

Sequential Testing for Full Credibility

by Michael Baron and Rui Xu

ABSTRACT

This article introduces sequential statistical methods in actuarial science. As an example of sequential decision making that is based on the data accrued in real time, it focuses on sequential testing for full credibility. Classical statistical tools are used to determine the stopping time and the terminal decision that controls the overall error rate and power of the procedure. As a result, a set of conditions is obtained under which an insured cohort becomes fully credible.

KEYWORDS

Credibility, frequency-severity model, likelihood ratio, sequential test, stopping boundary, asymptotic normality

1 Introduction

Sequential statistical methods refer to data that are collected in real time. The sample size is not fixed in advance. At any moment, either a decision is taken, or it is deferred until more data become available. For example, when a statistical hypothesis is tested sequentially, after each new observation, either the null hypothesis is rejected, or it is accepted, or no decision is made at this point, and sampling continues (e.g. [3, 9, 12]).

Actuaries routinely make decisions that are *sequential* in nature (e.g., [10]). During each insured period, the new claims and losses data are collected, and together with the new economic and financial situation and other factors, they are taken into account for the calculation of revised premiums and risks. For all the policyholders, several decisions are made regularly, at least once per each insured period: whether the coverage should continue, whether the coverage should be changed, and what premium should be charged.

Conventional methods tend to ignore the facts of sequentially developing data and recurrence of actuarial decisions. They use non-sequential statistical procedures that are based on the assumption of only one statistical inference that will not be repeated. This leads to lower than nominal confidence levels and higher than nominal probabilities of Type I and Type II errors. That is, the probability of committing an error at least once during repeated tests is substantially higher than the nominal level of each test. Controlling these errors at the specified levels is one of the benefits of the proposed methodology.

One important sequential problem is testing an insured cohort for *full credibility*. Credibility estimation combines the real data R and the hypothetical prior information H and uses a *compromise estimator*

$$C = ZR + (1 - Z)H, \tag{1.1}$$

where Z is the *credibility factor*. Assigning *full credibility*, $Z = 1$, means that the real data about a cohort are deemed sufficient for decision making, without referring to other sources of information. When the real data are not sufficient, *partial credibility* is used according to (1.1). Compromise estimators of type (1.1) are widely used for the estimation of *pure premium*, a part of the total premium, the expected amount that the insurer will have to pay to cover the policyholder's losses during the next insured period. The resulting estimator is also called *credibility premium*. Let us further refer to [1, 2, 4] for foundations of credibility; [6, 11] for the current status of credibility theory and practice.

When does a cohort deserve full credibility? This can be determined, among other methods, by the classical *limited fluctuation credibility approach*. According to it, the cohort is fully credible if its total loss S is likely to fall within a desired margin from its expected value $\mathbf{E}(S)$, the pure premium, namely,

$$\mathbf{P}\{|S - \mathbf{E}(S)| \leq c \mathbf{E}(S)\} \geq 1 - q. \quad (1.2)$$

Under the *aggregate loss frequency-severity model* [6], widely used in actuarial science, condition (1.2) for full credibility can take a simple form of hypothesis testing. The model assumes X losses during the insured period, and thus, the total loss amount S is the sum of a random number X of losses with severities Y_1, \dots, Y_X ,

$$S = \sum_{i=1}^X Y_i.$$

Assumptions. We assume a Poisson number of losses X with the frequency parameter $\lambda = \mathbf{E}(X)$. For the Poisson distribution, λ also equals $\text{Var}(X)$. The loss amounts are assumed independent of each other and independent of the number of losses, with expected value and variance $\mathbf{E}(Y_i) = \mu$ and $\text{Var}(Y_i) = \sigma^2$. Under this model, $\mathbf{E}(S) = \lambda\mu$, $\text{Var}(S) = \lambda(\sigma^2 + \mu^2)$, and the standardized total loss $(S - \mathbf{E}(S))/\text{Std}(S)$ has approximately standard

normal distribution, provided that the number of losses is sufficient for normal approximation ([4, 6], also see [5, 8]).

In view of this standardization, condition (1.2) becomes equivalent to inequality

$$\frac{c\sqrt{\lambda}}{\sqrt{1+\gamma^2}} \geq z_{q/2}, \quad (1.3)$$

where $\gamma = \sigma/\mu$ is the *coefficient of variation* of individual losses (severity parameter), and $z_{q/2}$ is the $(1 - q/2)$ th standard normal quantile.

Condition (1.3) is a statistical hypothesis that is being tested every insured period in order to decide on the full credibility. Conducting this test each time at a nominal level of 5% does not guarantee that the probability of at least one wrong decision is within 5%. For example, there is a probability of $1 - (1 - 0.05)^{10} \approx 40\%$ that a Type I error is made at least once during the course of 10 independent periods, when the probability of Type I error during each period is only 5%.

For this reason, we develop a *sequential method* for testing for full credibility. It specifies a *stopping time*, when full credibility can be awarded to a cohort, simultaneously controlling for the probability of Type I error (assigning full credibility to a cohort that does not deserve it) and the probability of Type II error (assigning only a partial credibility whereas the full credibility should be given). As a practical result, precise criteria are derived under which an insured cohort becomes fully credible.

For some cohorts, full credibility can be granted at an early stage, for others at a later stage, but regardless of the time, the proposed sequential tests control the probability of Type I error at a specified level α and the probability of Type II error at a specified level β . In other words, the probability that a cohort with insufficient credentials is (by a sampling error) considered fully credible is controlled at the desired level throughout the entire time

of data collection and decision making.

Sequential test for full credibility is developed in Section 2 as an application of classical sequential methods to credibility theory. Examples of using the new tool under the popular loss models are shown in Section 3. Proposed methods are then applied to the data generated by the Actuarial Loss Simulator in Section 4. Section 5 contains summary and conclusions, and all the lengthy proofs are found in Section 6.

2 Derivation of the Sequential Test

In this section, a level α sequential test for full credibility is derived. As shown in Section 1, the limited fluctuation condition for full credibility is equivalent to inequality (1.3), and thus, deciding between the full and partial credibility, an actuary needs to conduct a test of

$$H_0 : c\eta \leq z_{q/2} - \delta \quad \text{vs} \quad H_A : c\eta \geq z_{q/2}, \quad (2.1)$$

where $\eta = \frac{\sqrt{\lambda}}{\sqrt{1+\gamma^2}}$ is the tested parameter of interest, and a *region of indifference* of length δ separates the null and alternative hypotheses. An insured cohort is considered *fully credible* when there is sufficient evidence in the data that condition (1.3), equivalent to H_A , is satisfied.

Sequential test of (2.1) is derived through a series of steps. First, parameter η is estimated by a statistic T_n that is based on the first n time periods of observation. This statistic is shown to have an asymptotically Normal distribution for any n . For most of the loss models, the distribution of T_n contains unknown nuisance parameters which affect its variance only. In this case, the problem reduces to testing the location parameter of a Normal distribution with unknown variance, and the *sequential t-test* ([3], sec. 3.2; [9], sec. 3.6.2) is applied.

2.1 Test statistic and its distribution

We assume that the claim and loss data are collected for a large insured cohort over a sequence of days (weeks, months). During day k , a random number of claims X_k is observed, with random loss amounts Y_{k1}, \dots, Y_{kX_k} . The typical actuarial *frequency-severity model* assumes that the claims occur at random times, according to a Poisson process with intensity $\lambda = \mathbf{E}(X_k)$ claims per day, whereas the losses Y_{kj} are independent variables that follow distribution $F_Y(y)$ with expected value $\mu = \mathbf{E}(Y_{kj})$, variance $\sigma^2 = \text{Var}(Y_{kj})$, and coefficient of variation $\gamma = \sigma/\mu$, independently of claim times.

Based on the observed claim frequencies X_1, \dots, X_n and severities $Y_{11}, \dots, Y_{1X_1}; \dots; Y_{n1}, \dots, Y_{nX_n}$ over n days, parameter η is estimated by

$$T_n = \frac{\sqrt{\hat{\lambda}_n}}{\sqrt{1 + \hat{\gamma}_n^2}} = \frac{\sqrt{\bar{X}_n}}{\sqrt{1 + s_n^2/\bar{Y}_n^2}}. \quad (2.2)$$

Here parameters λ , μ , and σ are estimated at time n by the corresponding sample statistics available by that time, namely, \bar{X}_n , the average frequency or the average number of claims per time period; \bar{Y}_n and s_n , the sample mean and sample standard deviation of severity.

This yields a sequence of statistics T_1, T_2, \dots . Besides their main purpose of testing (2.1), they can also be used for estimating the partial credibility factor $Z = c\eta/z_{q/2}$ by $\hat{Z}_n = cT_n/z_{q/2}$. This factor is applied to the compromise estimation of the pure premium when H_0 is not rejected, and only partial credibility is assigned.

Next, we show that statistic T_n has asymptotically normal distribution under mild assumptions on the frequency and severity distributions.

Theorem 1. *Suppose that the frequency variables X_k and the severity variables Y_{kj} are not degenerate for $k = 1, \dots, n$ and $j = 1, \dots, X_k$; Y_{kj} is not a multiple of a Bernoulli random variable; $\mathbf{E}(X_k^2) < \infty$; and $\mathbf{E}(Y_{kj}^4) < \infty$.*

Then, as $n \rightarrow \infty$, statistic T_n is asymptotically normal, and

$$\sqrt{n}(T_n - \eta) \xrightarrow{d} \text{Normal}(0, V_T), \quad \text{as } n \rightarrow \infty,$$

where

$$V_T = \frac{\mu^2}{4\mu_2} + \frac{\mu_2^2 - \mu\mu_3}{\mu_2^2} + \frac{\mu^2(\mu_4 - \mu_2^2)}{4\mu_2^3}, \quad (2.3)$$

and $\mu_m = \mathbf{E}(Y_{kj}^m)$ are non-central moments of the severity distribution for $m = 2, 3, 4$.

The proof of this result is deferred to Appendix.

Two important conclusions follow from the form of T_n and its asymptotic variance.

Corollary 1. *The distribution of T_n is free of a scale parameter of the loss distribution.*

One can see that by noticing that each fraction in (2.3) has its numerator and denominator of the same dimension. This is expected because parameter η , estimated by T_n , depends on the loss distribution only through its coefficient of variation γ . In particular, Corollary 1 implies that our decision on full credibility is independent of the currency in which the losses Y_{kj} are valued.

The second conclusion is important for the practical implementation of the proposed sequential scheme. In typical actuarial data, a large portion of loss amounts are entered as zeros (Sect. 8.6 of [6]). In particular, this phenomenon is observed in the data obtained from the Actuarial Loss Simulator in Section 4. Zero payments are often caused by the claims exceeding policy limits or not meeting deductibles, and also by wrong coding or repeatedly submitted claims.

Allowing for a considerable probability of a zero loss clearly affects the modeling of both severities and frequencies. None of the common actuarial loss models (Gamma, Pareto, Loggamma, Lognormal, Burr, etc.) accounts for such a large number of zeros. Fortunately, the introduced test statistic T_n is independent of the inclusion of zero losses. According to

the next corollary, these losses can simply be deleted from the data when the test for full credibility is conducted.

Corollary 2. *Let \tilde{X}_k be the actual number of claims during the k -th time period and X_k be the number of claims excluding zero losses ($Y_{kj} = 0$) for all k . Consider statistics $\tilde{T}_n = \sqrt{\tilde{X}_n/(1 + \tilde{\gamma}_n^2)}$ and $T_n = \sqrt{X_n/(1 + \hat{\gamma}_n^2)}$, where $\tilde{\gamma}_n$ and $\hat{\gamma}_n$ are the sample coefficients of variation based on all the observed severities and based on the no-zero severities only. Then $\mathbf{P}\{\tilde{T}_n = T_n\} = 1$ for all n .*

This result follows immediately from an alternative expression for statistic T_n given below.

Lemma 1. *Statistic $T_n = \sqrt{X_n/(1 + \hat{\gamma}_n^2)}$ can also be expressed as*

$$T_n = \frac{1}{n} \sum_{k=1}^n \sum_{j=1}^{X_k} Y_{kj} \bigg/ \sqrt{\frac{1}{n} \sum_{k=1}^n \sum_{j=1}^{X_k} Y_{kj}^2}, \quad (2.4)$$

the average loss per day normalized by the root of the total of squared losses per day.

The proof is given in Appendix. Clearly, all $Y_{kj} = 0$ can be simply dropped in (2.4) without affecting any sums, and Corollary 2 follows. Lemma 1 is also used in the proof of Theorem 1.

2.2 Sequential scheme

Based on the asymptotically normal distribution of test statistic T_n , a sequential test for full credibility is developed in this section. The proposed procedure follows the ideas of Wald's sequential probability ratio test, or SPRT ([3, 9, 12]). This test is based on the log-likelihood ratio $\Lambda_n = \ln f(\mathbf{x}_n|H_A)/f(\mathbf{x}_n|H_0)$, where \mathbf{x}_n represents a vector of all the data collected by the time n . To attain the Type I error probability α and the Type II error probability β , the stopping boundaries $a = \ln((1 - \beta)/\alpha)$ and $b = \ln(\beta/(1 - \alpha))$ are then introduced, with

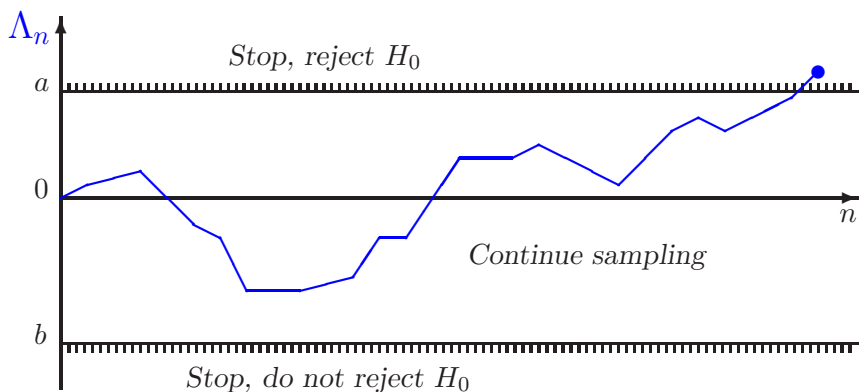


Figure 1: Example of an SPRT. Stop when Λ_n exceeds one of the boundaries.

the following sequential decision making strategy, shown in Fig. 1. Sampling continues as long as $\Lambda_n \in (b, a)$. When $\Lambda_n \geq a$, sampling stops, and H_0 is rejected in favor of H_A . When $\Lambda_n \leq b$, sampling stops, and H_0 is not rejected. For a test of simple hypotheses, when Λ_n does not contain unknown parameters, this SPRT guarantees the Type I and Type II error probabilities of α and β .

Suppose for a moment that the distribution of T_n is free of nuisance parameters. It happens, for example, when the loss model is parameterized only by a scale parameter, according to Corollary 1. In this case, the SPRT statistic for testing (2.1) is

$$\Lambda_n = \Lambda(n, \delta, t) = \ln \frac{f_{T_n}(t|H_A)}{f_{T_n}(t|H_0)} = \frac{n}{2V_T} \left\{ \left(t - \frac{z_{q/2} - \delta}{c} \right)^2 - \left(t - \frac{z_{q/2}}{c} \right)^2 \right\}, \quad (2.5)$$

where t is the value of T_n observed at time n .

Now turn to the general case of an arbitrary loss distribution. As seen in Theorem 1, additional parameters can affect the asymptotic variance but not the mean of T_n . To handle nuisance parameters, [12] proposed the method of *weight functions* $w(\theta)$, integrating both components of the likelihood ratio with respect to the unknown parameter θ ,

$$\Lambda_n = \ln \frac{\int f(\mathbf{x}_n|\theta, H_A)w(\theta)d\theta}{\int f(\mathbf{x}_n|\theta, H_0)w(\theta)d\theta}.$$

Then, the same stopping boundaries a and b (Fig. 1) control the *integrated* Type I and Type

II error probabilities, namely,

$$\int \mathbf{P} \{ \text{reject } H_0 \mid \theta, H_0 \} w(\theta) d\theta = \alpha, \quad \int \mathbf{P} \{ \text{do not reject } H_0 \mid \theta, H_A \} w(\theta) d\theta = \beta. \quad (2.6)$$

Applying this approach to our case of a normally distributed statistic with an unknown variance, a *sequential t-test* is detailed in [3], Sect. 3.2. This method essentially follows Wald's SPRT with a special choice of a constant weight function $w(\theta) = 1$ on $\theta \in (0, h)$. Letting $h \rightarrow \infty$ yields the test statistic

$$\Lambda_n = \ln \lim_{h \rightarrow \infty} \frac{\int_0^h f_{T_n}(t \mid \theta, H_A) d\theta}{\int_0^h f_{T_n}(t \mid \theta, H_0) d\theta} = \ln \lim_{h \rightarrow \infty} \frac{\int_0^h \frac{1}{\sqrt{V_T(\theta)}} \exp \left(-\frac{(t - \frac{z_{q/2}}{c})^2}{2V_T(\theta)/n} \right) d\theta}{\int_0^h \frac{1}{\sqrt{V_T(\theta)}} \exp \left(-\frac{(t - \frac{z_{q/2} - \delta}{c})^2}{2V_T(\theta)/n} \right) d\theta}. \quad (2.7)$$

As we shall see in Section 3, the limit in (2.7) can often be taken by the L'Hôpital's Rule, erasing both integrals. Multiple integrals appear in the case of a multi-dimensional nuisance parameter θ .

Based on Λ_n , one makes a decision according to the rule,

$$\left\{ \begin{array}{lll} \text{Reject } H_0 \text{ in favor of } H_A & \text{and} & \text{assign full credibility} & \text{if } \Lambda_n \geq a = \ln \frac{1-\beta}{\alpha}, \\ \text{Accept } H_0 & & \text{and assign partial credibility} & \text{if } \Lambda_n \leq b = \ln \frac{\beta}{1-\alpha}, \\ \text{Continue observation} & & \text{and defer decision} & \text{if } \Lambda_n \in (b, a). \end{array} \right. \quad (2.8)$$

Until Λ_n exits from the interval (b, a) (which eventually happens with probability 1), the null hypothesis (2.1) is neither rejected nor accepted. During this time, the final decision is deferred, and the partial credibility factor estimated by $\hat{Z}_n = cT_n/z_{q/2}$ is used. The same factor is applied if H_0 is finally accepted, and the partial credibility is assigned.

According to [12], testing scheme (2.8) attains

$$\mathbf{P} (\text{Type I error} \mid c\eta = z_{q/2} - \delta) = \alpha \quad \text{and} \quad \mathbf{P} (\text{Type II error} \mid c\eta = z_{q/2}) = \beta. \quad (2.9)$$

Here the Type I error occurs when full credibility is given to a cohort that does not deserve it, and the Type II error means not giving full credibility where it is deserved.

Probabilities α and β are to be chosen before the test. This choice affects the test accuracy as well as the time it will take to reach decision. For example, choosing smaller α and/or β results in lower probabilities of making a Type I or a Type II error. At the same time, smaller α and β result in wider stopping boundaries a and b , according to (2.8), expanding the continue-sampling region. Thus, it will generally take longer for the statistic T_n to reach the boundaries. Then, a decision will be based on a larger sample, increasing its accuracy, in terms of lower error probabilities.

When the method of weight functions is used in presence of nuisance parameters, the probabilities in the right-hand sides of (2.9) are understood in the integral form (2.6), which can also be interpreted as probabilities under the joint distribution of data and nuisance parameters.

3 Implementation for Common Loss Distributions

In this section, the introduced sequential testing scheme is elaborated for commonly used loss models [6] – Poisson number of losses X_k and Exponential, Gamma, Lognormal, Pareto, or Weibull loss amounts Y_{kj} . Each example requires calculating the asymptotic variance V_T according to Theorem 1, using it in the likelihood ratio statistic (2.5) or (2.7), and stating the sequential decision rule in terms of Λ_k or T_n statistics, according to (2.8).

3.1 Example 1: Exponential and Weibull Loss Models

As a special and simple case, consider the Exponential loss distribution. It is special because the asymptotic variance V_T appears free of nuisance parameters (by Corollary 1). Indeed, when $Y_{kj} \sim \text{Exponential}(\theta)$, the moments are $\mathbf{E}(Y_{kj}^m) = m! \theta^m$. Substituting them into (2.3), we obtain $V_T = 1/4$ for any θ .

By (2.5), the SPRT statistic in this case is

$$\Lambda_n = 2n \left(\left(T_n - \frac{z_{q/2} - \delta}{c} \right)^2 - \left(t - \frac{z_{q/2}}{c} \right)^2 \right) = \frac{4n\delta}{c} T_n - \frac{4n\delta z_{q/2}}{c^2} + \frac{2n\delta^2}{c^2},$$

with stopping boundaries $\{b, a\}$. That is, sampling continues while

$$b < \frac{4n\delta}{c} T_n - \frac{4n\delta z_{q/2}}{c^2} + \frac{2n\delta^2}{c^2} < a.$$

Solving for T_n , we obtain the stopping boundaries for the statistic T_n ,

$$\left\{ \frac{bc}{4n\delta} + \frac{z_{q/2}}{c} - \frac{\delta}{2c}, \frac{ac}{4n\delta} + \frac{z_{q/2}}{c} - \frac{\delta}{2c} \right\}.$$

When T_n attains the upper stopping boundary, the cohort is considered fully credible. If it attains the lower boundary, only partial credibility is assigned. Between the boundaries, the decision about the full credibility is deferred until one of the boundaries is crossed.

Similarly, the asymptotic variance V_T is constant for any Weibull distribution with a known shape parameter. For example, when $Y_{kj} \sim \text{Weibull}(\theta, 1/2)$, the m th moment is $\mu_m = \theta^m \Gamma(1 + 2m)$, which leads to $V_T = 17/12$ and the SPRT statistic

$$\Lambda_n = \frac{12n\delta}{17c} T_n - \frac{12n\delta z_{q/2}}{17c^2} + \frac{6n\delta^2}{17c^2}.$$

Reject H_0 and give full credibility after n insured periods when $\Lambda_n \geq a$, or, equivalently,

$$T_n \geq \frac{17ac}{12n\delta} + \frac{z_{q/2}}{c} - \frac{\delta}{2c}.$$

3.2 Example 2: Gamma Loss Model

Extending the Exponential model, let $Y_{kj} \sim \text{Gamma}(r, \theta)$, with the m th moment $\mu_m = \mathbf{E}(Y^m) = \theta^m \Gamma(r + m) / \Gamma(r)$. Substituting these moments into (2.3) and simplifying, we obtain

$$V_T = \frac{r^2 + r + 2}{4(1 + r)^2}.$$

As stated by Corollary 1, the asymptotic variance of T_n is free of the scale parameter of the underlying Gamma loss distribution, however, it contains the nuisance shape parameter r . It can be seen that both integrals in (2.7) diverge to ∞ as $h \rightarrow \infty$, justifying application of the L'Hôpital's Rule. Then, taking the limit as $h \rightarrow \infty$, we obtain the same formula as for the Exponential distribution,

$$\Lambda_n = \frac{4n\delta}{c} T_n - \frac{4n\delta z_{q/2}}{c^2} + \frac{2n\delta^2}{c^2},$$

and again, H_0 is rejected and full credibility is given after n insured periods if

$$T_n \geq \frac{ac}{4n\delta} + \frac{z_{q/2}}{c} - \frac{\delta}{2c}.$$

3.3 Example 3: Lognormal Loss Model

For $Y_{kj} \sim \text{Lognormal}(\mu, \sigma)$, the m th moment is $\mu_m = \exp(m\mu + m^2\sigma^2/2)$, and from (2.3),

$$V_T = 1 - e^{\sigma^2} + \frac{1}{4}e^{3\sigma^2}.$$

In this case, both integrals in (2.7) converge as $h \rightarrow \infty$, so that the nuisance parameter $\theta = \sigma^2$ is being integrated out with the improper prior $\Pi_{\sigma^2}(\theta) = 1_{(0, \infty)}$. Then the sequential

probability ratio test statistic is

$$\Lambda_n = \frac{\int_0^\infty (1 - e^\theta + e^{3\theta}/4)^{-1/2} \cdot \exp \left\{ -\frac{n}{2} \frac{(t - \frac{z}{c})^2}{1 - e^\theta + e^{3\theta}/4} \right\} d\theta}{\int_0^\infty (1 - e^\theta + e^{3\theta}/4)^{-1/2} \cdot \exp \left\{ -\frac{n}{2} \frac{(t - \frac{z-\delta}{c})^2}{1 - e^\theta + e^{3\theta}/4} \right\} d\theta}$$

In this case, there is no simple closed form for the stopping boundary for T_n , however, the inequality $b < \Lambda_n < a$ can be solved for T_n numerically.

3.4 Example 4: Pareto Loss Model

For the Pareto(r, θ) loss distribution, the m th moment is $\mu_m = \theta^m \Gamma(m+1) \Gamma(r-m) / \Gamma(r)$.

Substituting into (2.3), we obtain the asymptotic variance

$$V_T = 1 - \frac{3(r-2)(r-6)}{4(r-3)(r-4)},$$

and by the L^opital's Rule, the same test statistic

$$\Lambda_n = \frac{4n\delta}{c}t - \frac{4n\delta z_{q/2}}{c^2} + \frac{2n\delta^2}{c^2}$$

as in the Gamma case. Thus, full credibility should again be assigned as soon as

$$T_n \geq \frac{ac}{4n\delta} + \frac{z_{q/2}}{c} - \frac{\delta}{2c}.$$

Table 1 summarizes the stopping boundaries for test statistic T_n (the stopping boundary for T_n does not have a closed form in the case of a Lognormal distribution). These are conditions that allow to assign full credibility.

4 Analysis of Actuarial Data

In this section, we apply the developed sequential techniques to the actuarial data generated by the *CAS Public Loss Simulator*. This Simulator was written by Gooouon under the research

Table 1: *Rejection boundary (condition for full credibility) for some loss distributions*

Loss distribution family	Rejection boundary for T_n (Assign full credibility when the inequality holds)
Gamma(r, θ) Pareto(r, θ)	$T_n \geq \frac{ac}{4n\delta} + \frac{z_{q/2}}{c} - \frac{\delta}{2c}$
Weibull($\theta, \frac{1}{2}$)	$T_n \geq \frac{17ac}{12n\delta} + \frac{z_{q/2}}{c} - \frac{\delta}{2c}$

agreement with the Casualty Actuarial Society (CAS) and was made publicly available at http://www.gououon.com/loss_simulator_project.html. It generates data sets close to the real data by simulating all the features usually observed in actuarial records.

Below, we discuss four data sets that are generated in order to illustrate situations leading to different decisions on the full credibility. In these examples, we consider a single-payment model, where each loss event is shortly followed by valuation that results in either no payment or one payment. The probability of no payment is set to $p_0 = 0.2$, which means that roughly 20% of submitted claims are denied. As we know from Corollary 2, the test is independent of the appearance of zero losses in the data. The simulated insured records start on 01/01/2000 and continue for the next 12 years until the present time, with the insurance contract being renewed and premium being revised once a year. At these times of renewal, the credibility decisions are to be made.

Claim frequencies X_k are Poisson, and the non-zero loss amounts Y_{kj} are generated from Gamma and Pareto distributions. The limited fluctuation criterion (1.2) for full credibility is applied with the relative precision $c = 0.1$ and probability $q = 0.05$, and the corresponding hypothesis is tested with the indifference margin $\delta = 0.02$. With these parameters, test (2.1)

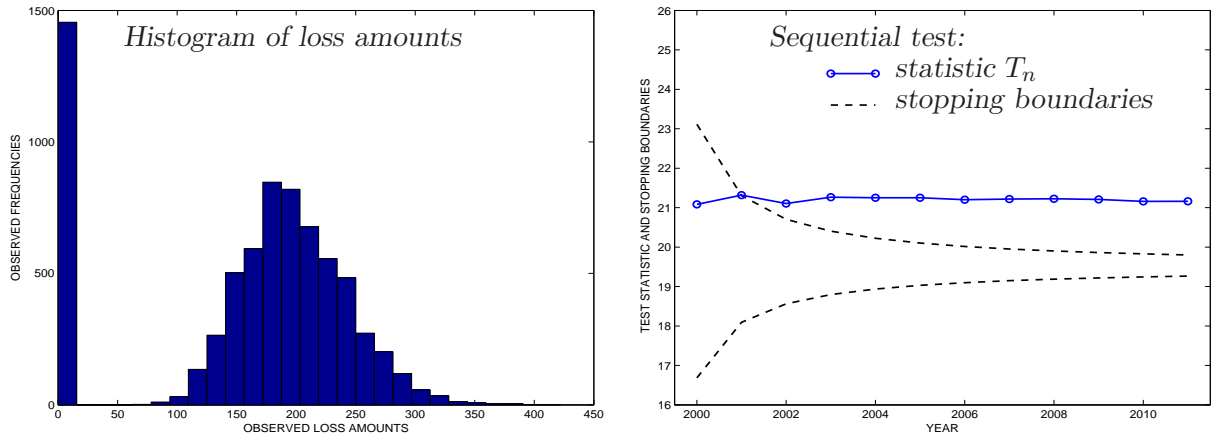


Figure 2: Generated Gamma distributed losses with $\lambda = 600$ claims per year and $p_0 = 20\%$ of claims not covered. Full credibility is awarded in 2001, at the first policy renewal.

reduces to testing

$$H_0 : \eta \leq \frac{z_{q/2} - \delta}{c} = 19.4 \quad \text{vs} \quad H_A : \eta \geq \frac{z_{q/2}}{c} = 19.6. \quad (4.1)$$

Thus, whether a cohort deserves full credibility is determined by the unknown parameter $\eta = \sqrt{\frac{\lambda}{1+\gamma^2}}$. Its sequential estimator T_n and the corresponding stopping boundaries are calculated according to equation (2.2) and Table 1, and the sequential test is conducted for each data sequence resulting in a definite decision on full credibility.

4.1 Gamma Loss Model

We start by generating a 12-year record of Gamma distributed losses with shape parameter $r = 20$, scale parameter $\theta = 10$ (average claim $\mu = \$200$), and the frequency of $\lambda = 600$ claims per year, 20% of which are not covered. This results in $\eta = 21.4$, high enough to support full credibility, according to the limited fluctuation credibility approach.

The histogram of observed losses is on Fig. 2, where a tall column on the left reflects a large portion of uncovered claims ($Y_{kj} = 0$). As we anticipated, the null hypothesis is rejected, claiming full credibility. The test statistic T_n converges to the true value of $\eta = 21.4$, supporting the result of Theorem 1, and quickly exceeds the upper rejection boundary. At

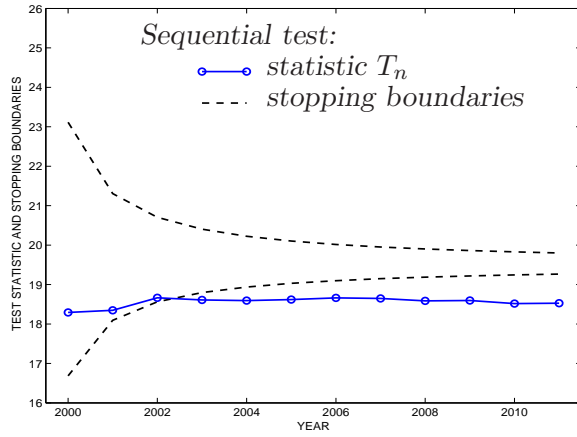


Figure 3: *Generated Gamma distributed losses with $\lambda = 450$ claims per year and $p_0 = 20\%$ of claims not covered. Full credibility is denied in 2003, after four years of observation.*

this time (in 2001, two years after signing the first contract), the null hypothesis is rejected, and full credibility is assigned.

Should fewer claims occur per year, their frequency may be insufficient for the full credibility even if the loss distribution is the same. In the next simulation, the loss model is still Gamma($r = 20, \theta = 10$), however, the frequency is only $\lambda = 450$ claims per year, of which 20% are not covered.

Based on this data, the test statistic T_n converges to the new value of $\eta = 18.5$, crossing the lower stopping boundary, as on Fig. 3. Hence, after three years of no decision, the full credibility is finally denied in 2003 due to a significance evidence of insufficient frequency of claims. Partial credibility will then be assigned with the credibility factor estimated by $\hat{Z}_{2003} = cT_{2003}/z_{q/2} = 0.949$.

4.2 Pareto Loss Model, Heavy Tail

How will the test results of Section 4.1 change if we replace the Gamma loss model with a heavy-tail distribution, say, Pareto? In this section, we again generate 12 years of claims, 01/01/2000 through 12/31/2011. As in our first example on Fig. 2, the frequency of claims

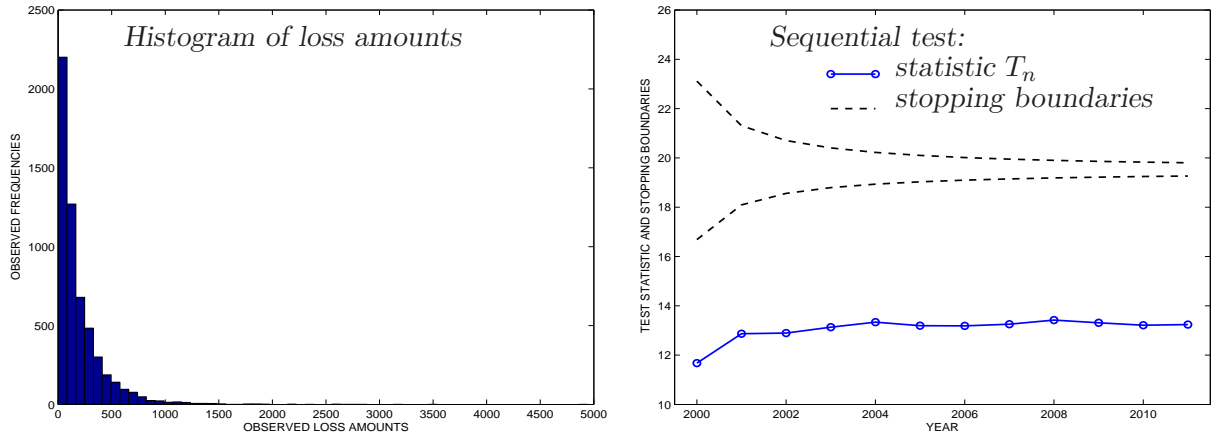


Figure 4: Generated Pareto distributed losses with $\lambda = 600$ claims per year $p_0 = 20\%$ of claims not covered, and the average covered loss amount $\mu = \$200$. Full credibility is denied after the first year of observation.

is $\lambda = 600$ per year, 20% of them are not covered, and the average covered claim amount is $\mu = \$200$. The only difference is that now the loss distribution is $Pareto(r = 5, \theta = 800)$ instead of Gamma. Notice that under this model, $\mathbf{E}(Y_{kj}) = \theta/(r - 1) = 200$ matching the $Gamma(20,10)$ expected value, however, the standard deviation is much higher due to the heavy tail, $\sigma = \frac{\theta}{r-1} \sqrt{\frac{r}{r-2}} = 258.2$ compared to Gamma's $\sigma = \theta\sqrt{r} = 44.7$.

The heavy-tail loss distribution implies a higher variability of the real data, so the limited fluctuation criterion for full credibility is no longer satisfied with the same frequency of claims. Parameter $\eta = 13.4$ is much lower than the tested value in (4.1). As a result, full credibility is denied at the first opportunity (Fig. 4), and partial credibility is applied with the credibility factor $\hat{Z}_{2000} = cT_{2000}/z_{q/2} = 0.60$. Due to high variability of loss amounts, only 60% of the compromise estimator of the pure premium is based on the actually observed data, and the remaining 40% have to come from other sources such as the prior distribution.

Recall that with the same frequency of claims and the same average loss, the full credibility was quickly awarded in the example on Fig. 2, where losses had Gamma distribution.

With Pareto losses, the frequency of claims should be more than twice higher in order

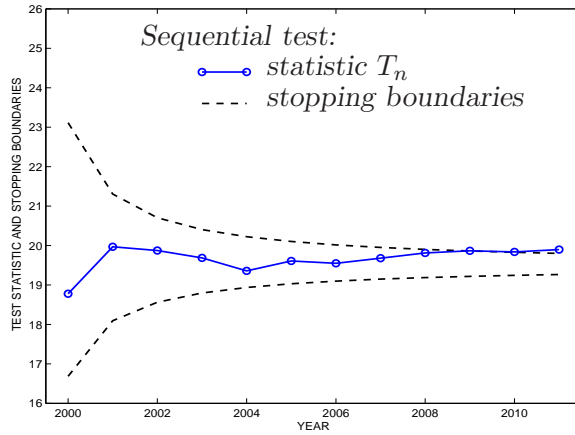


Figure 5: *Borderline case of Pareto loss model with $\lambda = 1300$ claims per year. Full credibility is awarded in 2009, after ten years of observation.*

to satisfy the condition for full credibility. On Fig. 5, the frequency is $\lambda = 1300$ claims per year, and again, 20% of claims are denied. This is almost the minimum λ required for full credibility, as the parameter $\eta = 19.75$ barely agrees with the alternative hypothesis H_A . As a result of this borderline η , it takes ten years until 2009 before the decision is made and full credibility is finally applied.

In all the considered examples, our scheme, with sequentially estimated parameters and sequentially re-tested hypotheses (2.1), eventually agreed with the standard limited fluctuation credibility approach that assumes known parameters. However, since the frequency and severity parameters are in practice estimated from the data, results can differ, and a Type I or a Type II error can occur. Probabilities of that are controlled by (2.9).

5 Conclusion

This article introduces sequential methods in Actuarial Science. It is motivated by a number of routine decisions that actuaries make at regular time intervals, based on sequentially collected data. Using elementary statistical tools without accounting for multiple tests, the error rate increases with the number of decisions made and considerably exceeds the nominal

significance level of each test. On the other hand, applying properly designed sequential techniques, one can control the error rate at the specified level. Level α and power $(1 - \beta)$ sequential tests for full credibility are derived in this article. As a result, conditions are determined under which an insured cohort becomes fully credible, according to the limited fluctuation credibility approach. A stopping rule is derived when the sequential test for full credibility rejects the null hypothesis, in which case full credibility can be assigned.

Although the introduced method is elaborated for credibility assessment, sequential methods in general and the proposed method of handling nuisance parameters can be applied to other areas. In any situation when a decision is made as soon as the data collected to the moment shows significant evidence, one should be looking for proper sequential methods in order to control the overall error rate.

6 Appendix

6.1 Proof of Lemma 1

Estimating parameter $\eta = \sqrt{\lambda/(1 + \gamma^2)}$ with $T_n = \sqrt{\bar{X}_n/(1 + s_n^2/\bar{Y}_n^2)}$, we use the mean frequency $\bar{X}_n = n^{-1} \sum_1^n X_k$ to estimate the frequency parameter λ_n , and the severity sample mean $\bar{Y}_n = \sum_{k=1}^n \sum_{j=1}^{X_k} Y_{kj} / \sum_{k=1}^n X_k$ and the severity sample variance

$$s_n^2 = \frac{\sum_{k=1}^n \sum_{j=1}^{X_k} (Y_{kj} - \bar{Y}_n)^2}{\sum_{k=1}^n X_k} = \frac{\sum_{k=1}^n X_k \sum_{j=1}^{X_k} Y_{kj}^2 - \left(\sum_{k=1}^n \sum_{j=1}^{X_k} Y_{kj} \right)^2}{\left(\sum_{k=1}^n X_k \right)^2}$$

to estimate the sample coefficient of variation $\gamma = \sigma/\mu$.

Substituting this version of the sample variance into T_n and simplifying, we obtain,

$$\begin{aligned}
T_n^2 &= \frac{n^{-1} \sum_{k=1}^n X_k}{1 + \left[\sum_{k=1}^n X_k \sum_{k=1}^n \sum_{j=1}^{X_k} Y_{kj}^2 - \left(\sum_{k=1}^n \sum_{j=1}^{X_k} Y_{kj} \right)^2 \right] / \left(\sum_{k=1}^n \sum_{j=1}^{X_k} Y_{kj} \right)^2} \\
&= n^{-1} \left(\sum_{k=1}^n \sum_{j=1}^{X_k} Y_{kj} \right)^2 / \sum_{k=1}^n \sum_{j=1}^{X_k} Y_{kj}^2, \tag{6.1}
\end{aligned}$$

and (2.4) follows.

6.2 Proof of Theorem 1

This proof of asymptotic normality of T_n is based on the Multivariate Central Limit Theorem ([7], Theorem 5.4.4), representing T_n in terms of the sums of independent random variables.

From Lemma 1, $T_n = \bar{S}_n / \sqrt{\bar{U}_n}$, where $S_k = \sum_{j=1}^{X_k} Y_{kj}$, $U_k = \sum_{j=1}^{X_k} Y_{kj}^2$, $\bar{S}_n = n^{-1} \sum_1^n S_k$, and $\bar{U}_n = n^{-1} \sum_1^n U_k$.

The last expression shows that statistic T_n is a smooth function of two sample means, \bar{S}_n and \bar{U}_n , that are based on i.i.d. variables S_1, S_2, \dots and i.i.d. variables U_1, U_2, \dots with

$$\mathbf{E}(S_k) = \mathbf{E} \left(\mathbf{E} \left\{ \sum_{j=1}^{X_k} Y_{kj} \middle| X_k \right\} \right) = \mathbf{E}(X_k \mu) = \lambda \mu,$$

$$\mathbf{E}(U_k) = \mathbf{E} \left(\mathbf{E} \left\{ \sum_{j=1}^{X_k} Y_{kj}^2 \middle| X_k \right\} \right) = \mathbf{E}(X_k \mu_2) = \lambda \mu_2,$$

$$\text{Var}(S_k) = \mathbf{E} \text{Var} \left(\sum_{j=1}^{X_k} Y_{kj} \middle| X_k \right) + \text{Var} \left(\mathbf{E} \left\{ \sum_{j=1}^{X_k} Y_{kj} \middle| X_k \right\} \right)$$

$$= \mathbf{E} X_k \sigma^2 + \text{Var}(X_k \mu) = \lambda \sigma^2 + \tau^2 \mu^2,$$

$$\text{Var}(U_k) = \mathbf{E} \text{Var} \left(\sum_{j=1}^{X_k} Y_{kj}^2 \middle| X_k \right) + \text{Var} \left(\mathbf{E} \left\{ \sum_{j=1}^{X_k} Y_{kj}^2 \middle| X_k \right\} \right)$$

$$= \mathbf{E} X_k (\mu_4 - \mu_2^2) + \text{Var}(X_k \mu_2) = \lambda (\mu_4 - \mu_2^2) + \tau^2 \mu_2^2,$$

and

$$\begin{aligned}
\text{Cov}(S_k, U_k) &= \mathbf{E} \text{Cov}(S_k, U_k | X_k) + \text{Cov}(\mathbf{E}\{S_k | X_k\}, \mathbf{E}\{U_k | X_k\}) \\
&= \mathbf{E} \text{Cov} \left(\sum_{j=1}^{X_k} Y_{kj}, \sum_{j=1}^{X_k} Y_{kj}^2 \middle| X_k \right) + \text{Cov} \left(\mathbf{E} \left\{ \sum_{j=1}^{X_k} Y_{kj} \middle| X_k \right\}, \mathbf{E} \left\{ \sum_{j=1}^{X_k} Y_{kj}^2 \middle| X_k \right\} \right) \\
&= \mathbf{E} X_k (\mu_3 - \mu \mu_2) + \text{Cov}(X_k \mu, X_k \mu_2) = \lambda (\mu_3 - \mu \mu_2) + \tau^2 \mu \mu_2,
\end{aligned}$$

for $k = 1, 2, \dots$, provided that the frequency distribution has two finite moments, and the severity distribution has four finite moments.

Random vectors $(S_k, U_k)'$ are independent, identically distributed for $k = 1, \dots, n$, with the finite mean vector $(\lambda \mu, \lambda \mu_2)'$ and finite covariance matrix

$$\Sigma = \text{VAR} \begin{pmatrix} S_k \\ U_k \end{pmatrix} = \begin{pmatrix} \lambda \sigma^2 + \tau^2 \mu^2 & \lambda (\mu_3 - \mu \mu_2) + \tau^2 \mu \mu_2 \\ \lambda (\mu_3 - \mu \mu_2) + \tau^2 \mu \mu_2 & \lambda (\mu_4 - \mu_2^2) + \tau^2 \mu_2^2 \end{pmatrix}. \quad (6.2)$$

The Bivariate Central Limit Theorem (e.g., Theorem 5.2.1 of [7]) holds for $(\bar{S}_n, \bar{U}_n)'$ if $\det(\Sigma) \neq 0$.

On the contrary, suppose that matrix Σ is singular. Then there exist coefficients u and v such that $u^2 + v^2 \neq 0$ and

$$u S_k + v U_k = \sum_{j=1}^{X_k} (u Y_{kj} + v Y_{kj}^2) = \text{const}$$

almost surely. This implies that $\text{Var}(u S_k + v U_k) = 0$. On the other hand, letting $W_{kj} = u Y_{kj} + v Y_{kj}^2$, we have

$$\text{Var}(u S_k + v U_k) = \mathbf{E} \text{Var} \left\{ \sum_{j=1}^{X_k} W_{kj} \middle| X_k \right\} + \text{Var} \mathbf{E} \left\{ \sum_{j=1}^{X_k} W_{kj} \middle| X_k \right\} = \lambda \text{Var}(W_{kj}) + \tau^2 \mathbf{E}^2(W_{kj}).$$

Since $\lambda > 0$ and $\tau^2 > 0$ by the conditions of this Theorem, the only possibility for $\text{Var}(u S_k + v U_k) = 0$ is that $\text{Var}(W_{kj}) = 0$ and $\mathbf{E}(W_{kj}) = 0$, which means that $W_{kj} = u Y_{kj} + v Y_{kj}^2 = 0$

with probability one. Solving this equation for Y_{kj} , we obtain that $Y_{kj} \in \{0, -u/v\}$ a.s. However, this implies that the loss variable Y_{kj} is either degenerate or it is a multiple of a Bernoulli random variable,

$$-\frac{v}{u} Y_{kj} \sim \text{Bernoulli}.$$

This contradicts assumptions of the Theorem, therefore, matrix Σ is non-singular.

Applying the Bivariate Central Limit Theorem, we obtain

$$\sqrt{n}(\bar{S}_n - \lambda\mu, \bar{U}_n - \lambda\mu_2)' \xrightarrow{d} \text{Normal} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right).$$

Next, applying the bivariate delta method (Lehmann 1999, Corollary 5.2.1) to sequences \bar{S}_n and \bar{U}_n and function $f(u, v) = uv^{-1/2}$, we obtain that

$$\sqrt{n}(T_n - \eta) = \sqrt{n}(f(\bar{S}_n, \bar{U}_n) - f(\lambda\mu, \lambda\mu_2)) \xrightarrow{d} N(0, V_T), \quad (6.3)$$

where the asymptotic variance V_T is derived by the bivariate delta method as follows.

For the function $f(u, v) = uv^{-1/2}$, we have $\partial f/\partial u = v^{-1/2}$ and $\partial f/\partial v = -uv^{-3/2}/2$.

Substituting $u = \mathbf{E}\bar{S}_n = \lambda\mu$ and $v = \mathbf{E}\bar{U}_n = \lambda\mu_2$, we obtain

$$\begin{aligned} V_T &= \left(\frac{\partial f}{\partial u} \right)^2 \Sigma_{11} + 2 \frac{\partial f}{\partial u} \frac{\partial f}{\partial v} \Sigma_{12} + \left(\frac{\partial f}{\partial v} \right)^2 \Sigma_{22} = \frac{\Sigma_{11}}{v} - \frac{u\Sigma_{12}}{v^2} + \frac{u^2\Sigma_{22}}{4v^3} \\ &= \frac{\lambda(\mu_2 - \mu^2) + \tau^2\mu^2}{\lambda\mu_2} - \frac{\lambda\mu(\lambda(\mu_3 - \mu\mu_2) + \tau^2\mu\mu_2)}{\lambda^2\mu_2^2} + \frac{\lambda^2\mu^2(\lambda(\mu_4 - \mu_2^2) + \tau^2\mu_2^2)}{4\lambda^3\mu_2^3} \\ &= \left(\frac{\tau^2}{\lambda} \right) \frac{\mu^2}{4\mu_2} + \frac{\mu_2^2 - \mu\mu_3}{\mu_2^2} + \frac{\mu^2(\mu_4 - \mu_2^2)}{4\mu_2^3}. \end{aligned}$$

Acknowledgements

Support of this research by The Casualty Actuarial Society is greatly appreciated.

References

- [1] H. Bühlmann. *Mathematical Methods in Risk Theory*. Springer, New York, 1970.
- [2] H. Bühlmann and A. Gisler. *A Course in Credibility Theory and its Applications*. Springer-Verlag, Berlin, 2005.
- [3] Z. Govindarajulu. *Sequential Statistics*. World Scientific Publishing Co, Singapore, 2004.
- [4] T. N. Herzog. *Introduction to Credibility Theory*. ACTEX Publications, Winsted,CT, 2004.
- [5] I. B. Hossack, J. H. Pollard, and B. Zehnwirth. *Introductory Statistics with Applications in General Insurance*. Cambridge University Press, New York, 1983.
- [6] S. A. Klugman, H. H. Panjer, and G. E. Willmot. *Loss Models: From Data to Decisions, 4th Edition*. Wiley, New York, 2012.
- [7] E. L. Lehmann. *Elements of Large-Sample Theory*. Springer-Verlag, New York, 1999.
- [8] L. H. Longley-Cook. An introduction to credibility theory. *PCAS*, XLIX(92):194–221, 1962.
- [9] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville. *Sequential Analysis: Hypothesis Testing and Change-Point Detection*. Chapman & Hall/CRC, 2014.
- [10] Y.-K. Tse. *Nonlife Actuarial Models: Theory, Methods and Evaluation*. Cambridge University Press, Cambridge, England, 2009.
- [11] G. G. Venter. *Credibility, in Foundations of Casualty Actuarial Science*. Casualty Actuarial Society, New York, 1990.
- [12] A. Wald. *Sequential Analysis*. Wiley, New York, 1947.