Education

Track: Education and Research

Moderator: STUART KLUGMAN
Panel: FRANK G. REYNOLDS
ANA STANGL
TRACI J. LILLY
WIL CHONG

Summary: The case study method has become an integral part of actuarial education. It is currently used extensively in Course 7 and is featured in issues of the North American Actuarial Journal. Unlike a traditional textbook, a case study can accomplish the following additional educational objectives:

- *Integration—the problem can be analyzed from beginning to end using a variety of techniques*
- *No unique correct answer—students can learn that not only is the answer not in the back of the book, but it is not unique, and in particular, may depend on the audience and the time and resources available*
- *Involve students in ways that textbook exercises cannot—solutions are often obtained by working in groups and presentation skills can be enhanced*

MR. STUART KLUGMAN: I teach actuarial science at Drake University in Des Moines, IA. We are small group so please introduce yourselves.

MR. FRANK G. REYNOLDS: I teach at the University of Waterloo, in Waterloo, ON.

MS. ANA STANGL: I work at ING in Des Moines, IA.

MS. TRACI J. LILLY: I work for General Cologne Re in Stamford,CT.

MR. WIL CHONG: I work for The WMA Corporation, a reinsurance company in Atlanta, GA.

MR. KLUGMAN: This is a hybrid presentation. Some of it comes from work in Course 7, where we use case studies, and I'll talk about that a little bit. It also comes from my role as case study editor for the *North American Actuarial Journal*.

In both we are trying to promote the use of case studies and to give you an idea of why they're good things and how they can help you learn. Also, the SOA's new Course 8 always features a case study as part of the educational material.

If you're wondering what a case study is, here's a definition. It comes from a book on teaching the case method. If you want to engage in case study teaching yourself, this is an excellent resource with a lot of ideas about how to be good at it.

"A case is a partial historical clinical study of a situation which has confronted a practicing administrator or managerial group. Presented in narrative form to encourage student involvement, it provides data—substantive and process—essential to an analysis of a specific situation, for the framing of alternative action programs, and for their implementation recognizing the complexity and ambiguity of the practical world." [C. Christensen and A. Hansen, *Teaching and the Case Method* (Boston: Harvard Business School Publishing, 1987, p. 27)]

That's a long sentence, but that's a definition of a case study and we'll talk a little bit more about what that might mean as we go along.

What's so great about all this? Case studies can do a lot of things: stimulate discussion, and I hope in my sample case study today we'll be able to do just that,create teaching moments that show up sometimes by surprise, sometimes intentionally and reinforce the notion that there aren't unique answers. In Course 7, that's the hardest thing to get our actuarial students comfortable with. There's no book, and in particular there's no back of the book with the answer in it. In fact, the answer you get doesn't have to be the answer I get and we might still like it. It's the logic of your thought process and your ability to analyze a variety of facets of the situation that are important. Matching the exact numerical answer is not critical.

One of the case studies we use in Course 7 has as its endpoint a number. We have students work on it overnight. In the five Course 7s I've done we've had five different answers from the five groups that prepared a solution, yet none of them were wrong. All of them had done very reasonable things to arrive at their number. Only one of them actually matched the number that was produced by the person who created the question and some of them got what I thought were better numbers.

For actuarial education, I like to think that this, versus a multiple-choice problem or a memorized list is good learning. Having you solve unstructured problems is the best way to go from the teaching perspective, education perspective, and learning perspective.

MR. FRANK G. REYNOLDS: There's a certain extent to which I can see where you're coming from, Stuart, but the problem that I see is the one that is encountered by every other profession. Namely, there's a basic body of factual material that must be memorized and then having memorized it, people must then learn how to use it. Most professions differentiate the two.In medicine, the first two years is usually straight memorization of fact, followed by two or three years of clinical studies where students start learning how to apply these facts. What we seem to be trying to do now is to say, "Let's not learn anything, let's just start applying it." How do you expect people to be able to remember what they never learned when they're faced with a factual situation? I've seen this. I can remember one person years ago coming into my office and demanding a systems project be approved on short notice. It turned out afterward he got himself into real hot water, because his boss complained to mine. I had to point out that the whole thing fell apart. Why? Because he had never bothered to learn the basics. He'd done this type of learning where you don't really memorize anything.

MR. KLUGMAN: With regard to actuarial training, that's why the case studies don't show up until Exams 7 and 8. We still, at the earlier part of the process, do things in a very traditional way. Similarly, in a lot of business schools there's still the technical material, particularly in accounting and finance and some of those areas that come first. Also, at our current Exam 8, as I understand it, we are not abandoning lists and multiple choice, but we also want to do something to move our students towards feeling comfortable with working with less structured situations. In Course 7 we make sure they have the technical background that goes with it. We don't expect them to invent solutions out of nothing because that would not be efficient. We make sure that they have the grounding in the subject matter on which the cases are based. A good Course 7 session does not allow students get to that point unless they have the background to do a good job of it.

MR. REYNOLDS: I guess I'm coming more from the Course 8 perspective where they don't have the material that they're trying to learn for the first time, the basic professional material. To me the early exams are the technical work, and now we're getting into Course 8, an introduction to what used to be two-thirds of the material. Before they even get the introduction to Course 5 straight, we're asking them to go to this learning process. Don't take me wrong, Stuart. I can see why there's a need for this innovation, but I'm afraid we've done the usual trek of throwing the baby and the bassinette out with the bath water.

MR. KLUGMAN: And I wouldn't be surprised. Our experience with examinations has been when we asked students at this level to do those things, they do very poorly because they're not ready to do well. I do know when we put open-ended questions on the life contingencies exam, which we tried at least two times in our history, they were abandoned because the scores were so dismally low it was frustrating for the graders to write questions that people weren't ready to handle.

MR. REYNOLDS: Well, I come from a background of having had the licensing exam as the Chairman of Education & Examination. We used to find that the biggest problem was sitting there for another day and a half because the people who'd ask the cognitive-type questions couldn't agree on the answers. When the student had written something down, they couldn't even agree on what it was worth. It used to take them twice as long as anybody else to do it. The cognitive-type questions didn't even correlate with the two thought-type questions.

MR. KLUGMAN: Let me also add, part of my promotion of case studies today is the continuing education perspective, particularly through the *North American Actuarial Journal* and some other things I'll mention later.

One thing that I do believe is true, and this may not be for beginning actuaries, but when I see good actuaries doing good work, I'm continually amazed at their ability to provide inspired and efficient solutions to unstructured problems. They pull the facts together, make do with what's available, and come up with something pretty darn clever in the end. I see that skill over and over again. Actuaries develop either because of or in spite of our education system. Let's not debate that one.

Why should cases work well? Well, they force you to think in terms of models. You have to take the complexity of the real situation and simplify it. Actuarial science has turned into the science of building models.

Three of our exams have modeling or models in the title: Actuarial Models, Actuarial Modeling, and Seminar in Applied Modeling. And we all use models; the life table is a model. The key principle used so often is expected present value, which is a model-based concept. When the health and casualty actuaries use credibility, there's an underlying model that drives the formulas. Anybody who generates interest rate scenarios has a model behind them that helps them do that work. We build large models to do asset liability management to forecast solvency.

How do we use cases in Course 7? First we do some background reading. Before the case begins, we ask the student to read something to give them some background. Usually they get these materials a month before the seminar, so they have some time to digest the material and practice some calculations on their spreadsheet so they're comfortable when they show up. Again, they can work on their own and try out a few ideas and to get comfortable. At the seminar, they'll have an interactive session with the instructor to make sure that they're not confused, answer any questions that remain from the preliminary reading, talk more about the question, but don't get the answer, learn a few more things, get comfortable with the problem, maybe isolate some of the issues, but then we send them to work in groups in the evening to actually complete the solution to the question that's been asked. One of the things they get to do is practice their teamwork and some presentation skills.

The next morning they have to present the solution to the full group. We not only criticize them on the technical work that they've done, but we also offer comments and praise on the quality of their presentation when appropriate. And we talk about presentation issues such as having white letters on a darker background. Most every session I've attended at this meeting, the presenters have used dark letters on a light background. We have a not-so-meaningful, but fun, debate on which one is better. I think in the end, most people decide that this looks a little nicer. We also discuss more serious presentation issues.

Part of a good modeling case study is to do all the elements of the modeling process. This includes: collecting data (which might just consist of prior knowledge); postulating a model; calibrating the model (collecting more data as needed); refining the model; using the model to answer the question (which may be a prediction of what might happen or an explanation of what's been happening); and then going back to the beginning and starting all over again.

One of the things we might like to do for our education system is what the Australians do. Instead of having a Course 7, like ours, that is one-week long, they do something that takes a year. They spread things out so that you have time to think and formulate solutions, which is much more time than we allow our students to do careful thinking. In that year they can work on a variety of problems from a number of application areas and spend time making sure their tools are sound and that they have the necessary skills. They do both teaching and case study work as they solve problems. They have time to do the feedback loop to get some results and then say, "Here's some more information; make your model better, revise your predictions to do the control that's part of an actuary's work."

We do it in some lines of our business more than others. Of course, for pensions it's automatically a part of the business as you do gain and loss analysis and adjust your contributions every year to keep the plan on track. We also do that for participating policies where we adjust our dividends every year, as need be, to keep things on track. We can do that in any modeling exercise where you can keep looping through and improving your model.

What's the problem we're going to look at today? It's based on a real problem. A good case study normally comes from something that really happened. The best case studies we've had in Course 7 are ones that came out of the presenters' own experiences. The very best ones are ones where somebody else first comes up with an ineffective solution with lots of errors so they could talk about traps they fell into and then various things they had to do to arrive at a good solution to their problem.

I found out about two weeks ago that this problem I worked on has entered the courts and so I've been forbidden to tell anyone about it, so I, unfortunately for today, have to abstract the problem. I've had to turn this into a statistics problem with no content and I apologize about that. I didn't have time to think up another good problem that you might enjoy talking about. Here's the problem. You have 400 urns. Each urn has some marbles in it and we know how many. We do know that some of the marbles are blue and some of the marbles are red, but we don't know how many blue and red there are. For some reason it's important to know which of those 400 urns have 90% or more blue marbles in them.

Your task is to design a sampling scheme, because some of these urns have tens of thousands of marbles and you don't have time to count them all. The sampling scheme uses as small a number of draws as possible, because let's assume that getting a marble out of the urn and recording its color is expensive and time consuming. Your task is to design a sampling scheme to, with minimal errors, classify these 400 urns. What should we do?

MR. JOHN C. MICHAEL: Could we start out by at least taking an overall peek inside an urn to get an idea if it's mostly blue or mostly red?

MR. KLUGMAN: Good. One of the problems here is that we're approaching this case study from total ignorance about what's in these urns, about the only way we can have a clue about what's going on is to look at a few. We know we're going to have to look at some anyway, and if I were to make this a wonderfully interactive case study, which I didn't do, but it would be kind of fun to design, we could have the challenge of seeing who can classify the most urns correctly at the least cost. You would come up and say, "I want a sample of size six." I'd use a random number generator that gives you your sample and you'd go off and do your calculations and you'd come back and give me your final scheme. Then you'd make your decisions and we'd see what kind of score you got based on the number of correct decisions and the number of balls sampled. One thing we can agree on for this problem, the only sensible way to make our decisions is to sample a certain number of marbles from each urn, and if at least so many from the sample are blue, we'll declare that urn to be 90% or more blue. If it's under that number, we'll say no.

One of the other things we do in Course 7 is talk about data issues and data quality. That's not a part of this case study, because when we think about it, the only data issue would be to make sure that we correctly recorded the color of each marble as we drew it out, so that's not going to affect our work.

Anyway, here are the results of the sample. Remember, there are 400 urns; we took five balls out of each of them; 155 of the samples of size five had no blue marbles; 27 of them were one blue and four red; 17 were two and three; 26 were three and two, 28 were four and one, and finally 147 of them had all five blue. What do you observe? At least in our samples, there is a strong tendency to get either all blue or all red. We need to wonder whether or not that might help our work.

Tell me from a statistical perspective, what assumptions can we make about the data collection process? Think about some terms from your statistics education that might apply, things that we can make be true or at least close to true about our sampling.

MS. LILLY: It's got to be randomized.

MR. KLUGMAN: It's got to be random and we need to be confident that we're truly drawing the balls out at random. I've worked on sampling problems for third-party health administrations where the issue was to sample claims and have them carefully audited to see how accurate the administrators were. When you do that, you need to make sure that you're drawing them at random and that you've got a viable scheme for going into the records and giving each claim an equal chance of being selected. Here we would hope we can find some way to reach into those urns and give each marble a fair chance of being selected. It's not always done in the best way. The most famous example of nonrandom sampling is the 1970 draft lottery in the U.S., where the birthdays were not selected at random and people born in January had much less chance of being awarded a low draft number than those born in December, because they didn't actually ever mix up the capsules.

That translates effectively in statistical terms to the observations being independent. Whatever you get on one draw has no impact on what you get on the next draw. We'll assume then that there's a probability (*p*) that it will be blue and that probability is fixed for a particular urn, but could vary from one urn to the other. Tell me what probability distribution are we talking about when we have a fixed number (*n*) of draws and we're counting the number of blue marbles and we have a fixed probability (*p*) of each draw being blue?

MS. STANGL: Binomial?

MR. KLUGMAN: Yes. Now, for those of you who want to be technically, exactly, and perfectly correct, you might have come up with a slightly different answer that I know you all knew once upon a time. We might ask, is it actually hypergeometric? What's the difference? Each time you draw a marble, the probability isn't quite *p* for the next one because you've reduced the total number of marbles by one. Depending on what you got, the blue or red have diminished by one, and so the true distribution is hypergeometric. I made a decision and my decision was this: 1) I told you there was a fairly large number of balls in each urn, so that effect is probably small. 2) Let's assume I'm under a time constraint and the hypergeometric distribution is a lot messier than the binomial distribution and I may never be able to complete my work if I have to work with the hypergeometric distribution. I also know that if you have hypergeometric variables, the variances are smaller, so by using a binomial distribution with a larger variance, I'll be overstating any accuracy that I come up with in my results, so I'm taking a conservative approach to building the model.

MR. REYNOLDS: Stuart, at this point I'll ask a question. I can see what you're doing. Why would you have not picked one marble from each urn, got an overall estimate of *p*, done your distribution work, and said that given whatever end is necessary to get a 95%confidence level, and worked from there? It would seem to be much simpler than what you are doing, although I'm not saying that what you are doing is wrong. I'm just simply saying it's quicker and cheaper. You've obviously got a reason.

MR. KLUGMAN: Good point. Let's look at one way we might proceed with our analysis that is equivalent to the confidence interval approach that you just suggested. We'll see that it's quick in the sense that we're going to solve the technical problem quickly, but we will end up taking big samples with this approach. This is the one-at-a-time approach. We could think of this as 400 separate problems, because each urn has a different proportion of blue balls in it. I could treat this as 400 unique problems and we know that there's a hypothesis test ready to go for the particular problem. Either the proportion of blues in the urn is above or below 0.9 and we're asked to choose, so this problem was set up for a

hypothesis test because there are only two things to choose from: either we believe there's more than 90% or less than 90%.
The only thing we have to resolve is: Which one of them is to be the null hypothesis and which one of them is to be the alternative hypothesis? We know they're treated differently by the hypothesis testing method. Unfortunately, we can't answer that question with the way I've set up the case study because we don't know which of the two errors we are least willing to make. I'm going to have to tell you which one it is. We want to control the probability of declaring the urn to be under 90% blue, when in fact, it's over 90% blue. That's the error we don't want to make, or at least we want to control it at an acceptable level. That leads to a null hypothesis greater than or equal to 90% blue and the alternative that it's below 90% blue. The Type I error is the one we want to control. If we set it up at 5% significance level and do the standard test based on the normal approximation to the binomial, we'll find that if we sample n items, we should accept the null hypothesis and declare that we have 90% or more blue marbles in the urn if the number of blue marbles in the sample exceeds $0.9n$ minus 1.645 times the square root of $.09n$.

Suppose we ultimately take a sample of size 10. We would declare the urn to be 90% or more blue if there are eight or more blue in the sample. We know that the Type I error is about 5%. If I want to look at the Type II error, suppose the urn is 80% blue. I then have a 67% chance of making an error. That's kind of large. What if I made n 100, then I could get it down to about a 10% chance of an error, which might be reasonable. That's 100 draws from 400 urns or a total of 40,000 balls to look at. Remember, I said it was very expensive to do this. This scheme will work. It's statistically sound. The model is sound; it's just not particularly satisfactory. Is there something better we can do?

We have to ask a basic question: Is controlling the Type II error what we're really interested in? Did I ask the right question? I charged into a hypothesis test because all my statistics training and my Exam 2Course 110 background told me that's how you solve problems like that. We had a method that we were comfortable with and we went with it. Recall that we calculated Type II error probabilities for urns that are 80% blue. But our sample indicates that is unlikely. It is more likely that when p is below 90% it is closer to 10%. But what's the real question? The real question is, how many misclassification errors are we going to make on these 400 urns? How many urns are we going to get right and how many urns are we going to get wrong? And the key is we care about the actual 400 urns we have, not a set of hypothetical urns. The statistical analysis that was done was all hypothetical. If the urn looks like this, what's the chance of making a mistake? What if the urn looks like that?

What did the sampling tell us? Most of the urns are all blue, or a high proportion of blue, or have a high proportion of red. We're not going to make a lot of errors if we really believe that's true. We won't make the Type II error very often because when the probability isn't above 90%, it's probably closer to 10% rather than 80%. There aren't a lot of cases in the middle. The ones that are going to cause errors and we're going to have difficulty classifying are the urns that are mixed, those which have a fair number of red and blue in them. Most of the urns appear not to be that way.

What do we need? We need objective criteria and with luck, a smaller, yet effective, sample size. Here's the goal: determine a method that would make each kind of error 5% of the time on our 400 urns, that is admitting that we could have 20 urns wrong in each direction.

We're going to average our Type II errors. Instead of saying, let's control the Type II error for a particular urn that might have 80% blue in it, let's talk about the Type II errors for the kinds of urns and their proportion of blues that we actually have. What we need is a model for the distribution of p over the population of urns. What do those ps look like? That's where our sample helps out. And we need more than one from each one to take this approach to get a feel for how p is distributed from urn to urn and then maybe five is enough, maybe it isn't, but we'll address that issue in a moment.

Hypothesis Test?

- Is the control of Type II error what we are really interested in?

- Required sample sizes are much too large to be practical.

- What do we need?
  —A more objective criterion.
  —A smaller, yet effective sample size.

What are some of the desired properties for a model for p? Well, the distribution of p should go from zero to one. It should be continuous because any number between zero and one should be possible, and it probably should be u-shaped, that is with high probability on p values close to zero and close to one with low probability in the middle. Anyone remember the name of the distribution that has those characteristics, that the possible values range from zero to one continuously and it can be u-shaped? It's the beta distribution.

This is a beta distribution when the parameters at $a=0.3$ and $b=0.4$. By changing a and b, you can make it more u-shaped or less u-shaped, that is, how high the ends go versus how low the middle is. If a and b are equal to each other, the density is symmetric. In our example, with b being slightly larger, you'll notice that the u is higher on the left than on the right. Here's what I'm going to do for assumptions. Assume the 400 different p values are independently drawn from a beta distribution.

A Better Criterion

- How about a method that will make each kid of error 5% of the time?

- Note that the Type ii error probabilities are then averaged over the various true *p* values in our population of 400 policy forms.

- We need a model for the distribution of *p* over the population.

We could get the answers analytically, but time is short and I might get the integrals wrong. Instead, I will use simulations. This is done in the Excel worksheet. I arbitrarily selected $a=0.3$ and $b=0.4$, so this matches the graph that we had before. What does my simulation sheet do? First select *a* and *b*, and then select the sample size. I have used 10 here. Then a decision rule is needed. I have decided to declare that the urn is 90% blue if six or more in the sample are blue. Then I have 400 rows and in each I first simulate *p* from the beta distribution using *p*=BETAINV(rand(),.3,.4) and then I use *y*=CRITBINOM(10,*p*,rand()) to simulate a binomial sample from that distribution. Both simulations use the inversion rule. So in Urn 1 we sample 10 and got 10 blues. That urn would be classified as above 90% based on the sample. The actual *p* of 0.985064 is above 90% and so no error was made. In row 5 the true *p* value is 0.579756, but when 8 of 10 in the sample turn up blue, it is classified as more than 90% blue. As a result, a Type II error is made. In this simulation of the decision rule, I made one Type I error and 81 Type II errors for an overall error rate of 20.5%.

Now, I can play around with that. What if I go up to eight of 10 before declaring the urn to be over 90% blue? Now I make more Type I errors. As you expect, if I make more Type I errors, I'm probably going to make fewer Type II errors. I can play around with this, and being this is one simulation, I'd probably want to repeat the simulation more than one time to see what kind of changes you get. I could try various combinations until I discover a sampling scheme that produces error rates that I'm comfortable with. Again, I could actually calculate the expected number of errors analytically, but it's a little bit painful, and it was very easy to construct this spreadsheet to do the simulations so it's more a question of how much time I have and how confident I am in my ability to get those expected values.

What's the missing ingredient? What do I need to do to complete my analysis? I need to estimate *a* and *b*. I just made up the numbers we have been using. That's probably not a sound way to go. Now we really have to use our data. That was a good reason for collecting it. How are we going to estimate *a* and *b*? Let's try the method of moments. There are two parameters; and two moments gives you two equations and two unknowns. The equations are on the slide. The solutions are $a=0.155$ and $b=0.159$. But, we don't have very good methods of assessing the accuracy of method of moments estimators, so we might want to switch to maximum likelihood estimation. The likelihood function is kind of ugly for this problem and I did have to do an integral to make that work out. The point of our case study isn't to wonder about the technical aspects, but at some point we had to do that. One of the things that you can learn in the new Course 4 is that it's possible, for maximum likelihood, to both use your Excel spreadsheet to maximize the likelihood function and get the standard errors.

Equation 1

Methods of Moments Estimation

- Calculate the mean and the second moment of the observed *y* values.
- Match them to the mean and the second moment for our model.
- The equations are:

$$\bar{y} = E(Y) = E[E(Y|p)] = E(mp) = \frac{ma}{a+b}$$

$$\frac{1}{k}\sum_{j=1}^{k} y_j^2 = E(Y^2) = E[E(Y^2|p)] = E[mp(1-p)+(mp)^2] = \frac{mab+(ma)^2}{(a+b)(a+b+1)}$$

The maximum likelihood estimates are essentially identical to the method of moments estimates, which gives us confidence that I've executed the two methods properly. In addition, I have standard errors, so the estimate of *a*, plus or minus a couple of standard errors, tells us that we are within 0.019 of the true value. I could use that range to do sensitivity analysis.

Equation 2

Maximum Likelihood Estimation

- Maximize the product of the density function evaluated at each data point.

$$L = \prod_{j=1}^{k} f(y_j) = \prod_{j=1}^{k} \frac{m!\Gamma(a+b)\Gamma(a+y_j)\Gamma(m+b-y_j)}{y_j!(m-y_j)!\Gamma(a)\Gamma(b)\Gamma(a+b+m)}$$

- The values of *a* and *b* that maximize this function are the estimates.
- Second derivatives can provide estimation error.

Table 1

Results of Sampling

K	*a*	*Stdev(a)*	*b*	*Stdev(b)*
50	.056	.025	.085	.039
100	.109	.029	.117	.032
150	.137	.028	.146	.030
250	.167	.026	.176	.027
400	.153	.019	.159	.020

The other thing I can do that I know you all remember is a chi-square goodness-of-fit test to see if the distribution fits. It did fit quite well here. I got a test statistic of 2.0, with 3 degrees of freedom. That's a 0.55 *p*-value. We will easily accept the null hypothesis that this beta-binomial model fits our data well, so that also produces some confidence that we've done a good modeling job.

What would happen is, if we were doing this at a Course 7 seminar, the technical tools for doing the problem would have been presented. The students would have had a chance to practice and ask questions and would have those formulas and the technical skills necessary to deal with this issue. What we would probably do to keep their exercise simple is bring them to this point with our help and then say, "Now you have all of this at hand, what size sample are you going to recommend be taken?" And then write up a report that persuasively demonstrates that you've done an appropriate job of analyzing the problem and can convince someone that your sample size is reasonable. That's the way we would do the case study in Course 7. We would not have the students go out and do the integrals and the maximum likelihood estimation because we don't have time. We think they may have seen the ideas; that they could probably do it if they have enough time, but we want them to concentrate on the last step. They're getting a reasonable answer and reporting the results and then presenting it in an effective manner to the rest of the group.

Where does that leave us? We have the estimates for *a* and *b*, and we've estimated the standard error. We're almost there. Now we can go back to our spreadsheet, and as I said, we now have the tools at hand to help answer the question. With these *a* and *b* values, a sample size of 10, and 8 blues needed to declare an urn 90% or more blue, we have an overall error rate of about 7%. Happy? Again, I can recalculate and you'll notice each the simulations are hanging around 6-8%.

We might say, "We've already taken a sample of size 5; maybe we just don't need to do any more." What if you take samples of size 5, and if four or more are blue, you go with a 90% blue statement. We have an error rate of about 10%. And notice generally we've just made zero, one or two Type 1 errors, so one in about 400 on average are misclassified with the error we care about. It looks like our sample of size 5 might get the job done, so by careful analysis of the problem and by applying a few statistical facts, we've cut out sample down to five. That's 2,000 balls to look at from the 400 urns, to do a pretty good job of classification. Ten will do, maybe even less.

What have we learned? Thinking about the model and understanding its consequences helps by doing a little bit of lateral thinking. By not going with the first impulse that we've learned, we might be able to come up with a better solution. Having some data in hand helped a lot.

Now, if you're taking the new Courses 3 and 4, and particularly if you look at the credibility material, you'll find that none of the calculations that I did are beyond what's taught there. For any of you who took exams, which I would guess is everyone, under the older system, there's a good chance you did not see everything that I talked about. You would have come across the beta and binomial calculations if you did the old 4B exam on credibility, but that's about the only place that it appeared on the prior syllabus. There also wasn't as much emphasis on practical aspects of maximum likelihood estimation. This still isn't on the exams, but at least the reading material gets people to understand that numerical maximization on your spreadsheet can get the job done. You're probably more accustomed in your studies to only solving problems where you could take the derivative, set it equal to zero, and by luck, you've been handed a problem for which you could solve the equation, which you couldn't have done in this problem.

What can you do? If you like case studies, the Actuarial Education & Research Fund is sponsoring the creation of case studies. You can contact Judy Yore at the SOA Office if you'd like more information. I've also mentioned that the *North American Actuarial Journal* will have its first case study appear in the October issue and we hope to have one or two in each issue after that. If you'd like to learn more about that, you can contact me.

In the *Journal* we're going to encourage readers to submit alternative solutions to the ones that are presented in the article. We may even at times publish the question in one issue and publish various solutions in a subsequent issue. The idea is to again reinforce the notion that solutions aren't unique to the kind of problems we'd like to pose.