Combining Estimates

Thomas Struppeck, FCAS, ASA, CERA

Abstract:

The problem of combining two or more estimates into a single estimate appears in many applications, such as combining estimates based on paid losses and estimates based on incurred losses, or combining estimates for several accident years or lines of business into a single estimate. A methodology for performing such combinations which allows for correlation is described. An accompanying Excel spreadsheet illustrates the procedure.

Keywords: Reserving, Credibility, Solvency II, Ranges

1. Introduction

Actuaries often are faced with the task of combining two or more estimates into a single estimate. When estimating ultimate losses, there may be one estimate based on paid losses and another estimate based on incurred losses. As a second example, consider estimates for several lines of business that are to be combined into a single estimate. While these two examples are somewhat similar, there is a very important difference between them. In the first example, we have two different estimates for the same quantity. In the second example, we have estimates for different quantities. As we will see, in the first case we will actually be able to improve on both estimates. In the second case, we can only hope to not be much worse than the worst one (at least in terms of the width of our confidence interval).

Patel and Raws [PR] used a simulation technique to study the effects of different methods of combining estimates in the case where the estimates are for different quantities (example 2 above). We review their results in section 3.

The remainder of this paper is structured as follows:

- Combining two or more estimates for the same quantity
    - Precision vs. accuracy
    - An example: two estimators for one parameter
- Combining estimates for multiple components
    - Patel and Raws' simulation work
    - Sums of different component pieces
    - Why normal distributions?
        - Considerations in selecting correlations

- An example
- Concluding remarks

2. Combining Two Or More Estimates For The Same Quantity

First, consider the case where two independent estimates, $\hat{\theta}_A$ and $\hat{\theta}_B$, are available for an unknown parameter, θ. We will follow the usual notational convention of writing carats ("hats") over parameters to denote estimators of those parameters and of using Greek letters to denote unknown parameters. In a reserving context, θ might represent the unpaid claims on a block of business, and $\hat{\theta}_A$ and $\hat{\theta}_B$ might represent two different estimates of θ.

It should be noted that we think of θ as being a fixed (but unknown) number; it is not itself a random variable. On the other hand, $\hat{\theta}_A$ and $\hat{\theta}_B$ are often instances of random variables and as such have sampling distributions. We will follow customary notation and write, for example, $E[\hat{\theta}_A]$ for the expected value of (the sampling distribution of) $\hat{\theta}_A$.

2.1 Precision vs. Accuracy

The two terms "precision" and "accuracy" are often used interchangeably; however, there is a slight difference in their definitions. Accuracy refers to the proximity of the expected value of an estimator to the true value of the parameter (the "reference value" in the illustration), whereas precision refers to the spread of estimator values around its expected value. Figure 1, below, illustrates this for a normal density.[1]

[1] Used with permission under the terms of the GNU Free Documentation License, Version 1.2.

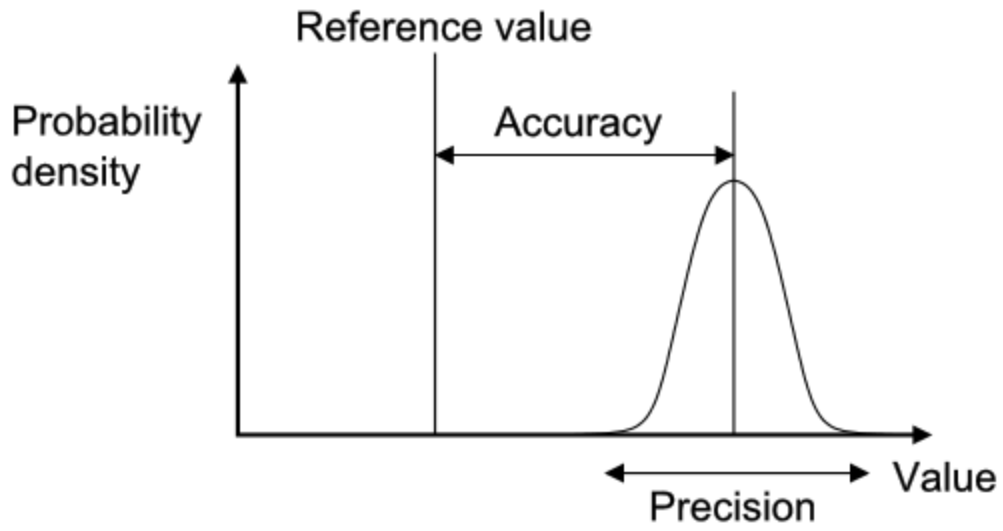

Figure 1: The distinction between accuracy and precision

2.2 An Example: Two Estimators for One Parameter

For our first example, we will attempt to estimate the outstanding loss reserve, θ. We have an unbiased estimator of θ, $\hat{\theta}_A$. For the moment assume that this estimator has a normally distributed sampling distribution with mean \$250 and standard deviation \$30. Then, since $\hat{\theta}_A$ is unbiased, a reasonable choice of point estimate for θ would be the expected value of $\hat{\theta}_A$, namely \$250. Since the sampling distribution is known[2], we can even give an approximate 95% confidence interval, \$250 ± 2(\$30) = (\$190, \$310)[3].

Now suppose that a second, independent, unbiased estimator of θ, $\hat{\theta}_B$, is available. Further suppose that the sampling distribution of $\hat{\theta}_B$ also is normal but with mean \$275 and standard deviation \$40. This second estimator has less precision than our first estimator (it has a larger standard deviation), and it suggests a different point estimate. Using just the second estimator we obtain another 95% confidence interval, \$275 ± 2(\$40) = (\$195, \$355).

We would like to create a single estimator that allows these two estimates to work in tandem in a way that maximizes the precision of the resulting estimator. The way to do this is to consider the one-parameter family of estimators obtained by taking weighted averages of $\hat{\theta}_A$ and $\hat{\theta}_B$:

$\hat{\theta}_t = (1-t)\hat{\theta}_A + t\hat{\theta}_B$ where $0 \le t \le 1$.

[2] In fact, we do not need to know the entire sampling distribution, only its 2.5th and 97.5th percentiles.
[3] We have used 2.0 instead of 1.96 for ease of exposition; about 95% of the area under a normal is within two standard deviations of the mean. The given interval is actually a 95.45% confidence interval.

Since we have assumed that both $\hat{\theta}_A$ and $\hat{\theta}_B$ are unbiased, each $\hat{\theta}_t$ is unbiased because expectation is a linear operator.

When t is zero, we get the first estimator, and when t is 1, we get the second estimator. In between, we get a family of estimators. Since each of $\hat{\theta}_A$ and $\hat{\theta}_B$ is normal, and they are independent, the weighted average $\hat{\theta}_t$ is normal with mean $(1-t)E[\hat{\theta}_A]+tE[\hat{\theta}_B]$, which in our case is $(1-t)*250+t*275=250+25t$, and variance $(1-t)^2Var[\hat{\theta}_A]+t^2Var[\hat{\theta}_B]$, which in our case is $(1-t)^2*30^2+t^2*40^2$.

We want to maximize the precision, which amounts to minimizing the standard deviation, which is the same as minimizing the variance. This is easily done by taking the derivative with respect to t and setting it to zero:

$$2(1-t)*30^2*(-1)+2*t*40^2=0\rightarrow(30^2+40^2)t=30^2\rightarrow t=\left(3\big/5\right)^2$$

So the minimum occurs when we let t=.36, i.e. we use 64% of estimator A and 36% of estimator B. This produces $\hat{\theta}_{0.36}$ which is normal with mean = 259 and standard deviation = 24. This estimator has the smallest standard deviation of any weighted average of our estimators and suggests a point estimate of $259 with an approximate 95% confidence interval of ($211, $307). This estimate has the most precision of any estimator in this class.

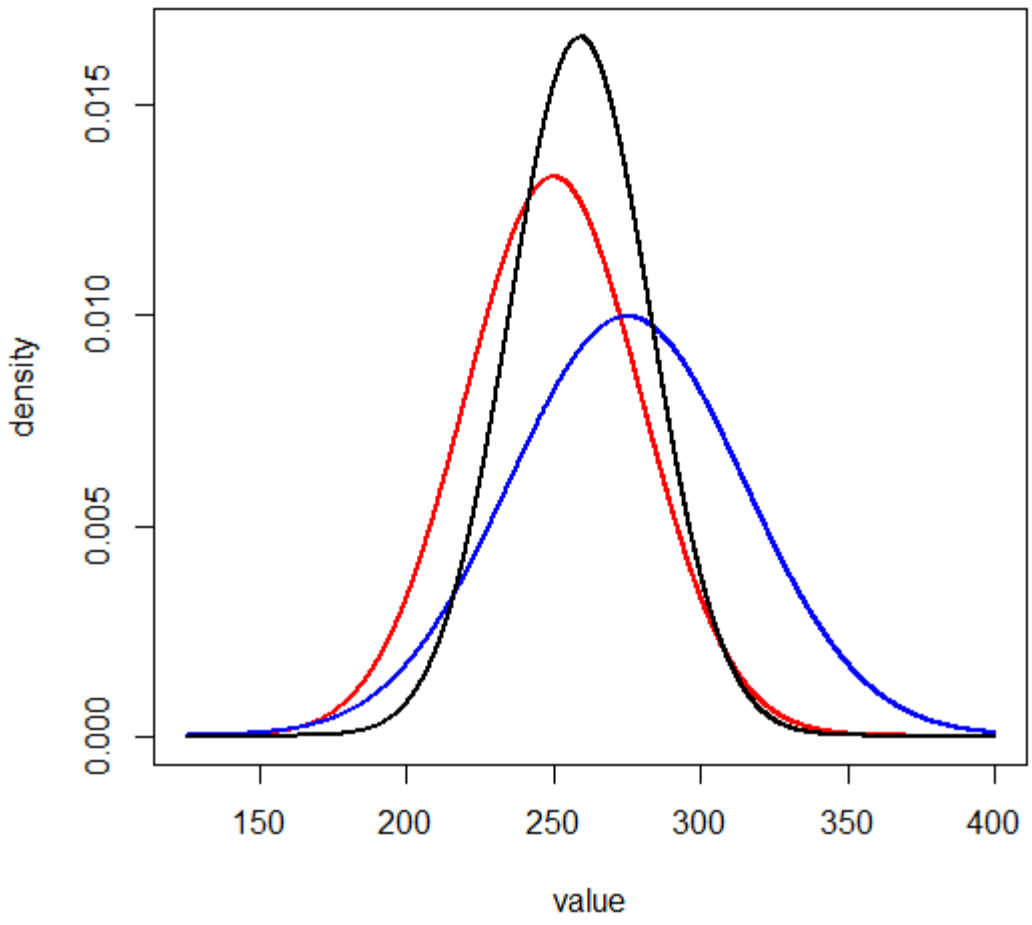

Figure 2: Densities for two estimators and their optimal combination

Figure 2, shown above, illustrates the density function for the two estimators, $\hat{\theta}_A$ (shown in red) and $\hat{\theta}_B$ (shown in blue), and the density function of their optimum weighted average, $\hat{\theta}_{0.36}$ (shown in black). The weighted average, having a more concentrated density, is more precise. Whether it is more accurate is a more delicate question.

Accuracy measures how far the true parameter is from our estimate. Since the true parameter is intrinsically unknowable, we will have to satisfy ourselves with a statement about confidence intervals. Since $\hat{\theta}_{0.36}$ lies between $\hat{\theta}_A$ and $\hat{\theta}_B$, whatever its true value, θ is closer to $\hat{\theta}_{0.36}$ than it is to at least one of $\hat{\theta}_A$ and $\hat{\theta}_B$. In other words, $\hat{\theta}_{0.36}$ is always more accurate than the worst of $\hat{\theta}_A$ and $\hat{\theta}_B$.

In fact, sometimes it is more accurate than either of them. Whenever $\theta > 254.5$, θ will be closer to $\hat{\theta}_{0.36}$ than to $\hat{\theta}_A$. And whenever $\theta < 267$, θ will be closer to $\hat{\theta}_{0.36}$ than to $\hat{\theta}_B$. Using the sampling distribution for $\hat{\theta}_{0.36}$, we see that the first interval is a 57.44% one-sided confidence interval for θ and that the second is a 63.06% one-sided confidence interval. The intersection of these two intervals is the interval $254.5 < \theta < 267$, which is a 20.49% confidence interval for θ.

In the example, we assumed that we had two estimators, that our estimators were normally distributed, and that they were independent. If we have more than two estimators, we can still find the optimum weighting by using multivariate calculus techniques (setting the gradient to zero, etc.). If the estimators are correlated and we have a good estimate of the correlation coefficient(s), we can still compute the standard deviation of the weighted averages and find the optimum weighting. In order to create a confidence interval, we need to know the distribution of the sum. If our summands are bi-normally or multi-normally distributed[4], then the sum will have a normal distribution. In that case, we can create our confidence interval in the usual way, namely by picking the appropriate point from a table of standard normal values (a z-score), multiplying it by the standard deviation and using this as a radius about the point estimate.

In practice, multiple estimates of the same quantity tend to be highly correlated. But, unless the correlation is 100%, some increase in precision will occur when they are combined.

3. Combining Estimates for Multiple Components

3.1 Patel and Raws' Simulation Work

In [PR], Patel and Raws considered the problem of estimating a total reserve from estimates of the component pieces. They used a simulation approach to compare several different possible distributions for the losses in each piece. Among the distributions that they examined were the uniform, triangle, normal, and log-normal distributions. In each iteration of the simulation, they generated losses by line from those distributions, summed them, and repeated this process many times to create the simulated distribution of the sum.

In their simulations, they assumed that the component pieces were independent. That assumption allowed them to select the losses for each line without having to explicitly correlate them. In the text they suggest that correlations between accident years could be adjusted for in the choice of distribution and that correlations between lines of business could be adjusted for similarly.

[4] It is possible for two or more jointly distributed normal random variables to have normal marginal distributions but not be multi-normally distributed.

Capturing the effects of correlation in a simulation is possible, but it can be tricky. If one is not careful, the marginal distributions can fail to be what is expected. Copulae can solve this problem.

A (d-dimensional) copula is a mapping from the d-dimensional unit cube onto the unit interval that is a joint cumulative distribution function with uniform marginals. Copulae are of interest to us because of Sklar's Theorem, which says that any d-dimensional random variable (i.e. d jointly distributed univariate random variables) can be expressed as a composition of its d marginal random variables and a copula that combines them. Furthermore, in the case of continuous random variables, this decomposition is unique. So, utilizing a copula, one can impose any possible correlation structure on a family of random variables, such as having correlation mainly manifest itself in the tails. More in depth discussions of copulae can be found in Mango and Sandor [MS] and Venter [V].

The ranges of estimates that Patel and Raws obtained did not vary greatly by choice of distribution. This suggests an alternative approach: instead of assuming independence and using multiple distributions, only use one distribution, but use one that allows for explicit correlations. One such distribution is the multi-normal distribution.

3.2 Sums of Different Component Pieces

The example in the first section illustrated how multiple estimates for the same parameter can be combined to create an estimate with greater precision than the original individual estimates. Often the quantity that we want to estimate is a sum of several parts, each of which has an associated estimate --- this is the problem that Patel and Raws examined. For instance, we could be interested in the total outstanding losses for a company that writes three lines of business, and we have estimates of the outstanding losses for each line.

We will not be taking a weighted average here, but rather we will just take a sum. The summing and averaging are closely related, but differ in an important way: in a weighted average, each summand gets multiplied by a number between 0 and 1, t and (1-t) in our example:

$$\widehat{\theta}_t = (1-t)\widehat{\theta}_A + t\widehat{\theta}_B \text{ where } 0 \le t \le 1$$

The variances of $\hat{\theta}_A$ and $\hat{\theta}_B$ got multiplied by $(1-t)^2$ and t^2, respectively, and these are strictly less than (1-t) and t (unless t is 0 or 1). This is how we obtained greater precision. There is no similar opportunity when the coefficient is equal to one as it is in a sum.

What this means is that as we add more and more independent pieces to our sum, our confidence intervals will get nominally larger.[5] By nominally larger, we mean that the width of the confidence interval will increase. In relation to the size of the reserve however, the intervals might be getting smaller. One possible measure of the relative size is the coefficient of variation (CV) --- the ratio of the standard deviation to the mean. The square of this, the ratio of the variance to the squared mean, is another measure which is sometimes easier to work with. It is called the squared coefficient of variation or SCV. Both the CV and SCV are dimensionless quantities.

When there is no correlation among the pieces and the number of pieces is large, the sampling distribution of the sum will start to look like a normal distribution--this is essentially the content of the Central Limit Theorem. Intuitively, we expect half of our pieces to be above their respective medians and the other half to be below their medians; when we add them all together the errors tend to cancel.

Often, however, there is reason to believe that there may be some correlation among the pieces. For instance, if we are estimating the total ultimate losses for a book consisting of several lines of long-tail business, we might model each of them separately. Future inflation (or deflation) might simultaneously increase (or decrease) each of those lines effectively creating correlation.

[5] If the pieces had very large negative correlations, it is possible that the confidence intervals could get smaller, but such instances are rare and would be quite unusual.

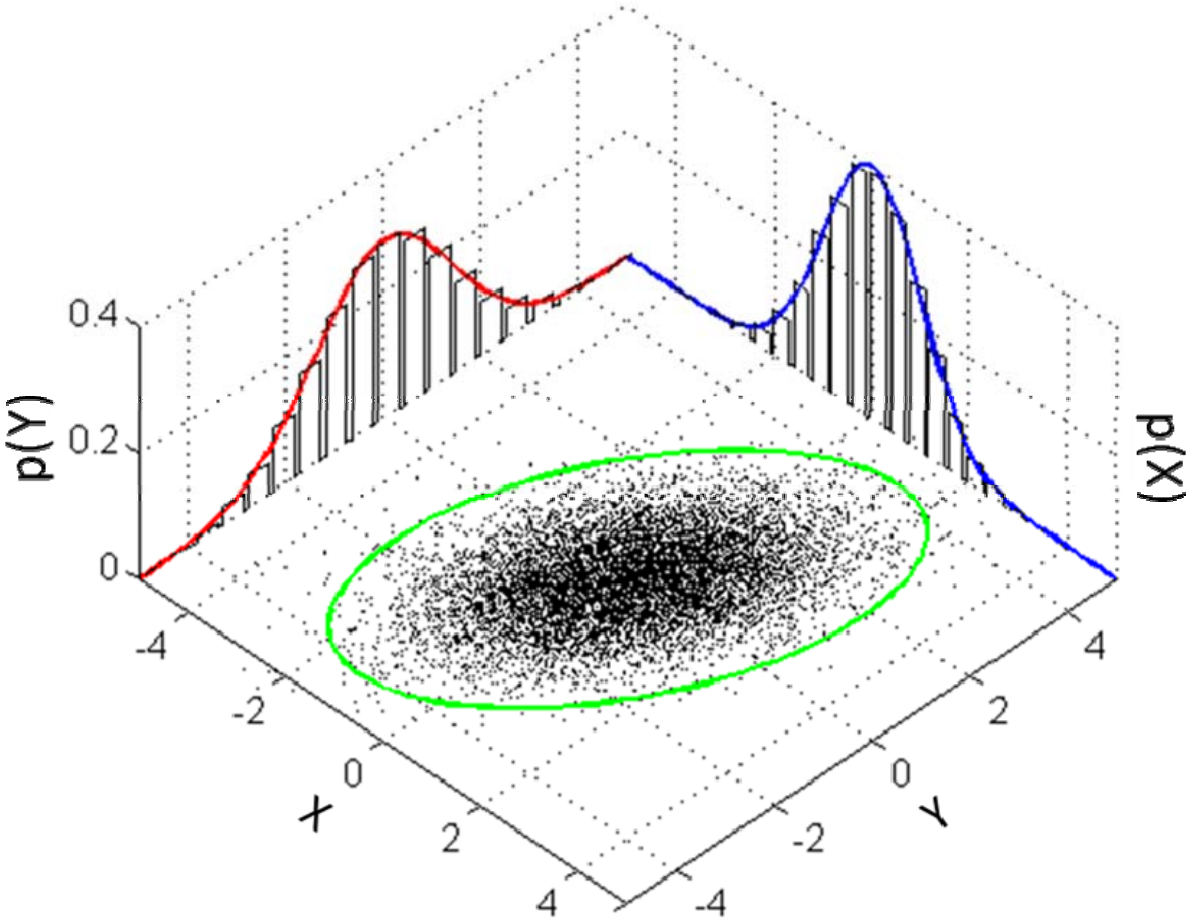

Figure 3: The marginal densities of a bi-normal pair of random variables

Figure 3, above, illustrates the marginal densities (in red and blue) of a bi-normally distributed pair of random variables[6]. The green ellipse represents a level curve of the joint distribution (in this instance, a 3-sigma ellipse[7]). If the correlation is positive, the distribution of the sum corresponds to the major axis of the ellipse (if the correlation is negative, the minor axis). This shows that the precision of the sum is smaller than the precision of the summands (unless the correlation is very negative).

[6] This image is in the public domain. The code to produce this graph can be found here: http://en.wikipedia.org/wiki/File:MultivariateNormal.png

[7] Symmetric confidence intervals for univariate normal variables have the property that the density function is equal at each end of the interval, i.e. the two endpoints of the interval form a level set of the density function. The analog for bivariate or multivariate normal random variables is to use a level set of the density function as the boundary for the analog of a confidence interval--a confidence region. In the case of bivariate normals, these level sets are ellipses. The green one shown corresponds to three standard deviations, i.e. approximately 99.7% of the probability is inside it.

Consider the following example: the company writes three lines of business, A, B, and C. You have estimated the outstanding losses for these three lines of business and selected ultimate losses for each. Your estimated outstanding losses along with some estimated ranges are:

Line of Business	Expected Losses	25th-percentile Losses	75th-percentile Losses
A	100	90	110
B	225	150	300
C	350	200	500

Further suppose that we believe that these estimates of the unpaid losses for lines A and B are 50% correlated, for lines B and C are 60% correlated and for lines A and C are 50% correlated.

Notice that we have selected our range of estimates for each line of business to a symmetric confidence interval (best estimate is in the center). This is because we are going to represent all three lines by a multi-normal random variable with mean vector, $\vec{\mu}$, and variance-covariance matrix, Σ.

First, we select the marginal distribution for line A. It is going to be normal with mean = 100 (the expected value) and some standard deviation, σ_A. We know that the 25th-percentile is 90, which is 10 less than the mean, so using the standard normal table we conclude that σ_A must be 14.82. This technique is just the familiar method-of-moments.

Similarly, for line B we obtain mean = 225 and σ_B = 111.2, and for line C, mean = 350 and σ_C = 222.4.

Since the expected value of a sum is the sum of the expected values, the mean for our total will be 675.

To create the variance-covariance matrix, we start with the correlation matrix and multiply it on both the right and the left by a diagonal matrix with the respective standard deviations down the diagonal.

The correlation matrix:
$$\begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.6 \\ 0.5 & 0.6 & 1 \end{pmatrix}$$

The variance-covariance matrix:
$$\Sigma = \begin{pmatrix} \sigma_A & 0 & 0 \\ 0 & \sigma_B & 0 \\ 0 & 0 & \sigma_C \end{pmatrix} \begin{pmatrix} 1 & \rho_{BA} & \rho_{CA} \\ \rho_{AB} & 1 & \rho_{CB} \\ \rho_{AC} & \rho_{BC} & 1 \end{pmatrix} \begin{pmatrix} \sigma_A & 0 & 0 \\ 0 & \sigma_B & 0 \\ 0 & 0 & \sigma_C \end{pmatrix}$$

The variance-covariance matrix for our example:

$$\Sigma = \begin{pmatrix} 14.82 & 0 & 0 \\ 0 & 111.2 & 0 \\ 0 & 0 & 222.4 \end{pmatrix} \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.6 \\ 0.5 & 0.6 & 1 \end{pmatrix} \begin{pmatrix} 14.82 & 0 & 0 \\ 0 & 111.2 & 0 \\ 0 & 0 & 222.4 \end{pmatrix} = \begin{pmatrix} 219.6 & 824.0 & 1,648.0 \\ 824.0 & 12,365.4 & 14,838.5 \\ 1,648.0 & 14,838.5 & 49,461.8 \end{pmatrix}$$

The correlation matrix is easier to interpret than the variance-covariance matrix. One reason for this is that the correlation matrix is dimensionless, meaning that it has no units, whereas the variance-covariance matrix has units, in this case square dollars. It is generally easier to work with dimensionless quantities when possible.

Line of Business	Expected losses	25th-percentile losses	75th-percentile losses	St. Dev. (Est.)	Estimated CV
A	100	90	110	14.8	0.148
B	225	150	300	111.2	0.494
C	350	200	500	222.4	0.635
Naïve Total	675	440	910	348.4	0.516
With Covariance Adjustment	675	465.3	884.7	310.9	0.461

The range for the "naïve total" is obtained by summing the endpoints of the intervals for each line of business. This corresponds to comonotonicity[8], which in the case of normal random variables means 100% correlation. See, for example, [S].

The last line labelled "with covariance adjustment" shows the 25th- to 75th-percentiles for the sum using the given correlations. The term covariance adjustment is taken from the US Statutory Risk Based Capital (RBC) calculation.

This calculation is easily reproduced using only Lines A, B, and C in the accompanying Excel spreadsheet, the use of which is described in detail later in this paper.

[8] Incidentally, the copula corresponding to this is the upper Fréchet–Hoeffding bound.

3.3 Why Normal Distributions?

Normal distributions are not the only distributions with the property that they are closed under addition[9]. Gamma distributions and many other families of distribution also have this property, at least when the summands are independent and possibly with some restrictions on the parameters. We choose to use normal distributions to approximate the individual distributions that we are going to combine because:

- Many naturally arising estimators have large-sample normal distributions. In particular, large-sample bootstrap estimators are asymptotically normal.
- The Central Limit Theorem suggests that averages (and hence sums) of independent observations will become normally distributed when the sample sizes get large.
- We can easily incorporate correlation and interpret it.
- It tends to produce results that seem reasonable.

3.3.1 Considerations in Selecting Correlations

Correlations are related to the R^2 statistic that comes from performing a simple regression of one of the two variables on the other. This statistic gives the proportion of the variation in the response variable that is explained by the explanatory variable. The square root of R^2 is an estimator for the correlation between the two variables. So, an R^2 of 49%, which means that about half of the variation on one variable is explained by the other, corresponds to an estimated correlation of 0.70. Higher R^2 values correspond to even larger correlations.

Correlations can be estimated from historic data, if available, or they can be selected judgmentally. Some caution should be exercised when using historic data, as common estimation methods can severely underestimate correlations, especially when the correlation is large. The choice of a value for the correlation coefficient can be influenced by how the result is to be used. If the goal is to obtain a central estimate, correlations based on historic levels may be adequate. On the other hand, if the goal is to obtain estimates in the tails, higher correlation selections may be justified, because correlations tend to be higher in the tails--when it rains, it pours.

4. An Example

The accompanying Excel spreadsheet illustrates how this technique can be used. The spreadsheet accepts up to eight lines-of-business and combines them into a single total. The user

[9] A family of distributions is said to be "closed under addition" if, whenever two members of the family are added together, the resulting sum is in the same family, but possibly with different parameters.

gives two points on a normal curve for each line-of-business. The user also specifies a correlation matrix for the various lines-of-business.

The reserve being estimated is outstanding losses. This purely hypothetical example uses five lines. Lines A, B, and C are taken from the above example[10]. Line D is a large loss, which the claim handler estimates will settle for $250 million, although there is a chance that it could realistically be as little as $150 million or as high as $400 million. Selecting the 20th-percentile to be $150 million and the 90th-percentile to be $400 million produces an expected loss close to $250 million; this is shown on the Output Page. Tweaking the 90th-percentile to be the 89.67th-percentile trues up the expected loss to be $250 million. Line E represents the unpaid losses from a recent catastrophe. They are currently estimated to be $450 million, but are expected to grow to $500 million. There is a small chance that the ultimate losses will turn out to be much worse--we selected a 95th-percentile loss of $1,155 million. We have chosen to enter $500 million as the 50th-percentile (which is the mean for symmetric distributions, such as the normal distribution.) The spreadsheet uses the method of percentile matching (see, for example, [KPW]) and a univariate normal is determined by two parameters (say, the mean and the standard deviation), so we can specify two percentiles for each line.

For the correlations among Lines A, B, and C, we will use the correlations we used above. We believe that Line D will act less like Line A or B than Line C, so we select 25% for the first two and 50% for the third correlation. Finally, we feel that an increase in Line E would come from a general adverse change in insurance loss reserves in general (less friendly courts, unexpected inflation, etc.), so we select 50% for the correlation with each of Lines A, B, C, and D.

The resulting correlation matrix is positive definite, so there is a multi-normal distribution that has our selected correlation matrix and that has our modelled losses for our lines of business as its marginals. The sum of our five lines of business has a normal distribution with mean and standard deviation computed by the spreadsheet. It is now an easy matter to select the mean and the two specified percentiles. These are shown on the Output Page along with the ranges corresponding to no correlation and comonotonicity ("100% correlation").

The spreadsheet will calculate a value for any given percentile; however, it is designed for estimates somewhat close to the center of the distribution. Solvency II calls for calculations at specified percentiles such as the 99.5th-percentile. Caution should be exercised in estimating such high percentiles using these methods.

[10] To reproduce the earlier example, simply set the Line D and Line E losses at the 25th- and 75th-percentiles to zero.

5. Concluding Remarks

Estimates for a given quantity often are obtained by combining other estimates. In the case of multiple estimates for the given quantity, we can combine them and obtain an estimate with more precision than any of the individual estimates. On the other hand, if our estimate is for a sum and we have estimates for the summands, we cannot hope to obtain (nominal) precision better than the worst of our summands, and in fact we cannot even do that well. Often, however, it is possible to improve the precision in a relative sense, using a measure such as the coefficient of variation.

Patel and Raws' simulation work showed that the choice of distribution did not matter very much, but they had to assume independence. We give up the choice of distribution, always selecting a normal distribution, but in exchange we recapture the ability to select the correlation structure. Since, in many cases, independence cannot reasonably be assumed, this is a real advantage.

Acknowledgement

The author would like to thank the reviewers whose individual and collective comments greatly improved this paper.

Supplementary Material

There is an accompanying Excel spreadsheet which contains the examples from this paper. It can be found on the CAS website. The URL is: http://www.casact.org/pubs/forum/14fforum/.

References

[KPW] Klugman, Stuart A; Panjer, Harry H.; and Willmot, Gordon E.; *Loss Models: From Data to Decisions*. New York: Wiley, 1998. Print.

[MS] Mango, Donald F. and Sandor, James C.; "Dependence Models and the Portfolio Effect", *CAS Forum*, Winter 2002, p. 57-72

[PR] Patel, Chandrakant C. and Raws, Alfred; "Statistical Modeling Techniques for Reserve Ranges: A Simulation Approach", *CAS Forum*, Fall 1998, p. 229-255

[S] Struppeck, Thomas, "Correlation", *CAS Forum,* Spring 2003, p. 153-176

[V] Venter, Gary G., "Tails of Copulas", *PCAS* **2002**, Vol. LXXXIX, p. 68-113

Biography of the Author

Tom is a lecturer at the University of Texas at Austin, where he teaches statistics. He is a Fellow of the CAS, an Associate of the SOA, and a CERA. Currently he serves on the examination committee where he advises on statistics.