In practice, it was not quite this simple, in that there was another set of $A$, $B$, and $C$ values estimated for excess losses simultaneously, and both sets of parameters were indexed for inflation and for state differences in loss severity. This can be accomplished by indexing $B$ and $C$ only, in that then the credibilities will not change if $P$ increases according to the index.

The (extended) Bailey-Simon method thus obviates the need for estimating variances, and it also gives the credibility formula that would have worked best in the past. To the extent that subtle violations of the model assumptions are operating, the resulting formula could work better even than one with the actual variances, if they were known.

## How Good Is Least Squares Credibility?

### *Linearization Error

In Example 6.3, the percentage change in territory pure premium was estimated using least squares credibility. It is often felt appropriate to use this credibility for percentage changes in pure premium, and for loss ratios, but not for pure premiums themselves by territory or by class. One reason for this is that pure premium differences among territories and among classes may be greater, especially in the extremes, than would be accounted for in this theory. Since least squares credibility relies on the first two moments, it may not be well adapted for use with highly skewed distributions, for example. In this section it is shown that least squares credibility can perform quite poorly with highly skewed distributions. Taking the logs of the data before applying credibility is explored as an alternative.

As discussed earlier, the function of the $X_{iu}$'s that optimizes the expected squared error in $X_{g0}$ is the predictive expected value $E(X_{g0}|$ the $X_{iu}$'s). The best *linear* function in this sense is the least squares credibility estimate.

Whenever the predictive expectation is a linear function of the observations, it is thus the same as the credibility estimator, and several examples of this were seen above. Jewell (1973) has shown that this occurs when the conditional density of $X$ given the unknown parameter $R$ is a member of the linear exponential family, i.e., when the density can be written in the form $f(x|R) = exp[xp(R) + b(-x) + q(R)]$ for functions $b$, $p$, and $q$, as long as $R$ is distributed as the natural conjugate prior of $f$. The gamma-Poisson, inverse gamma-gamma, and normal-normal conjugate pairs are useful examples.

It is widely speculated that probability densities of the above form are the only cases where the credibility estimate is the same as the predictive expectation (e.g., see Goel (1982)). In all other cases, the difference between these two estimates can be described as linearization error, since it is the minimal additional error that can arise from using a linear estimator instead of the predictive expectation.

Many of the distributions that arise in casualty insurance practice, such as the lognormal, Weibull, Pareto, and inverse gamma are not in the linear exponential family, and in fact quite a bit of accuracy may be sacrificed by the use of linear estimators for variables having these distributions, as the following examples show. Taking logs before using credibility and then exponentiating the answer, as an alternative to standard credibility practice when estimating such variables, will be illustrated as well.

*Example 7.1*

The first example will consider the lognormal distribution with a lognormal prior in the constant exposure model. This might be a reasonable model for claim severity, for example. The predictive mean and credibility estimate will both be calculated and compared to each other to determine the extent of linearization error.

The lognormal distribution can be parameterized to have two parameters, $b$ and $c^2$, such that $1/f(x) = cx\sqrt{\pi 2}exp([ln(x/b)]^2/2c^2)$ and $E(X^j) = b^j exp(j^2c^2/2)$. In this parameterization, $b$ is the exponentiation of the usual parameter $\mu$. In particular, $EX = bexp(c^2/2)$ and $VarX = b^2[exp(2c^2) - exp(c^2)]$.

The following properties of the lognormal can be derived by the methods used in Example 2.4. If $B$ is lognormal in $(v, q^2)$ (the prior distribution), and $X|B$ is lognormal in $(B, c^2)$ (the conditional distribution), then $X$ is unconditionally lognormal in $(v, c^2 + q^2)$ (the mixed distribution). If $n$ observations are made, denoted as $\{X_i\} = X_1, \ldots, X_n$, then the posterior distribution for $B|\{X_i\}$ is the lognormal in $v'$ and $c^2q^2/(c^2 + nq^2)$, where $ln(v') = [c^2 ln(v) + q^2 \Sigma_i lnX_i]/[c^2 + nq^2]$. The predictive distribution for $X|\{X_i\}$, being the mixture of the conditional by the posterior, is then lognormal in $v'$ and $c^2 + c^2q^2/(c^2 + nq^2)$. The predictive mean is then $E(X|\{X_i\}) = v'exp([c^2/2 + c^2q^2/2(c^2 + nq^2)])$. Letting $z = nq^2/(c^2 + nq^2)$:

$$E(X|\{X_i\}) = exp(zlnX_. + (1 - z)ln(v))exp(.5[c^2 + (1 - z)q^2]) \qquad (7.1)$$

where $lnX_.$ denotes the average value of the log of the observations. Since $z = n/(n + k)$, where $k = c^2/q^2$, this predictive mean can be seen to be a constant factor $M$ times the exponentiation of the credibility estimate of $lnX$, where the constant factor is $M = exp(.5[c^2 + (1 - z)q^2])$.

To find the credibility estimate of $X$ given the $X_i$'s, the lognormal moment formulas are applied to the conditional and prior distributions. The credibility is $Z = n/(n + K)$, where

$$\begin{aligned} K &= EVar(X|B)/VarE(X|B) \\ &= E(B^2[exp(2c^2) - exp(c^2)])/Var(Bexp(c^2/2)) \\ &= v^2exp(2q^2)[exp(2c^2) - exp(c^2)]/exp(c^2)v^2[exp(2q^2) - exp(q^2)]. \end{aligned}$$

This simplifies to:

$$K = [exp(q^2)][exp(c^2) - 1]/[exp(q^2) - 1] \qquad (7.2)$$

Table 7.1 illustrates a simulation test of the resulting credibility estimate vs. the predictive mean, in order to illustrate the potential linearization error for the lognormal. Parameter values of $c^2 = 4$, $v = 1$, and $q^2 = 2$ were used in this example. The $c^2$ value might correspond to the loss size distribution in a heavy tailed line. The $q^2$ value provides quite a large dispersion among classes or risks, perhaps more than would ordinarily be expected.

For the test, 10,000 risks were placed at each of the following percentiles of the prior distribution: 1, 10, 50, 75, 90, 99. Thus in this test, the $b$ parameter value is known for each risk, and so the different ways of estimating it can be tested. For each risk, 50 claims were simulated from the lognormal severity arising from the $B$ value at that percentile and the selected value $c^2 = 4$. Then the sample mean, the credibility estimator, and the predictive mean were computed from the 50 claims. Each of these was considered to be an estimator of the known conditional mean $E(X|B)$, and the absolute value and square of the estimation errors were recorded for each risk. The table gives the averages of these estimates and errors for the 10,000 risks at each percentile.

*Table 7.1*

**Lognormal Linearization Error**

| Percentile of B: | 1 | 10 | 50 | 75 | 90 | 99 |
|---|---|---|---|---|---|---|
| B | .037 | .163 | 1.00 | 2.60 | 6.13 | 26.8 |
| E(X\|B) | .275 | 1.21 | 7.39 | 19.2 | 45.3 | 198 |
| | | | Average Estimate | | | |
| Sample | .276 | 1.21 | 7.41 | 19.0 | 45.3 | 200 |
| Predictive Mean | .382 | 1.49 | 7.90 | 19.2 | 42.3 | 165 |
| Credibility | 11.2 | 11.7 | 14.4 | 19.6 | 31.3 | 100 |
| | | Average of Individual Risk Absolute Errors | | | | |
| Sample | .131 | .592 | 3.56 | 9.03 | 21.9 | 97.1 |
| Predictive Mean | .114 | .369 | 1.66 | 4.02 | 9.45 | 46.6 |
| Credibility | 11.0 | 10.5 | 7.04 | 3.88 | 18.6 | 113 |
| | | Average of Individual Risk Squared Errors | | | | |
| Sample | .058 | 1.35 | 41.5 | 319 | 1,760 | 42,000 |
| Predictive Mean | .021 | .241 | 4.81 | 26.0 | 136 | 3,020 |
| Credibility | 120 | 110 | 57.8 | 63.8 | 544 | 18,000 |

To compute these estimates, $K = 62$ can be found using (7.2) with $c^2 = 4$, $q^2 = 2$, so $Z = 50/112 = .45$. Similarly, $z = .96$ resulted for the credibility estimate of $lnX$, with a $K$ of $2 = c^2/q^2$. The constant $M = exp(.5[c^2 + (1 - z)q^2])$ from (7.1) is equal to 7.68; also, $EX = 20.1 = v exp([c^2 + q^2]/2)$, and $E(lnX) = ln(v) = 0$. With these values, $(1 - Z)EX = 11.1$, so this is a lower bound for the credibility estimate $(1 - Z)EX + Z\bar{X}_{..}$.

The table shows that the average sample estimate is generally quite good, as the sample estimate is unbiased. The error of the sample mean can be large for any particular risk, however. The predictive mean shows up to be by far the best estimator in terms of either absolute or squared errors. Thus the linearization error is quite substantial in this case.

The sample mean is also linear in the observations, and it may appear more favorable than the credibility estimate overall in this example. By the expected squared error criterion, the credibility estimator is better: the sample squared errors at the upper percentiles are much larger, so the average squared error over the entire distribution is higher than that of the credibility estimator, even though the credibility errors are higher at the lower percentiles. The better performance of the credibility estimator by this criterion does not appear to carry over to absolute errors, or to percentage errors, however. This raises doubt about whether the overall expected squared error is really an appropriate criterion when heavy tailed distributions are involved.

A possible alternative is to use least squares credibility to estimate the logs of the observations. This will minimize percentage errors when converted back to full values, and thus may be more appropriate, even in cases when, unlike the lognormal, it does not give the predictive mean. The next example looks at such a case.

With the distributions in the above example, an ad hoc method, such as using limited fluctuation credibility with $n_F = 50$, would probably appear better than the least squares credibility estimate for most risks, even though such a low value for full credibility would appear strange by usual limited fluctuation standards. In some aspects of workers' compensation ratemaking, e.g., serious pure premiums, as a matter of fact, such low full credibility values have been selected over the years, based on performance of the ratemaking method. While this has given rise to actuarial suspicion, the above gives a theoretical context for its applicability.

*Example 7.2*

In this example, the conditional and prior from Example 2.4 are reversed. Now X is inverse gamma in (c, Y), and Y *is gamma in* (r, b). The unconditional or mixed distribution for X is the Beta2 or generalized Pareto, in (r, c, b).

With $n$ observations $X_1, \ldots, X_n$, the posterior distribution is then also gamma, but now with parameters $(r+nc,\ a)$, where $a = [b^{-1} + \Sigma X_i^{-1}]^{-1}$. Thus the predictive distribution is Beta2 in $(r+nc,\ c,\ a)$, and the predictive mean is $a(r+nc)/(c-1)$.

This is to be compared to the usual credibility estimate and credibility based on the log transform. The $K$ value can be found to be $(r+1)/(c-2)$ by the usual approach. To do credibility in logs, $EVar(W|Y)$ and $VarE(W|Y)$ must be found, where $W = lnX$. These will rely on some lesser known facts about the gamma and inverse gamma distributions:

1. $E(W|y) = ln(y) - \Psi(c)$, where $\Psi(c)$ is the digamma function, the derivative of the log of the gamma function. This is a well tabulated function, e.g., see Abramowitz and Stegun (1965).

2. $Var(W|y) = \Psi'(c)$, the derivative of the digamma, which is called the trigamma function and is again well tabulated. Note that this does not depend on y.

3. $Eln(Y) = ln(b) + \Psi(r)$

4. $Var(lnY) = \Psi'(r)$

From these it follows that $EVar(W|Y) = \Psi'(c)$ and $VarE(W|Y) = \Psi'(r)$, and so for log credibility, $k = \Psi'(c)/\Psi'(r)$. Also, $ElnX = ln(b) + \Psi(r) - \Psi(c)$. Taking $z = n/(n+k)$, the log credibility estimate can be written as $Mexp[(1-z)ElnX + (z/n)\Sigma lnX_i]$, where $M$ is a constant needed to make the estimate unbiased. $M$ can be determined by computing the expected value of this estimate without the $M$ and then setting the estimate equal to $EX$. It can then be found that the log credibility estimator is:

$$nb^{1-z}(r-1)!(c-1)!^n exp[(z/n)\Sigma\ lnX_i]/$$
$$(c-1)(r+z-1)!(c-1-z/n)!^n.$$

For an example, 10,000 trials for each percentile shown were taken for a sample of $n = 25$, with parameters $b = 100$, $r = 0.5$, and $c = 4.0$. The results are given in Table 7.2.

*Table 7.2*

**Inverse Gamma Linearization Error**

| Percentile of Y: | 1 | 10 | 50 | 75 | 90 | 99 |
|---|---|---|---|---|---|---|
| Y | .008 | .789 | 22.7 | 66.2 | 135 | 332 |
| E(X\|Y) | .003 | .263 | 7.58 | 22.1 | 45.1 | 111 |
| **Average Estimate** | | | | | | |
| Sample | .003 | .263 | 7.58 | 22.0 | 45.1 | 110 |
| Predictive Mean | .003 | .267 | 7.67 | 22.2 | 45.1 | 108 |
| Credibility | .488 | .741 | 7.84 | 21.9 | 44.2 | 108 |
| Log Credibility | .003 | .266 | 7.60 | 22.0 | 45.0 | 110 |
| **Average of Individual Risk Absolute Errors** | | | | | | |
| Sample | .000 | .028 | .817 | 2.38 | 4.86 | 11.9 |
| Predictive Mean | .000 | .021 | .608 | 1.75 | 3.54 | 8.63 |
| Credibility | .485 | .478 | .797 | 2.33 | 4.85 | 12.0 |
| Log Credibility | .000 | .022 | .639 | 1.86 | 3.79 | 9.30 |
| **Average of Individual Risk Squared Errors** | | | | | | |
| Sample | .000 | .001 | 1.13 | 9.59 | 40.1 | 241 |
| Predictive Mean | .000 | .000 | .594 | 4.90 | 19.8 | 115 |
| Credibility | .235 | .230 | 1.13 | 9.07 | 38.5 | 235 |
| Log Credibility | .000 | .000 | .652 | 5.49 | 22.9 | 137 |

In this example, $\Psi(4) = 1.25612$, $\Psi(.5) = -1.96351$, $ln100 = 4.60517$, and so $ElnX = 1.38554$. Also $\Psi'(4) = .283823$, $\Psi'(.5) = 4.93480$ so $k = .0575146$, and $z = 25/(25+k) = .9977$. Since $K = .75$, $Z = .9709$, and $EX = br/(c-1) = 50/3$ can be computed.

For this sample, the log credibility performed almost as well as the predictive mean, and considerably better than linear credibility. In fact, due to lower absolute errors and percentage errors, the sample mean may be felt to be better than the linear credibility estimator, even though it can be shown to have a higher expected squared error overall. This again raises doubts about how appropriate the overall expected least squares criterion is for heavy tailed distributions.

To apply log credibility in practice, it is not necessary to use digamma and trigamma functions. At least in the constant exposure case, the empirical development of $K$ can be applied to the logs of the observations, and the resulting credibility estimators exponentiated. This will not be in general unbiased, but a factor $M$ can be developed so that the credibility estimators over all risks balance to the overall mean, simply by eliminating the off balance that would arise without this factor.

This procedure is applicable, for example, to loss severity, or to other instances of the equal exposure case. Before it could be applied in the more general situation, the relationship of $Var(lnX_{iu})$ to $P_{iu}$ would have to be determined, and resulting credibility formulas worked out. For this it may be reasonable to use the simplest model $Var(lnX_{iu}|R_i) = s(R_i)^2/P_{iu}$, in fact. For a gamma distributed variable in $(c,b)$, the sum of n independent observations is gamma in $(nc, b)$, and the variance of the log of this sum is thus $\Psi'(nc)$. This is reasonably well approximated by $1/nc$, at least when nc is not too small. (Exercise. Prove this by twice differentiating Sterling's approximation for the log of the gamma function.) Thus assuming that the conditional variance of $lnX$ is approximately inversely proportional to the exposure $n$ makes sense at least in this case.

The principle of working with a transformed loss value, having a more manageable distribution, has been applied in casualty actuarial practice. An example, mentioned in the opening section to this chapter, is the multi-split experience rating plan, first introduced in 1940 in New York. For small and medium sized risks, credibility was applied only to losses transformed by a "primary value" function $p(X)$. An example in recent use is:

$$p(X) = X \qquad \text{if } X \leq \$2000$$

$$p(x) = 10{,}000X/(X + 8000) \text{ if } X > \$2000$$

The primary value is always less than \$10,000, which effectively limits the tail of the distribution.

## Expected Squared Error of Credibility Estimate

Another approach to the question, "How good is least squares credibility?" is to compute the expected squared error of the estimate, either conditional on the parameters $R_i$ or overall. First the conditional expected squared error is calculated. Through straightforward algebraic manipulation of probabilities, in the equal exposure case it is possible to show that, for any weight Z:

$$E\{[ZX_{i.} + (1-Z)m - E(X_{i0}|R_i)]^2|R_i\} =$$
$$Z^2Var(X_{i0}|R_i)/n + (1-Z)^2[m - E(X_{i0}|R_i)]^2 \qquad (7.3a)$$

and thus, taking the expected value of this, the overall expected squared error is:

$$E[ZX_{i.} + (1-Z)m - E(X_{i0}|R_i)]^2 =$$
$$Z^2EVar(X_{i0}|R_i)/n + (1-Z)^2VarE(X_{i0}|R_i) =$$
$$Z^2s^2/n + (1-Z)^2t^2 \qquad (7.3b)$$

(Minimizing 7.3b with respect to Z gives $Z = n/(n+K)$, as it should.)

Taking Z = 1 gives the squared error for the sample mean. For any particular class, this may be greater or smaller than the credibility squared error, depending on how close that class really is to the overall mean, as (7.3a) shows. For Example 7.1, the conditional expected squared errors are given below for the credibility value Z = .45, the sample estimate Z = 1, and an arbitrary choice of Z = .90.

*Conditional Expected Squared Errors*

| **Percentile of B:** | 1 | 10 | 25 | 50 | 75 | 90 | 99 |
|---|---|---|---|---|---|---|---|
| B | .037 | .163 | .385 | 1.00 | 2.60 | 6.13 | 26.8 |
| E(X\|B) | .275 | 1.21 | 2.85 | 7.39 | 19.2 | 45.3 | 198 |
| Z = 1 | .08 | 1.6 | 8.7 | 59 | 395 | 2,198 | 42,000 |
| Z = .9 | 4 | 4.8 | 10 | 49 | 321 | 1,789 | 34,400 |
| Z = .45 | 120 | 110 | 91 | 60 | 79 | 942 | 17,800 |

The overall expected squared errors from (7.3b) for these Z's are 3200, 2600, and 1400, respectively. This again illustrates that the expected squared error can be dominated by the large values of the parameter, which is not necessarily the way an estimation process should be evaluated. The last row can be compared to the simulated values from Table 7.1.

Even though the overall average squared error is lower for the credibility estimator, the individual (conditional) expected squared error can be much lower for the sample mean at some percentiles. This occurs even in the case of normal distributions, as Efron and Morris (1972) point out. An adaptation they discuss is placing a selected limit on how much the credibility estimator can differ from the sample value. This "limited translation estimator" effectively gives additional credibility to the observations furthest from the grand mean.

Since non-extreme risks can sometimes produce extreme observations, this procedure increases the overall expected squared error, but reduces the individual squared errors at the extremes, where they are often largest. Efron and Morris show how to compute the increase in overall expected squared error and the decrease in individual squared error when the limit is a constant difference from the sample value. This makes a lot of sense in their model, as all risks have the same conditional variance. For heavy tailed distributions, a percentage limit may be worthy of consideration, since the coefficient of variation is more likely to be constant across risks, as in the above examples.

For example, in a classification ratemaking context, workers' compensation class rates can range from 10 cents to $100 per $100 of payroll. Allowing a 10% error would seem more equitable than allowing a fixed $5 error in such a case.

## Further Topics

A number of different extensions of the above methods have been worked out by various authors. A very limited selection of those with applications in casualty insurance is presented in this section, in the form of an annotated bibliography. In general, references noted in earlier sections are not included here.

Mahler, Howard C. 1986, An actuarial note on credibility parameters *PCAS* 73.

The least squares and limited fluctuation formulas for Z can be matched fairly closely with the right choice of parameters. Mahler shows that taking $n_F = 8K$ makes the credibilities close and the credibility estimators even closer.

Klugman, Stuart. 1987 Credibility for classification ratemaking via the hierarchical normal linear model, *PCAS* 74.

A Bayesian approach to estimating $s^2$ and $t^2$ in the normal-normal model is presented, using both diffuse and proper priors. The choice of prior turns out to have little effect on the outcome. In a test using live insurance data, improved results are obtained. Although numerical integration is needed, error estimates are provided which incorporate the uncertainty about $s^2$ and $t^2$.

Venter, Gary G. 1986, Classical partial credibility with application to trend *PCAS* 73.

The limited fluctuation paradigm is extended to trend estimates. A trend projection is given full credibility if the p-confidence interval around it has radius no greater than 100k% of the projected value. This radius is a function of the goodness of fit of the trend line.

Hachemeister, Charles A. 1975 Credibility for regression models with application to trend in *Credibility—Theory and Applications*, Kahn, ed. New York: Academic Press.

Least squares credibility is applied to regression models in a very general setting. In a particular application, trend lines for several states are credibility weighted against each other, with number of claims used as the basis of credibility.

Miller, Robert B., and Hickman, James C. 1975 Insurance credibility theory and Bayesian estimation in *Credibility—Theory and Applications*, Kahn, ed. New York: Academic Press.

Least squares credibility and Bayesian analysis are applied to the collective risk model for frequency, severity, and total claims.

Heckman, Philip. 1980 Credibility and solvency in *Pricing property and casualty insurance products*, New York: Casualty Actuarial Society.

A hierarchical least squares credibility scheme is presented in which risk means are weighted against class means, which in turn are weighted against the overall mean. The two levels of credibility are computed in an integrated manner, but these turn out to be separable steps.

DuMouchel, William H. 1983 The Massachusetts automobile insurance classification scheme *The Statistician*: 32.

Credibility is applied in a two dimensional setting. Class rate relativities by territory are initially modeled as a statewide class relativity plus the product of a territory relativity with a second statewide class relativity. The deviations of the individual cell relativities from this model are assumed normal, which leads to a credibility weighting between the sample cell relativity and the model expectation.

Venter, Gary G. 1985 Structured credibility in applications—hierarchical, multi-dimensional and multivariate models *ARCH*:2.

The hierarchical model described by Heckman is applied to calculating workers' compensation loss severity by class within hazard group, and two other models are described. A multivariate model is used to estimate several correlated quantities by class, e.g., frequency for different injury types. A two-dimensional model is developed to estimate class by state relativities by weighting the individual class-state cell simultaneously against the other classes and other states, without assuming any particular underlying additive or multiplicative structure between state and class.

de Jong, Piet and Zehnwirth, Ben 1983a Credibility and the Kalman filter," *Insurance Mathematics and Economics* 2.

The Kalman filter is a generalized model which is shown to include most of the known credibility models as special cases, while also allowing for changes in structural parameters over time.

de Jong, Piet and Zehnwirth, Ben 1983b Claims reserving, state-space models and the Kalman filter *Journal of the Institute of Actuaries*: 110.

Accident year loss payout timing can be modeled by various curves, such as exponential decay, generalizations of the inverse power curve, etc. Curve parameters are estimated based on payment triangles. Parameters are allowed to change from one accident year to another, but they are "credibility weighted" against the parameters for other years using the Kalman algorithm. Practical methods for how to use this algorithm are discussed. Standard errors for the resulting loss reserve estimates are provided by this method.

Robbin, Ira. 1986 A Bayesian credibility formula for IBNR counts *PCAS* 73.

For a given set of loss exposures, the claims emerging in each reporting period are postulated to arise from a process (Poisson for example) with a vector of parameters u (probably not precisely known). After some of these reporting periods, M claims have been observed and the number of claims yet to emerge is to be estimated. Credibility weights are specified for three estimators of this IBNR:

(a) the observed claims to date times a development factor

(b) the ultimate claims expected originally (prior to claims emergence), less the observed claims to date.

(c) the number of claims originally expected to emerge after this date.

To derive the credibility weights. the variance of $M$ can be broken down into $EVar(M|u)$ and $VarE(M|u)$. The second component is further split out by defining $n$ and $q$ (independent functions of the parameters $u$) so that $E(M|u) = n(1 - q)$. The interpretation is that $n$ *is the conditional expected ultimate number of claims, and* $q$ is the expected proportion of IBNR claims at this date, given $u$.

Since for independent $X$ and $Y$, $Var(XY) = E(X^2)VarY + (EY)^2 VarX$, it follows that $VarE(M|u) = E(n^2)Var(1 - q) + [E(1 - q)]^2 Var(n)$.

Thus three components of $VarM$ are identified, and the three estimators above end up being weighted in proportion to these three components.

Estimator (a) has weight proportional to $Var(n)$. If the ultimate number of claims is not well known, $Var(n)$ is high and a large weight is given to the developed observed claims. The weight on (b) is proportional to $Var(1 - q)$. If the development pattern is poorly known, (a) and (c) are less reliable. Finally, if $M$ itself is highly random, little weight can be given to it. Thus if $EVar(M|u)$ is higher, the weight on (c) increases.

## Conclusion

Credibility as a topic has been with the CAS since the first volume of the *Proceedings*, and it is by no means a finished one, as witnessed by the flurry of papers recently. The best form for credibility estimators has yet to be determined for many applications. It is not unreasonable to expect that the final volume of the *PCAS*, whenever it is published, will also contain a paper addressing credibility theory.

## Postlogue

by Charles C. Hewitt, Jr.

### Number of Observations

*Sometime in the 1920s, Sinclair Lewis, the author, was to receive an award as a distinguished alumnus of Yale University. In accepting the award, Lewis took the occasion to assert his known atheism.*

> *"I do not believe there is a God," said Lewis (in substance). "If, in fact, there is one, let him strike me down here and now!" And, of course, nothing happened. However, several days later the noted newspaper columnist, Arthur Brisbane, took Lewis to task.*
>
> *"Lewis, you poor misguided fool," wrote Brisbane (in substance). "You remind me of the ants who lived along the right-of-way of the Atchison, Topeka, and Santa Fe Railroad. This colony of ants depended for its existence upon the crumbs thrown from the dining cars of the railroad trains as they passed by.*

*"It came to pass that the ant colony fell upon hard times because, through chance, no crumbs were thrown out near its particular place along the right-of-way. The situation became desperate and the colony decided to hold a meeting. It was suggested that they all pray to the president of the Atchison, Topeka, and Santa Fe Railroad to send more dining cars so that crumbs would be thrown off in their area.*

*"So they did pray and the following day they waited, but no crumbs were thrown off where they lived. So the ants concluded that there was no such person as the president of the Atchison, Topeka, and Santa Fe Railroad."*

**Variance of the Processes**

*Four boys decided to play "hooky" from school, because they knew there was to be a test that morning. About 11 A.M. their consciences got the better of them and they decided to show up at school after all. Upon reaching their classroom, they explained to the teacher that they had been on their way to school in a car, but the car had a flat tire. This made them late because they had to have the flat tire fixed.*

*"No problem!", said the teacher. "Just come back here during the lunch hour and I'll give you a make-up test." At lunch time, when they reported back to the classroom, the teacher instructed the four boys to take seats in the four corners of the room.*

*"Now," said the teacher, "there's only one question on this make-up test. Which tire was flat?"*

**Variance of the Hypotheses**

**Television interviewer:** *"Do you believe in miracles?"*

Guest: *"Of course!"*

**Television interviewer:** *"Have you ever seen a miracle?"*

Guest: "No."

**Television interviewer:** *"Do you know anyone who has actually seen a miracle?"*

Guest: *"No, but that doesn't prove anything!"*

# References

Abromowitz, M. and Stegun, I. 1965. *Handbook of mathematical functions*. New York: Dover.

Bailey, A. 1945. A generalized theory of credibility, *PCAS* 32.

Bailey, A. 1950. Credibility procedures, *PCAS* 37.

Bailey, R. and Simon L. 1959. An actuarial note on the credibility of experience of a single private passenger car, *PCAS* 46.

Bailey, Robert. 1961. Experience Rating Reassessed. *PCAS* 48.

Berger, J. 1985. *Statistical decision theory and Bayesian analysis*, 2nd ed. New York: Springer Verlag.

Box, G. and Tiao, G. 1973. *Bayesian inference in statistical analysis*, Reading, MA: Addison-Wesley.

Bühlmann, Hans. 1967. Experience Rating and Credibility. 4 *ASTIN Bulletin*.

Bühlmann, H. and Straub E. 1970. Glaubwürdigeit für schadensätze (Credibility for loss ratios), English translation, *ARCH* 1972(2).

DePril, N. 1986. Improved recursions for some compound Poisson distributions, *Insurance Mathematics and Economics* 5.2.

DeVylder, F. 1981. Practical credibility theory with emphasis on optimal parameter estimation, *ASTIN Bulletin* 12.2.

Dropkin, L. 1959. Some considerations on automobile rating systems utilizing individual driving records, *PCAS* 46.
1960. Automobile Merit Rating and Inverse Probabilities. *PCAS* 47.

Efron, B. and Morris, C. 1972. Limiting the risk of Bayes and empirical Bayes estimators. Part II: The empirical Bayes case, *JASA* 67 (March).

—-. 1975. Data analysis using Stein's estimator and its generalizations, JASA 70 (June).

—-. 1977. Stein's paradox in statistics, *Scientific American* (May).

Ericson, W. 1970. On the posterior mean and variance of a population mean, *JASA* 65.

Gerber, H. and Jones, D. 1975. Credibility formulas of the updating type. In *Credibility: Theory and Applications,* New York: Academic Press.

Goel, P. 1982. On implications of credible means being exact Bayesian, *Scandanavian Actuarial Journal.*

Heckman, P. and Meyers G. 1983. The calculation of aggregate loss distributions from claim severity and claim count distributions, *PCAS* 70.

Hewitt, C. 1967. Loss ratio distributions a model, *PCAS* 54.

Hogg, R. and Klugman, S. 1984. *Loss distributions,* New York: Wiley.

Insurance Services Office. 1980. *Report of the credibility subcommittee: development and testing of empirical Bayes credibility procedures for classification ratemaking,* New York: ISO.

Jewell, W. 1973. Credible means are exact Bayesian for simple exponential families, *ORC*: 73-21, Operations Research Center, University of California, Berkeley, CA. (Also in *ASTIN Bulletin* 8:1, September 1974).

Keffer, R. 1929. An experience rating formula, *Transactions of the Society of Actuaries* 30.1.

Klugman, S. 1985. Distributional aspects and evaluation of some variance estimators in credibility models, *ARCH* 1985.1.

Klugman, S. 1987. Credibility for classification ratemaking via the hierarchical normal linear model, *PCAS* 74.

Lundberg, Ove. 1940. On Random Processes and Their Application to Sickness and Accident Statistics. Uppsala (Sweden): Harald Cramer Jubilee Volume.

Mahler, H. 1987. Review of "An analysis of experience rating", *PCAS* 74.

McDonald, James B. (1984). "Some Generalized Functions for the Size Distribution of Income," *Econometrica,* Vol. 52, pp. 647-663.

Mayerson, Allen L. 1964. A Bayesian View of Credibility. *PCAS* 51.

Meyers, G. 1985. An analysis of experience rating, *PCAS* 72.

Meyers, G. and Schenker N. 1983. Parameter uncertainty in the collective risk model, *PCAS* 70.

Mong, S. 1980. Estimating aggregate loss probability and increased limit factors in *Pricing Property and Casualty Insurance Products*, New York: Casualty Actuarial Society.

Morris, C. 1983. Parametric empirical Bayes inference: Theory and applications, *JASA* 78.

Mowbray, A. 1914. How extensive a payroll exposure is necessary to give a dependable pure premium? *PCAS* 1.

Nelson, D. 1969. Review of "The credibility of the pure premium", *PCAS* 56.

Panjer, H. 1981. Recursive evaluation of a family of compound distributions, *ASTIN Bulletin* 12:1.

Pentikäinen, T. 1977. On the approximation of the total amount of claims, *ASTIN Bulletin* 9.3.

Perryman, F. 1932. Some notes on credibility, *PCAS* 19.

Raiffa, Howard and Schlaifer, Robert. 1961. Applied Statistical Decision Theory. Cambridge: Harvard University Press.

Savage, Leonard J. 1954. The Foundations of Statistics. New York: John Wiley & Sons.

Schlaifer, Robert. 1959. Probability and Statistics for Business Decisions. New York: McGraw-Hill.

Simon, L. 1969. Review of "The credibility of the pure premium", *PCAS* 56.
1962. An Introduction to the Negative Binominal Distribution and Its Applications. *PCAS* 49.

Seal, H. 1977. Approximations to risk theory's F(X,T) by means of the gamma distribution, *ASTIN Bulletin* 9.C.

Stellwagen, H. 1925. Automobile rate making, *PCAS* 11.

Venter, G. 1983a. Review of "The calculation of aggregate loss distributions from claim severity and claim count distributions", *PCAS* 70.

—-. 1983b. Transformed beta and gamma distributions and aggregate losses, *PCAS* 70.

Whitney, A. 1918. The theory of experience rating, *PCAS* 4.

# Table of Distributions

**Preliminaries**

Gamma function: $\Gamma(r) = \int_0^\infty y^{r-1}e^{-y}dy = (r-1)!$ The partial integral can be evaluated by a series:

Incomplete Gamma function: $\Gamma(r;x) = \int_0^x y^{r-1}e^{-y}dy \div \Gamma(r) = \frac{e^{-x}x^{r-1}}{(r-1)!}\sum_{i=0}^{\infty}\prod_{k=0}^{i}\frac{x}{r+k}$

Beta function: $\beta(r,s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} = \int_0^\infty \frac{t^{r-1}dt}{(1+t)^{r+s}} = \int_0^1 u^{r-1}(u-1)^{s-1}du$, where $u = \frac{t}{1+t}$

Incomplete Beta function: $\beta(r,s;x) = \int_0^x u^{r-1}(u-1)^{s-1}du \div \beta(r,s) = \int_0^{x/(1-x)} \frac{t^{r-1}dt}{(1+t)^{r+s}} \div \beta(r,s) =$

$= \frac{x^r}{\beta(r,s)}\left[\frac{1}{r} + \sum_{i=1}^{\infty}\frac{x^i}{r+i}\prod_{k=1}^{i}1-\frac{s}{k}\right] = F(\frac{s}{r}\frac{x}{1-x})$, where $F$ is the distribution function of the F distribution with $2r$ and $2s$ degrees of freedom. Thus either the series expansion or a package with the $F$-distribution can be used to compute the incomplete beta function. Note that the incomplete beta and gamma functions are both increasing functions, with range [0,1), and thus can be used as distributions.

When possible, the distributions below are parameterized with a multiplicative scale parameter $b$, so that $E(X^y) \propto b^y$. This is usually more convenient for applications, although the parameterizations given in Hogg and Klugman (1984) are easier to estimate by maximum likelihood. After estimation they can be readily translated to the forms below, however. In the following, $F$ is the distribution function, and f the density.

**1. Transformed Beta distribution ($r,s,a,b$)**

$F(x) = \beta(r,s; \frac{x^a}{x^a + b^a})$; $f(x) = \frac{(a/b)(x/b)^{ar-1}}{\beta(r,s)(1+(x/b)^a)^{r+s}}$,

$E(X^y) = \frac{b^y\beta(r+\frac{y}{a}, s-\frac{y}{a})}{\beta(r,s)}$, $-ar < y < as$.

**1a. Burr Distribution (s,a,b,) (r = 1)**

$F(x) = 1 - (1 + (x/b)^a)^{-s}$; $f(x) = \frac{s(a/b)(x/b)^{a-1}}{(1 + (x/b)^a)^{1+s}}$,

$$E(X^y) = \frac{b^y \frac{y}{a}!(s - 1 - \frac{y}{a})!}{(s - 1)!}, \; -a < y < as.$$

**1ai. Loglogistic Distribution (a,b) (r,s = 1)**

$F(x) = \frac{x^a}{x^a + b^a}$; $f(x) = \frac{(a/b)(x/b)^{a-1}}{(1 + (x/b)^a)^2}$,

$$E(X^y) = b^y \frac{y}{a}!(-\frac{y}{a})!, -a < y < a.$$

**1b. Generalized Pareto (r,s,b) (a = 1) (also called Generalized-F)**

$F(x) = \beta(r,s; \frac{x}{x+b})$; $f(x) = \frac{b^{-r}x^{r-1}}{\beta(r,s)(1 + (x/b))^{r+s}}$;

$$E(X^y) = \frac{b^y \beta(r + y, s - y)}{\beta(r,s)}, -r < y < s.$$

If $y$ is an integer $n$, $E(X^n) = b^n \prod_{i=1}^{n} \frac{r+i-1}{s-i}$.

The Pareto is given by $r = 1$, $F(x) = 1 - (1 + \frac{x}{b})^{-s}$.

**1c. Generalized Loglogistic (r,a,b) (s = 1) (also called inverse Burr)**

$F(x) = \left[\frac{x^a}{x^a + b^a}\right]^r$; $f(x) = \frac{r(a/b)(x/b)^{ar-1}}{(1 + (x/b)^a)^{r+1}}$,

$$E(X^y) = \frac{b^y(r - 1 + \frac{y}{a})!(-\frac{y}{a})!}{(r-1)!}, \; -ar < y < a.$$

The special case $a = 1$ is the inverse Pareto, which only has moments $-r < y < 1$, and thus no mean!

**2. Transformed Gamma (r,a,b)**

$F(x) = \Gamma(r;(x/b)^a)$; $f(x) = \frac{a(x/b)^{ar-1}e^{-(x/b)^a}}{b\Gamma(r)}$;

$$E(X^y) = \frac{b^y\Gamma(r+\frac{y}{a})}{\Gamma(r)}, y > -ar.$$

**2a. Weibull $(a,b)$ $(r=1)$**

$F(x) = 1 - e^{-(x/b)^a};$ $f(x) = (a/b)(x/b)^{a-1}e^{-(x/b)^a};$

$E(X^y) = b^y\frac{y}{a}!, y > -a.$

**2b. Gamma $(r,b)$ $(a=1)$**

$F(x) = \Gamma(r; x/b);$ $f(x) = \frac{(x/b)^{r-1}e^{-(x/b)}}{b\Gamma(r)},$

$E(X^y) = \frac{b^y\Gamma(r+y)}{\Gamma(r)}, y > -r.$ For an integer $n$, $E(X^n) = b^n \prod_{i=0}^{n-1} r+i.$

**3. Inverse Transformed Gamma $(s,a,b)$ $(1/X)$ Transformed Gamma)**

$F(x) = 1 - \Gamma(s;(x/b)^{-a});$ $f(x) = \frac{(a/b)e^{-(x/b)^{-a}}}{(x/b)^{as+1}\Gamma(s)};$

$E(X^y) = \frac{b^y\Gamma(s-\frac{y}{a})}{\Gamma(s)}, y < sa.$

**3a. Inverse Weibull (log-extreme-value) $(a,b)$ $(s=1)$**

$F(x) = e^{-(x/b)^{-a}};$ $f(x) = (a/b)(x/b)^{-a-1}e^{-(x/b)^{-a}};$

$E(X^y) = b^y(-\frac{y}{a})!, y < a.$

For $a = 1$, the inverse exponential has only moments $y < 1$.

**3b. Inverse Gamma $(s,b)$ $(a=1)$**

$F(x) = 1 - \Gamma(s; b/x);$ $f(x) = \frac{(1/b)e^{-b/x}}{(x;b)^{s+1}\Gamma(s)};$

$E(X^y) = \frac{b^y\Gamma(s-y)}{\Gamma(s)}, y < s.$ For an integer $n$, $E(X^n) = b^n + \prod_{i=1}^{n} s-i.$

**4. Lognormal $(b,c^2)$**

$F(x) = N\left[\frac{ln(x/b)}{c}\right];$ $f(x) = \frac{e^{-[ln(x/b)]^2/2c^2}}{cx\sqrt{2\pi}};$ $E(X^y) = b^y e^{y^2c^2/2}.$

Note that all moments exist. McDonald (1984) shows that the lognormal is a limiting case of the transformed beta, transformed gamma, and inverse transformed gamma. ($N$ is the standard normal distribution function.)

X, given Y, is distributed according to the conditional distribution. The distribution of Y is given by the prior distribution. The combination of these two gives the unconditional, or marginal, or mixed distribution of X. On taking $n$ observations from this distribution, $x_1, \cdots, x_n$, the revised opinion on Y is given by the posterior distribution, and the new combined distribution for X is the predictive distribution. The distributions are conjugate when the posterior is of the same form as the prior.

| Conditional | Prior | Mixed | Posterior | Predictive |
|---|---|---|---|---|
| TG $(r,a,y)$ | ITG $(s,a,b)$ | TB $(r,s,a,b)$ | ITG $(s+nr,\ a,\ [b^a + \sum_1^n x_i^a]^{\frac{1}{a}})$ | TB $(r, s+nr, a, [b^a + \sum_1^n x_i^a]^{\frac{1}{a}})$ |
| ITG $(s,a,y)$ | TG $(r,a,b)$ | TB $(r,s,a,b)$ | TG $(r+ns,\ a,\ [b^{-a} + \sum_1^n x_i^{-a}]^{-\frac{1}{a}})$ | TB $(r+ns, s,\ a,\ [b^{-a} + \sum_1^n x_i^{-a}]^{-\frac{1}{a}})$ |
| G $(r,y)$ | IG $(s,b)$ | GP $(r,s,b)$ | IG $(s+nr,\ [b + \sum_1^n x_i])$ | GP $(r, s+nr,\ [b + \sum_1^n x_i])$ |
| IG $(s,y)$ | G $(r,b)$ | GP $(r,s,b)$ | G $(r+ns,\ [b^{-1} + \sum_1^n x_i^{-1}]^{-1})$ | GP $(r+ns,\ s,\ [b^{-1} + \sum_1^n x_i^{-1}]^{-1})$ |
| Weib $(a,y)$ | ITG $(s,a,b)$ | Burr $(s,a,b)$ | ITG $(s+n,\ a,\ [b^a + \sum_1^n x_i^a]^{\frac{1}{a}})$ | Burr $(s+n, a, [b^a + \sum_1^n x_i^a]^{\frac{1}{a}})$ |
| IW $(a,y)$ | TG $(r,a,b)$ | GLL $(r,a,b)$ | TG $(r+n,\ a,\ [b^{-a} + \sum_1^n x_i^{-a}]^{-\frac{1}{a}})$ | GLL $(r+n,\ a,\ [b^{-a} + \sum_1^n x_i^{-a}]^{-\frac{1}{a}})$ |
| Expntl $(y)$ | IG $(s,b)$ | Pareto $(s,b)$ | IG $(s+n,\ [b + \sum_1^n x_i])$ | Pareto $(s+nr,\ a,\ [b^a + \sum_1^n x_i])$ |
| IExpntl $(y)$ | G $(r,b)$ | IPareto $(r,b)$ | G $(r+n,\ [b^{-1} + \sum_1^n x_i^{-1}]^{-1})$ | IPareto $(r+n,\ [b^{-1} + \sum_1^n x_i^{-1}]^{-1})$ |
| LN $(y,c^2)$ | LN $(b,q^2)$ | LN $(b,c^2+q^2)$ | LN $\left(e^{\left[\frac{c^2 ln(b) + q^2 \sum_{i=1}^n ln(x_i)}{c^2+nq^2}\right]}, \frac{c^2q^2}{c^2+nq^2}\right)$ | |

(Predictive) $\quad$ LN $\left(e^{\left[\frac{c^2 In(b) + q^2 \sum_{i=1}^n In(x_i)}{c^2+nq^2}\right]},\ c^2 + \frac{c^2q^2}{c^2+nq^2}\right)$

This page inadvertently left blank.