Zeitschrift: Mitteilungen / Schweizerische Vereinigung der Versicherungsmathematiker = Bulletin / Association Suisse des Actuaires = Bulletin / Swiss Association of Actuaries

Herausgeber: Schweizerische Vereinigung der Versicherungsmathematiker

Band: - (1990)

Heft: 2

Artikel: Up the misty staircase with credibility theory

Autor: Jewell, William S.

DOI: https://doi.org/10.5169/seals-967246

B. Wissenschaftliche Mitteilungen

WILLIAM S. JEWELL, Berkeley

Up the Misty Staircase with Credibility Theory*

1 Introduction

Distinguished colleagues, grateful doctoral students and warm friends of Professor Doctor Bühlmann, dear Hans! It is a great privilege to be here to share with you in the celebration of the sixtieth birthday of this enthusiastic, hard-working yet fun-loving, persuasive yet open-minded man of broad interests and wide influence who means so much to us all.

In thinking about what remarks might be appropriate today, I could not help but recall how I first came into Hans' orbit; I hope you will permit me a few reminiscences. In 1969, consulting for an insurance company led me into contact with that arcane applied statistician, the actuary. I immediately saw the value of these models and methods, hidden away in obscure journals, to my own field of operations research and I began to study and teach from *Seal's* book (1969), which had just appeared. In this way, I learned about the ASTIN Section of the International Actuarial Association and their Colloquia on insurance mathematics, and became familiar with the names of the giants in the field. Because a U.C. colleague had met and exchanged ideas with *Erwin Straub,* we were later visited in Berkeley by *Jörg Hofmann,* also of the Swiss Reinsurance Company; from him I heard about the intensive actuarial activity in Switzerland. A trip to Europe in 1972 gave me the opportunity to visit Zürich for the first time, and so I offered to give a talk – I believe it was at the Swiss Life Insurance and Pension Company. Here I met Hans and many others of you here today and made plans to attend the 10th ASTIN Colloquium in Essex, England the next year, where my acquaintance with Hans deepened. By then I was well aware of classical credibility theory and began to present my own ideas at subsequent ASTIN Colloquia and to participate in other European actuarial activities. In spite of the fact that I am an engineer, I was, as we say in English, "hooked" on actuarial science.

Since that time, my friendship with Hans Bühlmann has grown immeasurably, helped, no doubt, by the fact that we are of the same generation and have

* Festreferat, gehalten am Symposium aus Anlass des 60. Geburtstages von Prof. Hans Bühlmann, 2. Februar 1990 an der ETH Zürich

many of the same values but also because Hans studied at the University of California, brought his new bride to the Bay Area, and because their daughter was born there! In slightly altered circumstances, we might have met many years earlier. But we have since made up for lost time through our reciprocal visits to Zürich and Berkeley.
What I would like to do today is to describe Hans' great influence on the subject of credibility theory and upon my own research, and to put this influence into context in the broader realm statistical inference, both classical and Bayesian. Although this particular topic may be of only passing interest to many of you, I hope that it will serve as a paradigm for the many ways in which Hans' ideas have broken new ground in the fields of statistics, insurance, and education. If it helps you in your own remembrances on this happy occasion, so much the better.

2 The Misty Staircase of Traditional Statistics

The title of my talk is taken from *Mosteller/Tukey* (1968/1977), whose analogy (Fig. 1) is worth quoting at some length:

> Before Student's time, every analysis of data that considered "what might have been" resembled a long staircase from the near foreground to the misty height. One began by calculating a primary statistic, a number that indicated quite directly what the data seemed to say about the point at issue. The primary statistic might, for instance, have been a sample mean. Then one faced the question of "How much different might its value have been?" and calculated a secondary statistic, a number that indicated quite directly how variable (or perhaps how stable and invariable) the primary statistic seemed to be. The secondary statistic might have been an estimate of the standard deviation of such a sample mean. After this step one again needed to face the question of "How much different?", this time for the secondary statistic, which again and again turned out to be less stable (of itself) than the primary statistic whose stability it indicated. In principle, one should have gone on to a tertiary statistic, which indicated the variability or stability of the secondary statistic, then to a quaternary statistic, ..., and so on up and up a staircase which, since the tertiary was a poorer indicator than the secondary, and the quaternary was even worse, could only be pictured as becoming mistier and mistier. In practice, workers usually stopped with primary and secondary statistics.

They then go on to describe the contribution made by "Student" (*W.S. Gossett*) with the famous *t-ratio* which, under normality assumptions, describes the standardized variability of the sample mean as a function only of the number of samples.

> This approach cuts off the misty staircase after the third step – indeed, almost after the second step. For, in order to tell us about the population mean, the data were asked to provide only:

1. the sample mean – a primary statistic,
2. the sample estimate of variance – a secondary statistic,
3. the sample size – a tertiary statistic, one that was easy to obtain and remarkably stable, at least so long as one compared this sample with other samples of the same size.

All else was provided by the assumption of exact normality.

Mosteller/Tukey then discuss the use of the *t-ratio* and other classical statistics that led to the development of the "whole machinery of significance testing and almost all the machinery used in practice to set confidence intervals", concluding with the observation that, in the 1930's and 1940's, people learned to short-cut the staircase through "non-parametric" or "distribution-free" procedures.

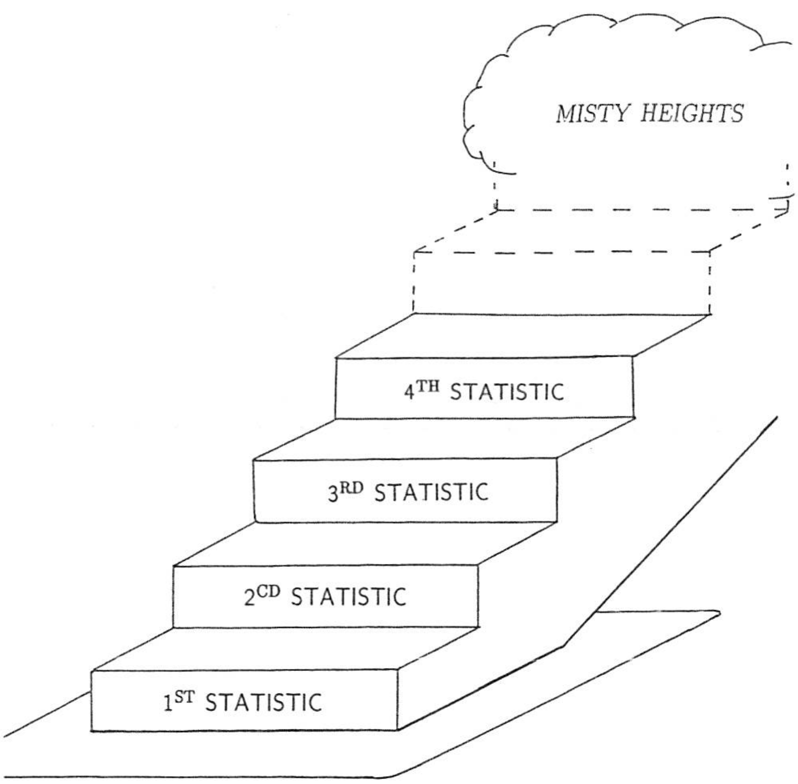

Figure 1 The Misty Staircase of Traditional Statistics

By describing the misty staircase of classical statistics, I do not mean to imply that the procedures of significance testing and confidence intervals have always stood the test of time and the challenge of new schools of thought; my own views on this issue are somewhat different and would only lead me into an argument with those of you of the frequentist persuasion. But it is interesting and instructive to see how imaginative but empirically-based procedures lead to increasingly esoteric argumentation that must eventually be supplanted by new paradigms.

3 The Development of Credibility Theory

To describe the corresponding *misty staircase of credibility theory* we now need to review briefly its historical development. The name *credibility* was given by American casualty actuaries to heuristic formulae developed over 70 years ago for insurance ratemaking with the original ideas developed especially by *A.W. Whitney* and *A.H. Mowbray* (sometime professor at U.C., Berkeley!). If we suppose that we have n years of *experience data*, $\mathscr{D} = \{x_1, x_2, \ldots, x_n\}$, for a given *individual risk* (insurance contract), the basic problem of the actuary is to *rate* the risk, that is, to find the *fair premium* (mean value) for the unknown outcome for next year, x_{n+1} (the x's may refer to the number of claims during one year, the total cost of such claims, the average cost per claim, etc.).
Under the assumption that the x's are independent random samples from some stationary distribution of outcomes for *this* risk with unknown mean μ, finding the fair premium is the same as estimating μ from the data.
These early pioneers knew of course that the sample mean, $\overline{x} = \sum x_t/n$, was an appropriate *experience premium*, i.e. a good estimator of μ, but they were worried about its variability, which they knew was large when n was small (sound familiar?). On the other hand, insurance companies usually have available a much larger historical data base for a *portfolio* of apparently similar risks to which the individual risk belongs. They also were aware that, no matter how hard they tried to group risks into portfolios that were similar from a statistical point of view, such a collection of contracts always contains some residual *heterogeneity;* in other words (and to anticipate later arguments), they already thought of the mean value of each risk as a kind of *random quantity*, $\widetilde{\mu}$, that varied over the portfolio. However, they believed that the average of these values, call it $m = \mathscr{E}\{\widetilde{\mu}\}$, obtained from a grand sample mean over a large amount of portfolio data, would be very stable; this value, m, they called the *manual premium*, since it was the value quoted in rate-making manuals circulated for the purpose of providing quotes on new, similar risks for which no individual experience data was available. Then, using a simple risk-pooling model, they argued for the adoption of an *experience-rated fair premium* of the form:

$$\mathscr{E}\{\widetilde{x} \mid \mathscr{D}\} = \mathscr{E}\{\widetilde{\mu} \mid \mathscr{D}\} \approx (1-z)m + z\,\overline{x}; \quad z = \left[\frac{n}{n+n_0}\right]. \tag{1}$$

The factor z was called the *credibility factor*, since it gives the relative weights to be attached to the manual premium and to the individual sample mean, depending upon the number of samples in the experience; it is thus a kind of *learning curve* that moves the estimate from the no-data estimate m to the classical estimator $\bar{x}$ as n increases. In the limit, the sample mean is "fully credible" and, by the law of large numbers, is almost surely equal to the unknown fixed value, $\tilde{\mu} = \mu$. The *credibility time constant*, n_0, was essentially determined by trial-and-error for different lines of insurance. The formula above was used successfully in American casualty rate-making for more than 50 years, with innumerable variation and elaboration, and remnants of it still remain today. Surveys with references may be found in *Longley-Cook* (1962) and *Hickman* (1975).

This period of development, the *empirical* phase, will form the first step for our *credibility staircase to the misty heights.*

4 Early Bayesian Ideas

The modern development of credibility begins with the resurgence of interest in Bayesian ideas, foreseen by *Bruno de Finetti* in the 1930's, and continued in the 1950's by *L. Jimmie Savage, Dennis V. Lindley, I.J. Good,* and many others. I have previously argued for this approach as the only complete and logical framework for making statistical inferences and decisions in the applied world (*Jewell*, 1980) and will not repeat myself here, except to point you towards *Lindley* (1978), *Barnett* (1982), and an important new book by *Howson/Urbach* (1989), all of whom make better arguments than I could.

The Bayesian view of the rate-making problem is that the unknown individual risk mean, $\tilde{\mu}$, is truly a random variable in the sense that, if we wish to find the conditional mean, $\mathscr{E}\{\tilde{\mu} \mid \mathscr{D}\}$, already implied in (1), then we must use the basic laws of *conditional probability* in the form $p(\mu \mid \mathscr{D}) \propto p(\mathscr{D} \mid \mu) \cdot p(\mu)$. The first term on the RHS is the usual *data likelihood*, $\prod p(x_t \mid \mu)$, formed by independent sampling from the *model density*, about which all statisticians are used to thinking. However, controversy arises because conditional probability requires knowledge of the second term, $p(\mu)$, the *prior density*, which in turn implies knowing about the possible values of μ and their relative occurrence *before the data is observed*. Thus, opining $p(\mu)$ means that the analyst must think carefully about the concrete meaning of the parameter in the problem at hand, and either have *extensive experience* with such risks or be able to live with a *subjective judgement* about the possible outcomes until relevant data can be obtained. In other words, thinking carefully about a *distribution* replaces the frequentist thinking about possible *procedures*. This is hardly a problem for actuaries, engineers, and other applied scientists, who routinely have to make experiential and subjective judgements, but seems to be difficult for classically – trained mathematical statisticians to accept, as they prefer to "let the data speak for itself" without physical interpretation, dimensions, or previous experience; however, the *ad hoc* procedures that result may lead to non-coherence with the laws of probability.

Bailey (1950) seems to be the first to have introduced Bayes' law explicitly into the rate-making model (although Hans cites an early book by *Ove Lundberg*) and showed that the credibility formula of (1) was *exactly* the predictive mean for the Binomial, Poisson, and Normal (with known variance) models, provided that certain special prior densities are used. In spite of a lively published discussion of Bailey's paper, it had little immediate impact upon the actuarial profession, perhaps because the Bayesian approach had not yet found its way into the statistical literature. (Some early actuarial exceptions are *Dropkin* (1950/1960), *Longley-Cook* (1962), *Bichsel* (1964), and *Fürst* (1964)). This prompted *Mayerson* (1964) to take up the theme again with a clear exposé of the same examples in more modern terminology, pointing out, for instance that the priors which give the credibility form are the so-called *natural conjugate priors*. If one were to criticize these early works, it is that they did not consider what might happen in models where parameters other than the mean were *a priori* unknown, and did not explore the meaning of the time constant n_0, because in their results it is just a combination of the (hyper)parameters from the priors.

This period will be referred to as the *early Bayesian* step on our credibility staircase to the misty heights.

5 Linear Least-Squares Approximations

We now are ready to consider the contributions of Hans Bühlmann to credibility. This is, as you can appreciate, a delicate task, since the statistician of whom I speak is seated here in front of me. However, if he disagrees with my attempt to reconstruct history and interpret his work, he should be the first to speak when I have finished!

The 1963 ASTIN Colloquium in Trieste must have been quite stimulating, if one can judge from the papers presented and the subsequent discussion. Here, for the first time, Hans (1964a) describes the connection between the classical collective theory of risk and the claims process (see *Bühlmann* (1975/1989)). Hans' analysis is of the total claims cost per risk, determining its mean and variance using *conditional expectation* under two main hypotheses, which he first discusses and whose consequences he then explores. For the simpler claims prediction problem of Section 3, these hypotheses translate as follows:

1. If one considers a *portfolio* of *r distinct risks,* their individual claims experiences over n years:

   $$\mathscr{D}_i = \{x_{i1}, x_{i2}, \dots, x_{in}\} \qquad (i = 1, 2, \dots, r)$$

   are *statistically independent* of each other, but are *similar,* because each experience is chosen independently from a very large *collective,* a kind of sample space of similar risks;
2. However, the experiences between *any two risks* are *different* because of residual *heterogeneity in the collective,* and hence in the portfolio;
4. The experience of any *single risk* is *homogeneous over time* because the claim amounts are chosen from the same distribution;
5. However, the claim amounts for a single risk *cannot be independent* of each other over time, but must be *exchangeable random variables.*

Clearly, the concept of exchangeable random variables, due originally to *Bruno de Finetti,* was on Hans' mind, as he had just completed his thesis (1960) on this topic. I also like to think that he already realized that one cannot predict a future outcome from past data assumed to be i.i.d. samples, because some form of dependency over time for the individual risk must be assumed, the simplest of which ist exchangeability. Perhaps Hans even realized that the variability in $\tilde{\mu}$, described previously, was the key to this assumption, since de Finetti's Theorem on $0-1$ variables can be extended to give, in an obvious notation, the desired *prior* dependence between outcomes from a single risk:

$$p(x_1, x_2, \dots, x_n) = \int \prod_t p(x_t \mid \mu) p(\mu)\, d\mu \,. \tag{2}$$

Note that the marginal densities of the x's are dependent *if we do not know the value of the unknown parameter,* but independent if we do! This distinction is at the heart of Bayesian modelling, namely, that our attitude towards independence and dependence *depends upon the knowledge at our disposal when we make the judgement!*

Then, Hans' (1964b) paper refers for the first time to the use of "sequential estimation" to determine the "true premium rate"(the particular μ for a given risk?) and the need for a *prior distribution* on this premium rate, viewed as a random parameter. Perhaps by this time the work of *Bichsel* (1964), *de Finetti* (1964), *Fürst* (1964), or *Mayerson* (1964) had sent him to Bailey's pioneering paper (1950), or perhaps he was familiar with the relationship of Jimmie Savage's work to that of de Finetti's. Hans: how did you discover the Bayesian approach? It was certainly not in the Statistics Department at Berkeley in the 1950's![1]

Many other pieces of the puzzle were in place by the time of the 1965 ASTIN Colloqium in Lucerne. Least-squares methods have been in the armorium of every scientist since the time of Gauss, LaPlace, and Legendre. Thus, Bailey's neglected (1945) paper on estimating unknown population means for different "characteristics" used linear regression to form classical precision-weighted estimates and then related them to the credibility form, with an interpretation of the time constant in terms of the underlying precisions. And Hans, who wrote the Introductory Report on Subject I, Experience Rating in Credibility for the Lucerne Colloquium, had already seen the paper of *Bichsel* (1967) and other related works by Lundberg and Franckx. So it is almost as an afterthought that Hans adds Sections 4 and 5 to his paper. In these, he clearly states that:

1. Each risk in the portfolio can be thought of a having its own abstract parameter θ and the observed samples for a given risk are drawn from the same *conditional distribution* with mean $\mu(\theta)$ and variance $\sigma^2(\theta)$, thus giving "homogeneity *in time,* but *not* in mass [the portfolio]". (Actually, Hans slips up a bit and says this means i.i.d. samples, when he intends exchangeable r.v.s., that is, independent samples, *given* θ);
2. The appropriate *actuarial* estimator for the unknown $\mu(\theta)$ is the *a posteriori* mean, $\mathscr{E}\{\mu(\tilde{\theta}) \mid \mathscr{D}\}$, justified under a postulate of equilibrium that "each class of risks with *equal observed risk performance* should pay its own way";
3. To calculate this conditional expectation one must average over the "*a priori* distribution" $p(\theta)$, which is described as the "structural function of the *portfolio*", i.e. the prior distribution over the collective;
4. "The credibility formula used by our American colleagues is nothing but a linearization of the ... estimator function", $\mathscr{E}\{\mu(\tilde{\theta}) \mid \mathscr{D}\}$.
5. And finally, after some simple least squares, Hans expresses the estimator in credibility form *and* determines that the time constant is just:

$$n_0 = \mathscr{E}\{\, \sigma^2(\tilde{\theta}) \,\} \mid \mathscr{V}\{\, \mu(\tilde{\theta}) \,\}.$$

[1] At the Symposium, the honoree responded to the effect that, when shown the Bayesian methodology at Berkeley, it was described as "the wrong approach".

Well, as Alexander Pope said about Sir Isaac Newton:

"Nature and Nature's laws lay hid in night: God said, Let Newton be! and all was light."

Suddenly, after the publication of Hans' paper in 1967, all the academically-inclined actuaries of the world understood the rationale for credibility, doctoral students at ETH and elsewhere had a new topic for theses, and working actuaries had a new approach to rate-making with which to experiment. I myself have benefitted from full employment and a great deal of pleasure in following up on these ideas, in particular being able to find a class of model densities and priors in which the credibility approximation is exact (1974/1975a), thus tying Hans' results to those of Section 4 and giving added support to the robustness of credibility approximations. Bühlmann's insight was an idea whose time had come.

The result has been a virtual explosion in "new-wave" credibility literature. *Jewell* classifies the important models as of 1980; *de Wit* (1986) attempts to survey the entire literature on credibility topics from 1855 through 1981; and the end is not yet in sight – I could keep you here all day describing the many extensions and elaborations that have been developed. Of course, many of these papers (including my own) are of the academic puzzle-solving variety. But there is still much "normal science" waiting to be done in applying least – squares Bayesian approximations to rate-making. It is a giant third step on our credibility staircase.

6 Apostasy in Structural Parameter Estimation

In a Bayesian formulation, one must specify both the model density, $p(x \mid \theta)$, and a prior density, $p(\theta)$, based upon physical evidence, actual experience, and informed judgement. On the other hand, a linear approximation to the predictive mean in the basic model requires only the first two model moments, $m(\theta)$ and $v(\theta)$, and three averages over the collective, $m = \mathscr{E}\{m(\tilde{\theta})\}$, $e = \mathscr{E}\{v(\tilde{\theta})\}$, and $d = \mathscr{V}\{m(\tilde{\theta})\}$. Because $\tilde{\theta}$ is now an abstract random parameter, $m(\tilde{\theta})$ and $v(\tilde{\theta})$ can be: strongly dependent, as in the QVF-NEF family of Morris (see *Jewell/Schnieper* (1985)); completely independent, as in the usual normal model with unknown mean and variance; or with $v(\tilde{\theta}) = v$ constant. The key point is that credibility applied to the basic model requires our opinion about only *three moment hyperparameters,* rather than about *two distributions.*

Immediately after Bühlmann's 1967 result, actuaries began to think about *procedures* to obtain *m*, *e*, and *d* from *cohort data* on other risks. Consider

a *portfolio of related risks* $(i = 1, 2, \ldots, r)$, each with its own abstract parameter, θ_i, drawn independently from the same collective structure density, $p(\theta)$; for convenience, we shall assume that each risk has the same data record length, n, so that the *total cohort data* is the $r \times n$ collection $\mathscr{D} = \{x_{it}; (1, 2, \ldots, r)(t = 1, 2, \ldots, n)\}$. Attention now focusses on how to use *all* of $\mathscr{D}$ to predict the next observation of *one* of the risks, say $\tilde{x}_{1,n+1}$.
An immediate difficulty is that, if the components of $\theta = [\theta_1, \theta_2, \ldots, \theta_r]^T$ are truly i.i.d., then the likelihood is $p(\mathscr{D} \mid \theta) = \prod\prod p(x_{it} \mid \theta_i) p(\theta_i)$, so that the data from risks $i = 2, 3, \ldots r$ *contain no predictive information for risk #1*! Nevertheless, actuaries still feel that there must be *some* use for this information. *Bühlmann/Straub* (1970), in a paper notable for its introduction of *volume-weighted samples,* propose forcing the inclusion of cohort data by using a *homogeneous* linear predictor to obtain the new credibility formula:

$$\mathscr{E}\{\tilde{x}_{1,n+1} \mid \mathscr{D}\} \approx (1-z)\bar{x}_{..} + z\bar{x}_{1.};$$
$$\bar{x}_{i.} = \sum_t \frac{x_{it}}{n}; \qquad \bar{x}_{..} = \sum_i \frac{\bar{x}_{i.}}{r} = \sum_i \sum_t \frac{x_{it}}{nr}, \tag{3}$$

but with the same credibility factor, $z = \frac{n}{n+n_0}$, $n_0 = \frac{e}{d}$. In effect, the collective mean hyperparameter, m, is replaced by the *portfolio grand mean,* $\bar{x}_{..}$, over all risks, including #1. They and other later authors then suggest that the natural way to implement the original credibility formula (1) to rate any individual risk is to use the classical estimates:

$$\hat{m} = \bar{x}_{..}; \qquad \hat{e} = \frac{1}{r}\sum_i \left[\frac{1}{n-1}\sum_t (x_{it} - \bar{x}_{i.})^2\right];$$
$$\hat{d} + \frac{1}{n}\hat{e} = \frac{1}{r-1}\sum_i (\bar{x}_{i.} - \bar{x}_{..})^2. \tag{4}$$

This approach is justified on the basis of being (classically) unbiased, and (asymptotically) "efficient", etc. The subsequent literature is then full of incredibly complex estimates of the hyperparameters needed for the increasing elaborate models to which credibility is applied (*Loimaranta* (1977), *De Vylder* (1978), *Sundt* (1979), *Norberg* (1980/1981), etc.). In fact many so-called papers on credibility theory from this period are mostly about sampling-school estimates for the necessary hyperparameters.
To my mind, this "new step" on the credibility staircase is a form of intellectual backsliding, in which traditional large-sample techniques are married to an approximation based upon Bayesian concepts. Not only does (3) not obey

the likelihood principle, it gives a mean-square error greater than that of the usual inhomogeneous form!

Further, there seems to be some confusion between the underlying concept of the *collective*, with its structure function, $p(\theta)$, and the *portfolio of a finite sample of risks*, $(i = 1, 2, \ldots, r)$, with its data set, $\mathscr{D}$. (See *inter alia, Straub* (1975), and the discussion to *Jewell* (1975c). As I remarked in 1980:

> But, what about the possibility of using the empirically derived [(4) with (1)] for arbitrary [n and r] – couldn't this also have some nice, robust properties? Well, yes, I would have to admit – but they haven't been demonstrated yet! In other words, if you propose to "ad hoc" up a complicated formula involving both sums, squared sums, and sums of squares of the data by appeal to two different schools of thought, then I can only be amazed by your ingenuity. But I suspect that you will have a difficult time in proving these properties analytically, and will have to resort to, say, simulation (that is, to experience).

In short, I believe that such hybrid approaches are, like the two-legged stool, doomed to failure because they rely on conflicting tenets, and can only lead to inconsistencies in either frame of reference. After all, a pure frequentist approach would select $\bar{x}_{1.}$ as the "best" point estimator for risk #1. We cannot rely upon such ad-hoc approaches for a solid fourth step to our credibility staircase.

7 The Hierarchical Model

In my view, if one wishes to include all of the data from a portfolio in a prediction for one of its risks, the *model* and hence the likelihood must be modified so that there is predictive information *between* risks. In other words, the i.i.d. assumption for the $\{\theta_i\}$ must be incorrect.

Based upon a normal model of *Lindley/Smith* (1972) and an idea of *Taylor* (1974), I suggested in 1975 that the compatible fourth-step extension of the basic model is a *hierarchical* generalization, in which *our* collective, the one from which *our* portfolio is drawn, is not necessarily the same as other supposedly similar collectives, for example, those of other insurance companies. This leads us to imagine that all such collectives are drawn from some *super-collective or universe of collectives*, or, if you prefer, from an *urn of urns*. Formally, this means that the individual risk parameters are drawn from a conditional prior density, $p(\theta \mid \phi)$, labelled by some abstract *unknown* hyperparameter, ϕ, which characterizes *our* collective. We now must visualize the manner in which the random $\tilde{\phi}$ varies over the universe of collectives by specifying a *hyperprior density*, $p(\phi)$. In this way, the various $\{\theta_i\}$ in *our* portfolio will become *exchangeable random variables*, governed by a joint density similar to (2). With this inter-risk dependence, the cohort data in $\mathscr{D}$ will enable us to "learn" about *our* value of $\tilde{\phi} = \phi$, and hence to improve any individual risk prediction in our portfolio.

Stated another way, if an actuary has no extensive experience or strong opinion relevant to the m, e, and d needed to rate an individual risk, he/she must treat them as unknown random quantities, $\tilde{m} = m(\tilde{\phi})$, $\tilde{e} = e(\tilde{\phi})$, $\tilde{d} = d(\tilde{\phi})$. *Jewell* (1975) then finds the optimal least-squares approximation from the data $\mathscr{D}$ to be the *joint* prediction:

$$\begin{aligned} \mathscr{E}\{\tilde{x}_{1,n+1} \mid \mathscr{D}\} &\approx (1 - z_1)\,\mathscr{E}\{m(\tilde{\phi}) \mid \mathscr{D}\} + z_1 \bar{x}_{1.}; \\ \mathscr{E}\{m(\tilde{\phi}) \mid \mathscr{D}\} &\approx (1 - z_0)m + z_0 \bar{x}_{..}, \end{aligned} \tag{5}$$

where now:

$$z_1 = \frac{n}{n + (f/g)}; \quad z_0 = \frac{r z_1}{r z_1 + (g/h)} = \frac{rn}{rn + n(g/h) + (f/g)}, \tag{6}$$

for which we now require *four hyper-hyper-parameters:*

$$m = \mathscr{E}\{m(\tilde{\phi})\}; \quad f = \mathscr{E}\{e(\tilde{\phi})\}; \quad g = \mathscr{E}\{d(\tilde{\phi})\}; \quad h = \mathscr{V}\{m(\tilde{\phi})\}, \tag{7}$$

found by averaging over the universe of possible collectives.

The interpretation of (5) is straightforward: beginning with the universal parameters, knowledge about *our* $m(\phi)$ is updated using the grand sample mean for our collective, $\bar{x}_{..}$; the updated value replaces the (previous mean) "m" in an individual risk credibility form that, of course, is the same for any individual risk from our portfolio. Note, however, that the ratios of the three *a priori* components of total variance, f, g, and h, are used in the credibility factors, and there is no explicit learning about portfolio or individual variability. The forecasts (5) are known to be exact Bayesian predicitive means in the normal-normal-normal, fixed variance hierarchical model of *Lindley/Smith* (1972), and for a heteroscedastic generalization in *Jewell* (1987). Additional interpretations may be found in *Jewell* (1975), and later generalizations in *Taylor* (1979), *Sundt* (1979), *Bühlmann/Jewell* (1987), and elsewhere.

This result provides further insight into the Bühlmann-Straub heuristic, namely, that $\hat{m} = \bar{x}_{..}$ can replace the previous assumed-constant collective mean, "m", in the original model *only* if $z_0 \approx 1$! But we see from (6) that this can happen *only* if the product rn is large, and *not* if n alone is large. In other words, we require data from a large number of risks so that this portfolio will be representative of *our* collective! If only n is large, it is true that $\bar{x}_{1.}$ will be "fully credible" for predicting risk #1, but we will still be uncertain about *our* $m(\phi)$.

The fact that the actuary must still provide the universal moments m, f, g, and h will probably disappoint those hoping to eliminate prior experience and opinion from credibility theory. However, even I would admit that we are now on more solid ground for using large-sample estimation methods, since the universe of collectives will be represented by a very large, perhaps nationwide data bank of many different portfolios from many different insurance companies. These higher-level statistics would then correctly initialize the learning process at the corporate and individual risk level, thus, as *Dennis Lindley* likes to say, "extending the conversation". Since the Bayesian position has perhaps been misunderstood, it should be made clear that no one proposes eliminating classical estimators entirely, but merely restricting their application to those portions of a model where the availability of large amounts of data guarantees that the estimates are, almost surely, close to their true values. After all, Bayesian point estimators also converge for large samples to the same underlying values as the classical estimators, in general – it's just that the Bayesian approach gives an estimate *with a probabilistic interpretation* for *any* sample size! In short, hierarchical models consistently use and show the importance of cohort data, thus forming a solid fourth step in our credibility staircase.

8 Decomposed Approximations to Credibility Prediction

Because many of my colleagues have remained unconvinced by the arguments above, I have continued to think about ways in which one might be able to get and use some approximate $\widehat{m}$, $\widehat{d}$, and $\widehat{e}$ *within the Bayesian framework*. The obvious obstacle is that the model must also be able to predict *variances*, not just *means*, for the individual risk.

In 1985 *René Schnieper* and I considered the joint prediction of both the first and second moments for a single risk using credibility approximations, and were able to obtain useful (and sometimes exact Bayesian) results with a three-dimensional formulation at a price, however, of requiring now *eleven* first through fourth hyperparameters! This was not surprising, since the more one requires of a model, the more one has to specify in advance, and we are still better off than having to specify two distributional forms. So, when Hans Bühlmann visited Berkeley in 1986, it was natural to discuss extending this model to the hierarchical case. The methodology to carry out this program was soon apparent to us, and we were able to obtain some intermediate results in a complicated and messy formulation. After three more years of development and supporting results (*Jewell* (1987/1989a)) plus extended computational testing, it was finally possible to shape these joint efforts into the article that appears in the Jubilee issue of the Bulletin of the Association of Swiss Actuaries.[2] You may see the results there, although, frankly, it is not light reading, nor was it intended to be much more than academic puzzle-solving. As the joke goes about the scientists who developed a dehydrated elephant: "It is not very useful, but it is interesting to see what can be done"!

But there ist a constructive and enlightening way in which the results of the paper might be used in a *decomposition approximation* to the credibility approximation. Suppose that the super-actuary in an industry-wide rating bureau had enough data to develop very good large-sample estimates for the 24 universal moments required by the model of the paper. These moments would then be given to the chief actuary of our company, who could solve the 8-dimensional credibility formula in Section 8 of the paper.

Now suppose that our chief actuary (being a very busy person) decides to compute only the *corporate-level predictands* in block 00, i.e. to solve for the 4×4 credibility matrix, $Z^{00} = R^{00}(C^{00})^{-1}$, the last equation in (9.1)*. Using the portfolio-level statistics, $[y_0, y_{00}, y_{0\times 0}, y_{0*0}]^T$, this will give him a forecast of $[M_1(\tilde{\phi}), M_2(\tilde{\phi}), M_{11}(\tilde{\phi}), M_1^2(\tilde{\phi})]^T$. No information has yet been lost, because the staff actuary in charge of rating individual risks could obtain individual-level results from the remainder of (9.1)*.

Now imagine that even this part of the task was beyond the powers of a journeyman actuary, who perhaps has just mastered the simple individual credibility formula, and needs only good values of m, e, and d to rate each risk. At the price of a *great simplification,* the chief actuary decides to decompose the problem by furnishing his subordinate with his best corporate estimates of $(\tilde{\phi})$, $e(\tilde{\phi})$ and $d(\tilde{\phi})$, and letting the subordinate use the standard formula with y_1 only, ignoring any predictive value in the seven other statistics, $[y_0, y_{00}, y_{0\times 0}, y_{0*0}; y_{11}, y_{1\times 1}, y_{1*0}]^T$!

By examining the relationships between central moments and moments about the origin, the chief actuary finds that:

$$m(\phi) = M_1(\phi); \quad e(\phi) = M_2(\phi) - M_{11}(\phi); \quad d(\phi) = M_{11}(\phi) - M_1^2(\phi), \tag{8}$$

[2] For the rest of this section we assume familiarity with the notation of *Jewell* (1989b), and denote equations from it as, e.g. (9.1)*.

so that predictions of these values can be found using the original formulation. But now suppose that the chief actuary decides to use only the *three* statistics of (4), which can be rewritten:

$$\widehat{m} = y_0; \quad \widehat{e} = y_{00} - y_{0\times 0}; \quad \widehat{d} = y_{0\times 0} - y_{0*0}. \tag{9}$$

As one of the predictands is dropped, there is a loss of portfolio-level information, and the chief actuary now has a 3×3 set of linear equations to solve that are a linear transformation of the original problem.
Using fraktur notation for the reduced problem, the chief actuary now predicts $\widetilde{\mathfrak{w}} = [m(\widetilde{\phi}), e(\widetilde{\phi}), d(\widetilde{\phi})]^T$ in terms of $\widetilde{\mathfrak{y}} = [\widehat{m}, \widehat{e}, \widehat{d}]^T$, using the 3-dimensional credibility formula:

$$\mathscr{E}\{\widetilde{\mathfrak{w}} \mid \mathscr{D}\} \approx (\mathfrak{I} - \mathfrak{Z})\mathfrak{m} + \mathfrak{Z}\mathfrak{y}, \tag{10}$$

with $\mathscr{E}\{\widetilde{\mathfrak{w}}\} = \mathscr{E}\{\widetilde{\mathfrak{y}}\} = \mathfrak{m} = [M(1), M(2) - M(11), M(11) - M(1;1)]^T$ and a 3×3 credibility matrix:

$$\mathfrak{Z} = (KR^{00}K^T)(KC^{00}K^T)^{-1} \tag{11}$$

where K is the dimension-reducing transformation matrix:

$$K = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & ^-1 & 0 \\ 0 & 0 & 1 & ^-1 \end{bmatrix}. \tag{12}$$

Now, *if* the restriction to using only y_1 in the basic credibility formula *and* the reduction of dimensionality in estimating $\widetilde{\mathfrak{w}}$ do not throw away significant information, and *if* the size of r and n is such that $\mathfrak{Z} \approx \mathfrak{I}$ for this reduced problem, *then* we could say that the sampling-empiristic-Bayesians have a valid approach. Otherwise, the first equation of (9.2)* would give a better result because; (i) it is valid for any r and n; (ii) it includes the conjoint variation between *four* individual statistics and *four* portfolio statistics.
Also, the only model known for which the decomposed approach is the same as the credibility approximation and the exact Bayesian result is the heteroscedastic normal-normal-normal model of *Jewell* (1987); with the exception that all *four* portfolio statistics, not just $\mathfrak{y}$, are needed to find the predictions of $e(\widetilde{\phi})$ and $d(\widetilde{\phi})$, which turn out to have a constant ratio, n_0!
Figure 2 shows the misty staircase of credibility theory that we have constructed thus far. There are many steps that could be added, but it

is increasingly easy to go astray in the misty heights, as more and more hyperparameters are required.

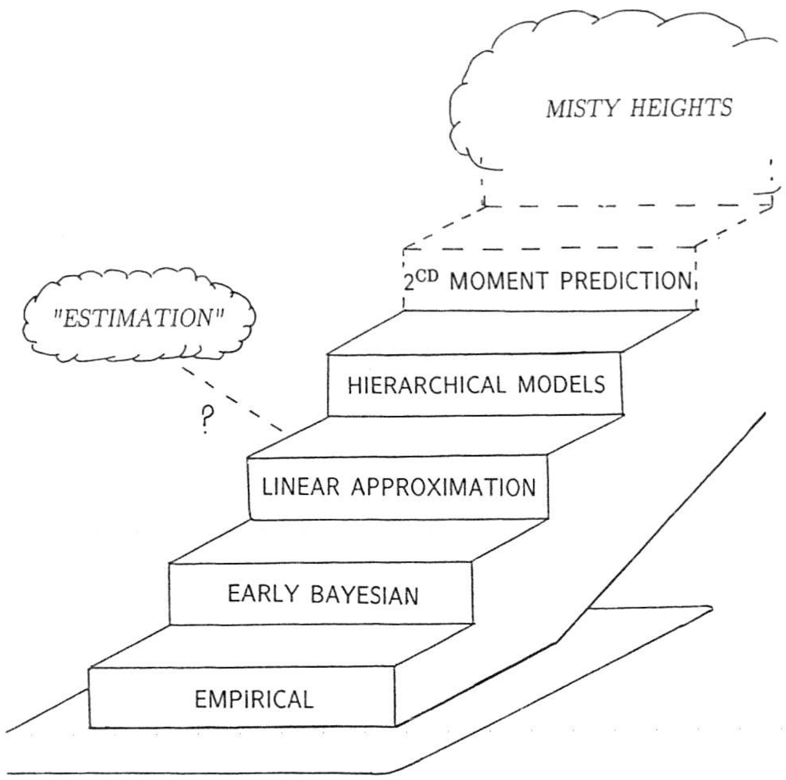

Figure 2 The Misty Staircase of Credibility Theory

9 The Bayesian Escalator

By now, the basic theme of this paper should be apparent. By following the logical progression of traditional statistics, we are led up a staircase of progressively shakier steps to misty heights that are far from reality. In credibility theory, too, although we start with a firm footing, we quickly reach esoteric models that require a knowledge of a large number of hyperparameters. Further, as we have seen in Section 6, it is easy to go off the track by trying to combine traditional estimation procedures with Bayesian approximations. The reason for this is, I think, because credibility theory has somehow been classified as an "empirical Bayes" approach, and the word "empirical" suggests that one can try anything that seems reasonable. It is only human to try to get something for nothing, so that, rather than thinking hard about the problem itself and admitting informed judgement on an equal footing with analysis, it seems easier to "cook up" procedures based only upon data.

Personally, I prefer to rely upon what might be called the *Bayesian escalator* to reach new heights of understanding (Fig. 3). This staircase begins with two giant fixed steps. The first is the acceptance of the laws of conditional probability *in toto,* examining all of my assumptions to see if any of them violate the basic axioms and trying to eliminate any that might lead to non-coherence. The next big step is to realize that this approach always means that I must posit some prior information or attitude about the problem *before* I begin the analysis. This prior information may be quite imprecise, but this simply means that, without many new observations, my opinion will not be sharpened very much. This admittance of experience, expert opinion, "know-how" etc. *can* use supporting large-sample methods when they are accurate, but *cannot* be *entirely* analytic – the analyst *is* part of the modelling process and must take responsibility for and make explicit *all* of his assumptions and judgements. Remember that, etymologically speaking, *experience* and *experiment* are closely related.

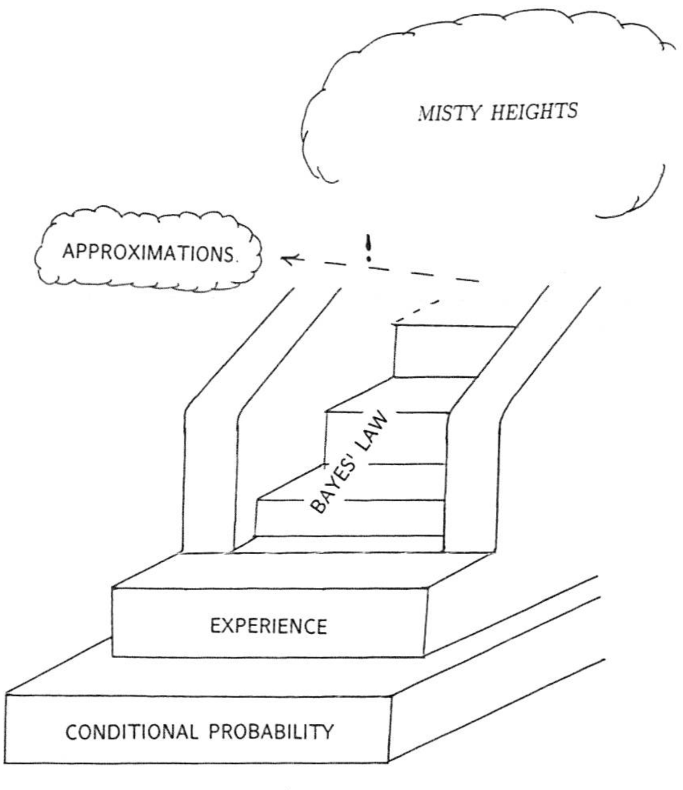

Figure 3 The Bayesian Escalator

From this point on, we have a Bayesian moving staircase whose logic will automatically help us to mount to the desired heights. More elaborate models at higher elevations are more difficult, to be sure, but these are mostly difficulties with dimensionality and computation of integrals, not in probabilistic interpretation. If the computational labor becomes too costly, we can then step off the Bayesian escalator and follow empirical simplifications and approximations without getting lost in the misty heights.

10 Forecasting Excess Losses using Credibility Theory

To illustrate in concrete terms the way in which the Bayesian paradigm interacts with and supports credibility theory, let us consider an important problem in reinsurance modelling – the forecasting of *excess-of-loss claims.* This problem has previously been analyzed using credibility by *Straub* (1971) and *Patrik/Mashitz* (1989); *Jung* (1964) and *Bühlmann* (1975) have underlined the dangers of using only extreme value data; and *Fürst* (1964) studied a simplified Bayesian model.

10.1 Basic Claims Model and Notation

The basic model of claims generation assumes that a variable number, n, of claims occurs during some fixed *exposure interval,* say T years, in amounts $\{x_1, x_2, \ldots, x_n\}$. The *total cost* of these claims, $w = \sum x_i$, is the quantity of interest. In the usual model, both the number of claims (the *frequency)* and the sizes of the claims (the *severities)* are assumed to be random variables, mutually independent, given the parameters. Specifically, we assume:

- the random number of claims, $\tilde{n} = 0, 1, 2, \ldots$, has a discrete counting density, $p_n(\lambda, T)$, that depends upon a *frequency* parameter, λ. In the sequel we shall use the *Poisson* (λT) for simplicity;
- Each random claim amount, $\tilde{x}_i \geq 0$ $(i = 1, 2, \ldots, n)$, has the same common density, $p(x \mid \varphi)$, that depends upon (one or more) *severity* parameters(s), φ;
- the $\tilde{x}_i$ are independent, given φ and n, and the $\tilde{x}_i$ and $\tilde{n}$ are statistically independent of each other, given λ and φ.

For later convenience, we define the following:

$$P(x \mid \varphi) = \mathscr{P}r\{\tilde{x} \leq x \mid \varphi\}; \quad Q(x \mid \varphi) = \mathscr{P}r\{\tilde{x} > x \mid \varphi\};$$

$$R(x \mid \varphi) = \int_x^\infty Q(u \mid \varphi)\, du; \qquad S(x \mid \varphi) = \int_x^\infty R(u \mid \varphi)\, du; \tag{13}$$

$$m(\varphi) = \mathscr{E}\{\tilde{x} \mid \varphi\} = R(0 \mid \varphi); \quad v(\varphi) = \mathscr{V}\{\tilde{x} \mid \varphi\} = 2S(0 \mid \varphi) - m(\varphi)^2.$$

P and Q are the usual cumulative and tail distribution functions; R is $\mathscr{E}\{\tilde{x} - x)^+ \mid \varphi\}$, the so-called *stop-loss fair premium,* whose second moment is $2S$; m and v are the usual conditional mean and variance for the claim severity.

Suppose that we wish to predict the mean total claim amount (the *total severity)* during some future interval of U years, using the well-known result:

$$\mathscr{E}\{\tilde{w} \mid \lambda, \varphi\} = \mathscr{E}\{\tilde{n} \mid \lambda, U\} \cdot m(\varphi) = \lambda U \cdot m(\varphi). \tag{14}$$

If the frequency and severity parameters were known, the result would be immediate; however, in the usual case, we are uncertain about the parameters, but can provide a prior opinion about their possible values, based upon our previous experience, industry-wide studies, etc. Making the usual assumption that $\tilde{\lambda}$ and $\tilde{\varphi}$, now treated like random quantities, are *a priori* independent, then our opinion will be summarized in the form of two prior densities, $p(\lambda)$ and $p(\varphi)$. But without further information it is clear that the prediction of mean severity is given simply by the product of the *a priori* average mean frequency times the *a priori* average mean severity.

Now, suppose that we can gather *all* the claim values from the contract(s) of interest during some previous *observation interval* of, say, T years during which the parameters had the same (unknown) values. This then provides data in the form $\mathscr{D} = \{n; x_1, x_2, \ldots, x_n\}$ from which one can "learn" more about the parameters and thence make a better prediction of a future $\tilde{w}$. To do this, we find first the likelihood of the data:

$$p(\mathscr{D} \mid \lambda, \varphi) = p_n(\lambda, T) \prod_{i=1}^{n} p(x_i \mid \varphi), \tag{15}$$

and then use Bayes' Law to form the posterior-to-data parameter density:

$$\begin{aligned} p(\lambda, \varphi \mid \mathscr{D}) &\propto p(\mathscr{D} \mid \lambda, \varphi) p(\lambda, \varphi) \\ &= \left[p_n(\lambda, T) p(\lambda)\right] \left[\prod_{i=1}^{n} p(x_i \mid \varphi) p(\varphi)\right] \propto p(\lambda \mid n) p(\varphi \mid \mathscr{D}). \end{aligned} \tag{16}$$

In other words, because of the assumption of independent priors and the factorization of the likelihood, the posterior-to-data random parameters will still be independent! Thus our updated prediction using (14) and (16) will also be the product of two estimates:

$$\mathscr{E}\{\tilde{w} \mid \mathscr{D}\} = \mathscr{E}\{\tilde{\lambda}U \mid n\} \cdot \mathscr{E}\{m(\tilde{\varphi}) \mid \mathscr{D}\}, \tag{17}$$

but now with posterior-to-data estimates of the mean frequency and mean severity. (There is also tacit use of the fact that past and future counts are independent, given λ).
Much is known about finding exact Bayesian and approximate credibility values for the predictive means in (17), see, e.g. *Hewitt* (1970), *Jewell* (1971)(1980), and *Bühlmann* (1974). So, in the case where we want to predict mean total claims in some future interval based on data from all claims during some past interval, the methodology is simple and straightforward.

10.2 Excess-of-Losses Model

In many forms of reinsurance treaties, on the other hand, attention focuses on the number and total amount of claims that are in *excess* of some *retention layer* value, L \$. In other words, the reinsurer *pays only for the truncated random severity:*

$$\tilde{y} = \{\tilde{x} - L \mid \tilde{x} > L\}. \tag{18}$$

In terms of the underlying "ground-up" claims just studied, this new random variable has density and conditional moments:

$$\begin{aligned} p(y \mid L, \varphi) &= \frac{p(L + y \mid \varphi)}{Q(L \mid \varphi)}; \\ m_L(\varphi) &= \frac{R(L \mid \varphi)}{Q(L \mid \varphi)}; \\ v_L(\varphi) &= 2\frac{S(L \mid \varphi)}{Q(L \mid \varphi)} - m_L^2(\varphi)\,. \end{aligned} \tag{19}$$

Also, only a fraction $Q(L \mid \varphi)$ of the original claims will exceed level L, so that $\tilde{r}$, the random *number of excess claims,* will usually be substantially less than $\tilde{n}$ and will have a counting distribution:

$$\begin{aligned} p_r(\lambda, \varphi, U, L) &= \mathscr{P}\!r\{\tilde{r} = r \mid \lambda, \varphi, U, L\} \\ &= \sum_{n=r}^{\infty} p_n(\lambda, U)\binom{n}{r}[P(L \mid \varphi)]^{n-r}[Q(L \mid \varphi)]^r, \end{aligned} \tag{20}$$

so that:

$$\mathscr{E}\{\tilde{r} \mid \lambda, \varphi, U, L\} = \mathscr{E}\{\tilde{n} \mid \lambda, U\} \cdot Q(L \mid \varphi);$$

$$\begin{aligned}\mathscr{V}\{\tilde{r} \mid \lambda, \varphi, U, L\} &= \mathscr{E}\{\tilde{n} \mid \lambda, U\} \cdot P(L \mid \varphi) \cdot Q(L \mid \varphi) \\ &\quad + \mathscr{V}\{\tilde{n} \mid \lambda, U\} \cdot [Q(L \mid \varphi)]^2.\end{aligned} \tag{21}$$

Given the Poisson assumption for the original claim count distribution, this leads not only to the obvious result that $p_r(\lambda, \varphi, U, L)$ is *Poisson* $(\lambda U \cdot Q(L \mid \varphi))$, but also to the well-known factorization result that the number of *primary* (uncovered) claims in the same interval, say, $\tilde{u} = \tilde{n} - \tilde{r}$, is *independent* of $\tilde{r}$, given the parameters, and has a *Poisson* $(\lambda U \cdot P(L \mid \varphi))$ distribution! This will greatly simplify the arguments below.

The objective of the reinsurer is to forecast the total dollar amount of covered claims in a future exposure interval of U years, that is, to predict the random compound sum, $\tilde{w}$, of the excess of all claims larger than L. Since this is the same as predicting the mean total severity using (19), we seek:

$$\mathscr{E}\{\tilde{w} \mid \mathscr{D}\} = \mathscr{E}\{\tilde{\lambda} U \cdot Q(L \mid \tilde{\varphi}) \cdot m_L(\tilde{\varphi}) \mid \mathscr{D}\} = \mathscr{E}\{\tilde{\lambda} U \cdot R(L \mid \tilde{\varphi}) \mid \mathscr{D}\}. \tag{22}$$

In other words, the objective is identical to predicting the product of the basic underlying claim frequency with the *stop-less premium*.

We now consider the use of credibility theory to approximate (22) for different types of claim data that might be available. The dilemma of the reinsurer (and the agony of his actuary) is that complete "ground-up" claims data are usually unavailable, or are too expensive to process. Naturally, some secondary claims data must be supplied before a treaty can be rated, but this data is quite sparse relative to the underlying experience as L is usually far out in the tail of any possible severity distribution. In terms of (22), this means that, while $m_L(\varphi)$ may be adequately estimated, λ and $Q(L \mid \varphi)$ are unfortunately not. A new analytic problem is that, with only secondary claim data available, the factorization of (17) no longer occurs, giving more complex prediction formulae.

To introduce more generality into the model, we shall henceforth suppose that it is possible to observe claims that exceed some *capture level,* $K\$(0 \leq K \leq L)$, during the observation interval of T years. If $K = 0$, we observe all claims of any size, or if $K = L$, we observe only those excess claims similar to those to be covered by the reinsurance treaty; a typical contractual arrangement might be to require reporting of all claims with capture level $K = 0.5L$. Since K is fixed in any case, an observed value $\tilde{y} = y$ in excess of K implies an underlying claim of $\tilde{x} = K + y$.

10.3 Counts Only Observed

First let us consider the case in which only the number of claims in excess of K during interval T are given, $\mathscr{D} = \{r\}$. It is clear that with such limited data, we cannot hope to predict the total severity, so we shall settle for a forecast of the number of claims under the treaty, that is, those in excess of L during the forecast interval U; thus the objective in this section only will be to estimate just $\mathscr{E}\{\tilde{\lambda}U \cdot Q(L \mid \tilde{\varphi})\}$.
The data likelihood contains all of the information about the parameters obtained from the observations:

$$p(r \mid \lambda, \varphi) = \frac{[\lambda T \cdot Q(K \mid \varphi)]^r e^{-\lambda T \cdot Q(K \mid \varphi)}}{r!}. \tag{23}$$

We see that the interaction between λ and φ is inseparable, given only secondary data, and that there are no unique separate maximum-likelihood estimates of the parameters, only the joint estimate, $\widehat{\lambda} \cdot Q(K \mid \widehat{\varphi}) = (\frac{r}{T})$. So, we have the paradoxical result that, if we collect data at some capture level K not equal to the retention level L, classical *MLEs* cannot provide the estimate $\widehat{\lambda}U \cdot Q(L \mid \widehat{\varphi})$ needed to rate the treaty!
Straub's model (1971) avoids this problem by assuming that the severity distribution is known, and hence that $Q_L = Q(L \mid \varphi)$ and $Q_K = Q(K \mid \varphi)$ are fixed. He also considers more than one risk to be under observation (and apparently the same number to be covered under the treaty), but this is the same as using different T and U in our model because of the *Poisson* assumption. We find for our generalization that:

$$\mathscr{E}\{\tilde{\lambda}U \cdot Q_L \mid \mathscr{D}\} \approx U\left[[1 - z(T)][\mathscr{E}\{\tilde{\lambda}\} \cdot Q_L] + z(T)\left[\frac{Q_L}{Q_K}\right]\left[\frac{r}{T}\right]\right], \tag{24}$$

where the credibility factor is:

$$z(T) = \left[\frac{T}{T + T_0}\right]; \quad T_0 = \left[\frac{\mathscr{E}\{\tilde{\lambda}\}}{\mathscr{V}\{\tilde{\lambda}\}Q_K}\right]. \tag{25}$$

Notice how the experience claims rate, r/T, during interval T is deflated using the known ratio Q_L/Q_K to obtain the frequency estimate at the treaty retention level. Otherwise, (24) is identical with *Straub's* result if $K = L$ and $T = U =$ number of risks exposed in one year. In particular, the credibility factor approaches unity with increasing T, so that the deflated experience estimate times U is ultimately "fully credible" for the mean number of excess

claims to be observed under the treaty. Of course, (24) is an exact Bayesian prediction if $\tilde{\lambda}$ has a Gamma density.
Patrik/Mashitz's contribution (1989) is to introduce the uncertainty about the severity parameter, $\tilde{\varphi}$, into the credibility estimation of the mean count. This means we must now treat $\tilde{Q}_L = Q(L \mid \tilde{\varphi})$ and $\tilde{Q}_K = Q(K \mid \tilde{\varphi})$ as (strongly dependent) random variables, with prior $p(\varphi)$. Following their development under the generalized observational protocol, we find easily the forecast:

$$\mathscr{E}\{\tilde{\lambda} U \cdot \tilde{Q}_L \mid \mathscr{D}\}$$
$$\approx U\left[[1 - z(T)]\left(\mathscr{E}\{\tilde{\lambda}\} \cdot \mathscr{E}\{\tilde{Q}_L\}\right) + z(T)\left(\frac{\mathscr{E}\{\tilde{Q}_L\}}{\mathscr{E}\{\tilde{Q}_K\}}\right)\left(\frac{r}{T}\right)\right], \tag{26}$$

where the credibility factor ist:

$$z(T) = \left(\frac{\mathscr{E}\{\tilde{Q}_K\}}{\mathscr{E}\{\tilde{Q}_L\}}\right)\left(\frac{\mathscr{V}\{\tilde{\lambda}\}\mathscr{E}\{\tilde{Q}_L\tilde{Q}_L\} + [\mathscr{E}\{\tilde{\lambda}\}]^2\mathscr{C}\{\tilde{Q}_K; \tilde{Q}_L\}}{\mathscr{V}\{\tilde{\lambda}\}\mathscr{E}\{\tilde{Q}_K^2\} + [\mathscr{E}\{\tilde{\lambda}\}]^2\mathscr{V}\{\tilde{Q}_K\}}\right)\left(\frac{T}{T + T_0}\right);$$
$$T_0 = \left(\frac{\mathscr{E}\{\tilde{\lambda}\}\mathscr{E}\{\tilde{Q}_K\}}{\mathscr{V}\{\tilde{\lambda}\}\mathscr{E}\{\tilde{Q}_K^2\} + [\mathscr{E}\{\tilde{\lambda}\}]^2\mathscr{V}\{\tilde{Q}_K\}}\right). \tag{27}$$

Patrik/Mashitz (1989) analyze the model $K = L$, in which case it can be seen that the first two terms in (27) are unity, giving an ordinary credibility factor in T with time constant T_0. Notice that we now require both the first and second moments of the frequency parameter. In their results they also assume that $p(\lambda)$ is *Gamma* (a, b), but this is not at all necessary in a linear approximation. Of course, (26) can never be an exact Bayesian prediction, whatever the priors, due to the complexity of (23).
In our generalization, the experience frequency is deflated as in (24), but now using the *a priori* average probabilities of exceedance. But perhaps the most surprising result in (27) is the fact that the credibility factor does *not* approach unity for large T when $K \neq L$, so the experience data is *not,* in the limit, "fully credible" for predicting the treaty claim frequency! In other words, a certain amount of the prior estimate, $\mathscr{E}\{\tilde{\lambda}\} \cdot \mathscr{E}\{\tilde{Q}_L\}$, will always be present in the prediction! To understand this result, we must consider what happens from a Bayesian point of view.
Using Bayes' Law in the form $p(\lambda, \varphi \mid \mathscr{D}) \propto p(\mathscr{D} \mid \lambda, \varphi) \cdot p(\lambda)p(\varphi)$, with the likelihood (23), we first find the *posterior-to-data* distribution of the parameters, from which $\mathscr{E}\{\tilde{\lambda}U \cdot \tilde{Q}_L \mid \mathscr{D}\}$ follows by integration. But we see immediately from the likelihood that $\tilde{\lambda}$ and $\tilde{\varphi}$ will now be *dependent* random variables after the observation. In fact if we assume that $\tilde{\varphi}$ represents a single

positive parameter, we see that the likelihood, viewed as a function of λ and φ in the positive quadrant, has a *ridge of constant height* along the line $\lambda \cdot Q(K \mid \varphi) = r/T$, no matter how large r and T become. In the limit, r/T converges almost surely to the true capture level claim rate, and the ridge becomes a degenerate "wall", with vanishing values elsewhere in the plane. In short, if r were larger (or smaller) than expected for some large but finite T, we could never be sure whether this means that $\tilde{\lambda}$ is larger (or smaller) than expected, whether $Q(K \mid \tilde{\varphi})$ is, or whether they both are. Now, for the case $K = L$, estimating mean future counts is done solely in the subspace along the ridge, so that the prior is, in the limit, "overcome" by the limiting likelihood. However, in the general case $K \neq L$, the forecast involves $\lambda \cdot Q(L \mid \varphi)$, whose constant values cross the ridge; thus, a better estimate can be obtained by always including some of the prior information!

10.4 K-Excess Counts and Severities Observed

Since so little information is provided by excess counts alone, let us consider the more usual case in which both the counts and claim severities are captured during the observation interval, that is, $\mathscr{D} = \{r; y\}$, where $y = [y_1, y_2, \ldots, y_r]^T$. The data likelihood is then

$$p(\mathscr{D} \mid \lambda, \varphi) = \frac{(\lambda T)^r e^{-\lambda T \cdot Q(K \mid \varphi)}}{r!} \prod_{j=1}^{r} p(K + y_j \mid \varphi). \tag{28}$$

Note that the factor $Q(K \mid \varphi)$, which was present in both terms in (23), now appears only in the exponential; however, the presence of this coupling term still means that the parameters are dependent, *a posteriori,* and makes a full Bayesian analysis difficult. A ridge is still present in the likelihood, but now it drops off from the (usually unique) maximum likelihood point because it is "shaped" in the φ direction by the last term, which contains information from the observed excess claims. If the priors are both unimodal then the joint posterior is also likely to be unimodal, giving more stable predictions.

Now let us return to the problem of predicting the mean total severity through the approximation of $\mathscr{E}\{\tilde{w} \mid r, y\} = \mathscr{E}\{\tilde{\lambda} U \cdot \tilde{R}_L \mid r, y\}$, where we set $\tilde{R}_L = R(L \mid \tilde{\varphi})$ for convenience. In many severity models of general interest, the sum of the observables, $Y = \sum y_j$, is a sufficient statistic for non-truncated variables and is, of course, also sufficient for truncated exponential variables; the statistic Y, appropriately deflated, would also be considered by many as the natural experience predictor for total severity. Thus it seems reasonable to approximate instead the simpler mean $\mathscr{E}\{\tilde{\lambda}U \cdot \tilde{R}_L \mid r, Y\}$. But what shall we do about r? One possibility would be to use a two-dimensional formulation to approximate the predictive mean as a linear function of both r and Y. But in ordinary credibility the number of samples enters in a distinctly nonlinear fashion, and this will certainly be true here. Fortunately, *Jewell* (1975) reminds us that we can also use *conditional* credibility theory, that is, we need only linearize on as many statistics as are necessary, retaining the others as conditioning parameters.

In our model this means that we will linearize on Y, but leave r as a conditioning variable in the moments, for which we will need $p(\lambda, \varphi \mid r) = p(\lambda \mid r, \varphi) \cdot p(\varphi \mid r)$, which is found using (23):

$$p(\lambda \mid r, \varphi) \propto [\lambda T]^r e^{-\lambda T \cdot Q(K \mid \varphi)} p(\lambda),$$
$$p(\varphi \mid r) \propto [Q(K \mid \varphi)]^r p(\varphi). \tag{29}$$

(Note that dependency upon K and T is suppressed). It turns out that we only need the mean conditional frequency rate from the first density, call it $\Lambda(r, \varphi) = \mathscr{E}\{\tilde{\lambda} \mid r, \varphi\}$.

The next step involves unconditioning arguments, whose details will be presented elsewhere. The final forecast turns out to be:

$$\mathscr{E}\{\tilde{w} \mid r, Y\} \approx U\left\{\mathscr{E}\{\Lambda(r, \tilde{\varphi})\tilde{R}_L \mid r\} + z(r)\left[\left(\frac{Y}{r}\right) - \mathscr{E}\{m_K(\tilde{\varphi}) \mid r\}\right]\right\}, \tag{30}$$

with a credibility factor:

$$z(r) = \left(\frac{\mathscr{C}\{\Lambda(r, \tilde{\varphi})\tilde{R}_L; m_K(\tilde{\varphi}) \mid r\}}{\mathscr{V}\{m_K(\tilde{\varphi}) \mid r\}}\right)\left(\frac{r}{r + r_0(r)}\right);$$
$$r_0(r) = \left(\frac{\mathscr{E}\{v_K(\tilde{\varphi}) \mid r\}}{\mathscr{V}\{m_K(\tilde{\varphi}) \mid r\}}\right). \tag{31}$$

This result should be compared with the count predictions (24) and (26). Note that the dimensions are: $[Y] = [R] = [m_K] = \$$, $[v_K] = \$^2$, $[\Lambda] = \text{year}^{-1}$, so that r_0 is dimensionless, $[z] = \text{year}^{-1}$, and the expression inside braces in (30) predicts the *covered claim loss rate* (\$/year).

If we make the usual assumption that $p(\lambda)$ is *Gamma* (a, b), then the updated density will be *Gamma* $(a + r, b + T \cdot Q(K \mid \varphi))$, so that:

$$\Lambda(r, \varphi) = \mathscr{E}\{\tilde{\lambda} \mid r, \varphi\} = \left(\frac{a + r}{b + T \cdot Q(K \mid \varphi)}\right). \tag{32}$$

With a general prior, one could find a credibility approximation to Λ that was linear in r; the main point is that $Q(K \mid \varphi)$ is going to enter into the forecast in a way that complicates an exact analysis.
Note that all moments needed for the forecast will use $p(\varphi \mid r)$, thus introducing a complicated dependency upon r, even in the time "constant", $r_0(r)$! Computing these factors will also be a little difficult, since powers of $Q(K \mid \widetilde{\varphi})$ are introduced into each expectation, including the normalizing factor. Thus, to find the mean severity of an observed excess claim, we must calculate:

$$\mathscr{E}\{ m_K(\widetilde{\varphi}) \mid r \} = \frac{\mathscr{E}\{ R(K \mid \widetilde{\varphi}) \cdot [Q(K \mid \widetilde{\varphi})]^{r-1} \}}{\mathscr{E}\{ [Q(K \mid \widetilde{\varphi})]^r \}},$$

where the $\mathscr{E}$ operator on the RHS means unconditional expectation using only $p(\varphi)$; similar results hold for the other terms. But this program is not too difficult for severity distributions of interest in reinsurance; details will be given elsewhere.
Thus an increase in r affects the forecast in *three* different ways: the accuracy of $\Lambda(r, \varphi)$ improves; more weight is attached to the statistic Y/r; and the moments, computed with $p(\varphi \mid r)$, change. The first change, if $T \gg \frac{b}{Q(K \mid \varphi)}$ and (32) applies, brings a term r/T outside the terms in (30) and (31), and leaves $(T\widetilde{Q}_K)^{-1}$ behind. The second change, if $r \gg r_0(r)$ (r_0 is usually a weak function of r), increases the credibility factor to its largest value, the first term in brackets in (31). Unless $K = L$, this term can never be unity, so "full credibility" is never attained with this model either. Finally, the third effect of increasing r is to attach more weight to smaller values of φ, in the usual case that $Q(K \mid \varphi)$ is decreasing in the parameter(s). If φ is a single scale parameter, the estimate shifts towards larger possible severities.

10.5 A New Recipe for the Mincing Machine

Jan Jung (1964) made a noteworthy statement about the rôle of information in insurance ratemaking:

> There is a natural law which states that you can never get more out of a mincing machine than what you have put into it. That is: If the reinsurance people want actuarially sound premiums, they must get a decent information about claims distributions.

We have seen that is also true for Bayesian predictions and their credibility approximations; although our prior opinion will always give us some kind of a result, we cannot hope to move from the collective means towards the true individual means unless more useful data is provided. The question then is: how can we improve excess-of-loss predictions? We could, of course, lower the data capture level K and require more claims data from the primary carrier; a study of prediction improvement along these lines would be most interesting. But, as we have already remarked, the gathering and analysis of more claims data is an increasingly difficult and costly undertaking.

A new recipe for the mincing machine that I believe hold promise is to collect also *the total number of all claims, n,* during the interval T. As before, the correct way to evaluate this proposal is through the Bayesian approach. With $\mathscr{D} = \{n, r; y\}$, the likelihood is now:

$$p(\mathscr{D} \mid \lambda, \varphi) = \frac{(\lambda T)^n e^{-\lambda T}}{n!} [P(K \mid \varphi)]^{n-r} \prod_{j=1}^{r} p(K + y_j \mid \varphi). \tag{33}$$

We see immediately that the information provided about the two parameters separates, so that, with independent priors, *the prediction of total severity factors* as it did for ground-up claims in (17), giving the *exact* result:

$$\mathscr{E}\{\tilde{\lambda} U \cdot \tilde{R}_L \mid n, r, y\} = \mathscr{E}\{\tilde{\lambda} U \mid n\} \cdot \mathscr{E}\{\tilde{R}_L \mid n, r, y\}. \tag{34}$$

This is already a great improvement, as the first term is easily calculated or approximated in the usual way, and we can concentrate on the prediction of the mean stop-loss premium, which we know will be a very small quantity.

If we take the conditional credibility approach to estimating $\mathscr{E}\{\tilde{R}_L \mid n, r, Y\}$ as a linear function of Y, we must first find the conditional density of $\tilde{\varphi}$, which is:

$$p(\varphi \mid n, r) \propto [P(K \mid \varphi)]^{n-r} [Q(K \mid \varphi)]^r p(\varphi). \tag{35}$$

This is the best we can hope for with n and r given, since the likelihood part already has a unique *MLE*, $Q(K \mid \tilde{\varphi}) = r/n$, which becomes sharper with increased r and n, that is, with increased T or decreased K.

We omit the details of calculation, and give the final credibility formula:

$$\begin{aligned} &\mathscr{E}\{\tilde{R}_L \mid n, r, Y\} \\ &\quad \approx U\left\{\mathscr{E}\{\tilde{R}_L \mid n, r\} + z(n, r)\left[\left(\frac{Y}{r}\right) - \mathscr{E}\{m_K(\tilde{\varphi}) \mid n, r\}\right]\right\}, \end{aligned} \tag{36}$$

with a credibility factor:

$$z(n,r) = \left[\frac{\mathscr{C}\{\widetilde{R}_L; m_K(\widetilde{\varphi}) \mid n,r\}}{\mathscr{V}\{m_K(\widetilde{\varphi}) \mid n,r\}}\right]\left(\frac{r}{r + r_0(n,r)}\right);$$

$$r_0(n,r) = \left(\frac{\mathscr{E}\{v_K(\widetilde{\varphi}) \mid n,r\}}{\mathscr{V}\{m_K(\widetilde{\varphi}) \mid n,r\}}\right). \tag{37}$$

Many of the remarks made in the last section for (31) apply here also, such as the implicit dependency upon n and r from (35), and there are many interesting computational and practical questions still unanswered about these models. However, my goal in this development of the excess-of-losses model was to convince you that actuarial mathematics has a great deal to learn from the Bayesian approach, even when dealing with simple approximations like credibility theory.

11 Conclusion

I hope that my remarks today have shown how Hans Bühlmann's early insights led to far-reaching advances in actuarial science and, in particular, how his work gave me a "jump start" in my own research. I am sure that all of you can add your own examples of ways in which his wisdom, help, and guidance have greatly influenced your own careers. In concluding my remarks in honor of our remarkable colleague, I can only think to add the wonderful paraphrase of Lucan made by Sir Isaac Newton:

> If I have seen further it is by standing on the shoulders of Giants.

Thank your for your attention.

William S. Jewell
Engineering Systems Research Center
University of California, Berkeley
Berkeley, CA 94720
USA

References

Bailey, A.L. (1945): A generalized view of credibility. Proceedings of the Casualty Actuarial Society, *32*, 13–20.

Bailey, A.L. (1950): Credibility procedures. LaPlace's generalization of Bayes' rule and the combination of collateral data with observed data. Proceedings of the Casualty Actuarial Society *37*, 7–23. Discussion by R. von Mises, E.C. Molina, J.E. Freund, M.V. Johns, Jr., W. Perks, L.H. Longley-Cook, *37*, 94–115.

Barnett, V. (1982): Comparative Statistical Inference, second edition. John Wiley & Sons, New York.

Bichsel, F. (1964): Erfahrungstarifierung in der Motor-Fahrzeughaftplficht-Versicherung. Bulletin of the Association of Swiss Actuaries *64*, 119–130.

Bichsel, F. (1967): Experience rating in subsets of risks. ASTIN Bulletin *4*, 210–227.

Bühlmann, H. (1960): Austauschbare stochastische Variablen und ihre Grenzwertsätze, Promotionsarbeit, Eidgenössische Technische Hochschule in Zürich. Also, University of California Publications in Statistics *3*, nr. 1, 1–35. University of California Press, Berkeley.

Bühlmann, H. (1964a): A distribution free method for general risk problems. ASTIN Bulletin *3*, 142–152.

Bühlmann, H. (1964b): Optimale Prämienstufensysteme. Bulletin of the Association of Swiss Acutaries *64*, 193–214.

Bühlmann, H. (1967): Experience rating and credibility. ASTIN Bulletin *4*, 199–207.

Bühlmann, H. (1969): Experience rating and credibility. ASTIN Bulletin *5*, 157–165.

Bühlmann, H. (1970): Credibility procedures. In: Preceedings 6th Berkeley Symposium on Math. Statistics and Probability, University of California, 1970, Volume 1, 1972, 515–525. University of California Press, Berkeley.

Bühlmann, H. (1970): Mathematical Methods of Risk Theory. Springer-Verlag, New York.

Bühlmann, H. (1974): A comparison of three credibility formulae using multidimensional techniques. ASTIN Bulletin *7*, 203–207.

Bühlmann, H. (1975): The mincing machine revisited. ASTIN Bulletin *8*, 135–143.

Bühlmann, H. (1989): Tendencies of development in risk theory. Paper presented at the Centennial Meeting of the Casualty Actuarial Society, Washington, D.C., June, 1989.

Bühlmann, H./Jewell, W.S. (1987): Hierarchical credibility revisited. Bulletin of the Association of Swiss Actuaries *87*, 35–54.

Bühlmann, H./Straub, E. (1970): Glaubwürdigkeit für Schadensätze. Buzlletin of the Association of Swiss Actuaries *70*, 111–133. Translated as "Credibility for Loss Ratios" by C.E. Brooks, with Appendix added. ARCH *1970*, nr. 2.

De Vylder, F. (1978): Parameter estimation in credibility theory. ASTIN Bulletin *10*, 99–112.

Dropkin, L. (1959): Some considerations on automobile rating systems utilizing individual drive records. Proceedings of the Casualty Actuarial Society, *46*, 165–177.

Dropkin, L. (1960): Automobile merit rating and inverse probabilites. Proceedings of the Casualty Actuarial Society *47*, 37–40.

Finetti, B. de (1964): Sulla teoria della credibilità. Giornale dell'Istituto Italiano degli Attuari *27*, 219–231.

Finetti, B. de (1964): Report on Subject I of the ASTIN Colloquium, Trieste, 1963. La théorie des plus grandes valeurs et son application aux problèmes de l'assurance. ASTIN Bulletin *3*, 116–131.

Fürst, Dario (1964): Formulation Bayesienne du problème des valeurs extrêmes en relation à la réassurance en "excédent de sinistres". ASTIN Bulletin *3*, 154–162.

Hewitt, C.C., Jr. (1970): Credibility for severity. Proceedings of the Casualty Actuarial Society *57*, 148–171. Discussion by H.U. Gerber and H. Bühlmann *58*, 25–31.

Hickman, J.C. (1975): Introduction and historical overview of credibility. In: Credibility – Theory and Applications, edited by P.M. Kahn, 181–192. Academic Press, New York.

Howson, C./Urbach, P. (1989): Scientific Reasoning: The Bayesian Approach. Open Court Publishing Company, La Salle, Illinois.

Jewell, W.S. (1973): Multi-dimensional credibility. Operations Research Center Report ORC 73–7, University of California, Berkeley. Also in ARCH *1973*, nr. 4.

Jewell, W.S. (1974a): Credible means are exact Bayesian for simple exponential families. ASTIN Bulletin *8*, 77–90.

Jewell, W.S. (1974b): Exact multidimensional credibility. Bulletin of the Association of Swiss Actuaries *74*, 193–214.

Jewell, W.S. (1975a): Regularity conditions for exact credibility. ASTIN Bulletin *8*, 336–341.

Jewell, W.S. (1975b): The use of collateral data in credibility theory: a hierarchical model. Giornale dell'Istituto Italiano degli Attuari *38*, 1–16.

Jewell, W.S. (1975c): Model variations in credibility theory, with discussion. In: Credibility – Theory and Applications, edited by P.M. Kahn, 193–247. Academic Press, New York.

Jewell, W.S. (1980): Models in insurance: paradigms, puzzles, communications and revolutions. In: Transactions of the 21st International Congress of Actuaries, Zurich and Lausanne, 1980, Volume S, 87–141.

Jewell, W.S. (1983): Enriched multinormal priors revisited. Journal of Econometrics *23*, 5-35.

Jewell, W.S. (1987): A heteroscedastic hierarchical model. In: Bayesian Statistics 3, Proceedings of the Third Valencia International Meeting, 1–5 June 1987, 657–663. Clarendon Press, Oxford.

Jewell, W.S. (1989a): A general framework for credibility prediction of multidimensional first and second moments. Insurance: Mathematics and Economics *8*, 127–136.

Jewell, W.S. (1989b): Credibility approximations for second moments in hierarchical models. Bulletin of the Association of Swiss Actuaries, Special jubilee issue in honor of Hans Bühlmann, *89*, nr. 2, 301–330.

Jewell, W.S./Schnieper, R. (1985): Credibility approximations for Bayesian prediction and second moments. ASTIN Bulletin *15*, 103–121.

Jung, J. (1964): On the use of extreme values to estimate the premium for an excess of loss reinsurance. ASTIN Bulletin *3*, 178–184.

Lindley, D.V. (1970): The Bayesian approach. Scandinavian Journal of Statistics *5*, 1–26.

Lindley, D.V./Smith, A.F.W. (1972): Bayes estimates for the linear model. Journal of the Royal Statistical Society B *34*, 1–41.

Loimaranta, K. (1977): On the calculation of variances and credibilities by experience rating. ASTIN Bulletin, *9*, 203.

Mayerson, A.L. (1964): A Bayesian view of credibility. Proceedings of the Casualty Actuarial Society *51*, 85–104. Discussion by C.C. Hewitt, Jr., *52*, 121–127.

Mosteller, F./Tukey, J.W. (1968): Data analysis, including statistics, Chapter 10. In: The Handbook of Social Psychology, Volume Two (Second Edition), edited by G. Lindzey and E. Aronson, 80–203. Addison-Wesley Publishing Company, Reading.

Mosteller, F./Tukey, J.W. (1977): Data Analysis and Regression. Addison-Wesley Publishing Company, Reading.

Norberg, R. (1979): The credibility approach to experience rating. Scandinavian Actuarial Journal *1979*, 181–221.

Norberg, R. (1980): Empirical Bayes credibility. Scandinavian Actuarial Journal *1980*, 177–194.

Norberg, R. (1981): On optimal parameter estimation in credibility. XVth ASTIN Colloquium, Loen, 1981.

Norberg, R. (1986): Hierarchical credibility: analysis of a random effect linear model with nested classification. Scandinavian Actuarial Journal *1986*, 204–222.

Patrik, G./Mashitz, I. (1989): Credibility for reinsurance excess pricing. Paper presented at the 21st ASTIN Colloquium, New York, November, 1989.

Seal, H.L. (1969): Stochastic Theory of a Risk Business. John Wiley & Sons, New York.

Sundt, B. (1979): A hierarchical credibility regression model. Scandinavian Actuarial Journal *1979*, 107–114.

Sundt, B. (1979): On choice of statistics in credibility estimation. Scandinavian Actuarial Journal *1979*, 115–123.

Straub, E. (1971): Estimation of the number of excess claims by means of credibility theory. ASTIN Bulletin *5*, 388-392.

Straub, E. (1975): Credibility in practice, with discussion. In: Credibility – Theory and Applications, edited by P.M. Kahn, 347–362. Academic Press, New York.

Taylor, G.C. (1984): Experience rating with credibility adjustment of the manual premium. ASTIN Bulletin *7*, 323–336.

Taylor, G.C. (1979): Credibility analysis of general hierarchical models. Scandinavian Actuarial Journal *1979*, 1–12.

Wit, G.W. de, ed. (1986): Special issue on credibility theory. Insurance: Abstracts and Reviews *2*, No. 3, North-Holland, Amsterdam.

Summary

A brief survey of statistical thought on estimation and prediction is given, with emphasis on the historical development of credibility theory and Hans Bühlmann's important rôle in the evolution of this area. The analogy of the misty staircase is from *Mosteller* and *Tukey*, who use it to describe the approach and limitations of the traditional statistical approach; the author continues by describing first the credibility staircase, and then the most recent contender, the Bayesian escalator. The importance of the Bayesian paradigm and its interaction with credibility theory is illustrated using an important problem in reinsurance – the forecasting of excess-of-loss claims.

Zusammenfassung

Im vorliegenden Festreferat wird ein kurzer Abriss der statistischen Denkprinzipien bei Schätz- und Vorhersageproblemen gegeben, unter besonderer Beachtung der historischen Entwicklung der Kredibilitätstheorie und der wichtigen Rolle, die Hans Bühlmann auf diesem Gebiet gespielt hat. Das Bild einer «nebelhaften Treppe» wurde von *Mosteller* und *Tukey* verwendet, um den Zugang zur traditionellen Statistik und ihre Schranken zu illustrieren. Der Autor führt dieses Bild fort und beschreibt die Kredibilitäts-Treppe und – als jüngsten Mitbewerber – die «Bayesianische Rolltreppe». Die Wichtigkeit der Bayesianischen Hypothese und ihre Wechselwirkung zur Kredibilitätstheorie wird an einem wichtigen Beispiel aus der Rückversicherung erläutert: der Vorhersage von Excess-off-loss-Schäden.

Résumé

L'auteur présente un rapide survol des principes statistiques sur lesquels reposent les opérations d'estimation et de prévision, en appuyant spécialement sur le développement historique de la théorie de la crédibilité et le rôle important joué par Hans Bühlmann dans l'évolution de ce domaine. L'image de l'"escalier dans les brumes" a été proposée par *Mosteller* et *Tukey* pour illustrer l'approche statistique traditionnelle et ses limitations. L'auteur développe cette image en décrivant tout d'abord l'escalier de la crédibilité puis son plus récent assaillant: l'escalier roulant bayesien. L'importance de l'hypothèse bayesienne et son interaction avec la théorie de la crédibilité sont illustrées au moyen d'une question de poids en réassurance: la prévision des sinistres en "excess-of-loss".