STATISTICAL THEORY OF A SIMULATION SCHEME FOR DETERMINING THE SURPLUS NEEDS OF A LIFE INSURANCE COMPANY

by

Aaron Tenenbein

New York University

The material presented by Prof. Tenenbein may be found in:

1. Johnson, M.E. and Tenenbein, A.:
   "A Bivariate Distribution Family with Specified Marginals", JASA Vol. 76 no. 373, March 1981.
2. Tenenbein, A. and Gargano, Michael:
   "Simulation of Bivariate Distributions with Given Marginal Distribution Functions" (Chapter 18 in 'Current Issues in Computer Simulation'), Academic Press, Inc., 1979.

The Discussion of Prof. Tenenbein's presentation follows:

VOICE FROM FLOOR: What do you mean by rank correlation.....

AARON TENENBEIN: Correlation in the usual sense refers to the Pearson product-moment correlation coefficient. The rank correlation for sampled data involves taking the ranks of your paired observations and finding the Pearson Product Moment Correlation between the two sets of variables. ρ_S , which I talked about, really represents the population measure upon which this is based. It is called the Spearman's rank correlation coefficient. It can be estimated from some data by merely taking the ranks of the two random variables and computing the correlation coefficient between them.

VOICE: I am interested in the population parameters of the distribution - if you have got two ranked variables.

AARON TENENBEIN: Basically it's a little hard to explain in the population but the idea is as follows:

There are two measures of rank correlation that we can talk about. There is τ and there is ρ_S . Now let me explain what τ is, ρ_S is a little more difficult to explain. τ is sometimes known as

Kendall's τ. If you start with (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3), basically what τ does is: first, find the probability of a concordance, which means the probability that $(x_1-x_2)(y_1-y_2)$ exceeds 0; the idea being that if two variables are positively related then this probability is one half because it is equally likely for them to go in either direction. If they are perfectly related, i.e. if y goes up when x goes up, this probability has to be equal to one. If you form two times this minus one you get τ; if they are independent then $\tau = 0$ (since $2 \times 1/2 - 1 = 0$). If they are perfectly dependent in the positive case (as x goes up, y goes up) then τ equals 1; if they are perfectly dependent in the negative case, then τ equals -1. Now this is what τ is. It is a little harder to explain ρ_s in a population sense, because it is 6 times the probability that $(x_1-x_2) \times (y_1-y_3)$ is greater than 0, minus three. And you can show that this has the same properties as τ. ρ_s is preferable, but τ can also be used in this scheme and I talk in the paper about the use of τ as opposed to the use of ρ_s as a measure of correlation. ρ_s is the more usual one; people have been able to identify more with a product moment correlation. Basically what you can show is that the rank correlation coefficient, the r_s value, will tend to ρ_s as n becomes large. That is what ρ_s is but it is a little bit harder to understand why that measures the correlation when you take a look at three pairs of random variables rather than two pairs.

VOICE FROM FLOOR: I think that what you have shown us is that when you go back to bivariate distributions, it satisfies certain properties.

AARON TENENBEIN: That's right.

VOICE: And how do you know that bivariate distribution is in some sense unique? If you don't are you going to consider all possibilities of the bivariate distribution?

AARON TENENBEIN: ... and you mean you are worried about the sensitivity? Is that it?

VOICE: Right. But there are lots of other distributions, not of your construction, which can be formed in other ways.

AARON TENENBEIN: That's a good point - that's a tough question to try to resolve. Basically what I looked at is really trying to scratch the surface of the problem; that is, trying to get an idea if there is any sensitivity. Most people just come up with one way of doing it and they usually use the normal distribution and they stop at that. They have not really attacked the sensitivity question. The problem that you raised is a very very important one which has really generated, and motivated, a lot of the work that I have done. Essentially the only way you can do that is perhaps to look at more G(t) values. What I tried to do is take a look at extreme cases, in order to gauge the sensitivity of the simulation model.

VOICE: But there are others, joint distributions, that satisfy those initial conditions that your method would never produce regardless of what data you put in.

AARON TENENBEIN: Well, the point is that my method produces the distributions in a systematic fashion. You can't really do a systematic sensitivity analysis by using the other distributions.

Johnston & Kotz have written a lot of monographs on univariate distributions, they also have one on bivariate distributions but the problem is that there is no real systematic way of simulating bivariate distributions with tried measures of dependence. This is what I really tried to do.