Part III

Credibility

Credibility theory provides the basic analytical framework for pricing insurance products. The importance of combining information about the recent experience of the individuals versus the aggregate past experience has been recognized in the literature through the classical approach. Rigorous analytical treatment of the subject started with Hans Bühlmann, and much work has been accomplished by him and his students. Bühlmann's approach provides a simple solution to the Bayesian method and achieves optimality within the subset of linear predictors. In this part of the book we introduce the classical approach, the Bühlmann approach, the Bayesian method, as well as the empirical implementation of these techniques.

6
Classical credibility

Credibility models were first proposed in the beginning of the twentieth century to update predictions of insurance losses in light of recently available data of insurance claims. The oldest approach is the limited-fluctuation credibility method, also called the classical approach, which proposes to update the loss prediction as a weighted average of the prediction based purely on the recent data and the rate in the insurance manual. Full credibility is achieved if the amount of recent data is sufficient, in which case the updated prediction will be based on the recent data only. If, however, the amount of recent data is insufficient, only partial credibility is attributed to the data and the updated prediction depends on the manual rate as well.

We consider the calculation of the minimum size of the data above which full credibility is attributed to the data. For cases where the data are insufficient we derive the partial-credibility factor and the updating formula for the prediction of the loss. The classical credibility approach is applied to update the prediction of loss measures such as the frequency of claims, the severity of claims, the aggregate loss, and the pure premium of a block of insurance policies.

Learning objectives

1 Basic framework of credibility
2 The limited-fluctuation (classical) credibility approach
3 Full credibility
4 Partial credibility
5 Prediction of claim frequency, claim severity, aggregate loss, and pure premium

6.1 Framework and notations

We consider a block of insurance policies, referred to as a **risk group**. Examples of risk groups of interest are workers of a company covered under a workers accident compensation scheme, employees of a firm covered under employees health insurance, and a block of vehicle insurance policies. The risk group is covered over a period of time (say, one year) upon the payment of a premium. The premium is partially based on a rate specified in the manual, called the **manual rate** and partially on the specific risk characteristics of the group. Based upon the recent **claim experience** of the risk group, the premium for the next period will be revised. Credibility theory concerns the updating of the prediction of the claim for the next period using the recent claim experience and the manual rate. The revised prediction determines the insurance premium of the next period for the risk group.

Credibility theory may be applied to different measures of claim experience. We summarize below the key factors of interest and define our notations to be used subsequently:

Claim frequency: The number of claims in the period is denoted by N.

Aggregate loss: We denote the amount of the ith claim by X_i and the aggregate loss by S, so that $S = X_1 + X_2 + \cdots + X_N$.

Claim severity: The average claim severity is the sample mean of $X_1, \cdots, X_N$, i.e. $\bar{X} = S/N$.

Pure premium: Let E be the number of exposure units of the risk group, the pure premium P is defined as $P = S/E$.

The loss measures N, X_i, S, $\bar{X}$, and P are random variables determined by uncertain events, while the exposure E is a known constant measuring the size of the risk group. For workers compensation and employees health insurance, E may be measured as the number of workers or employees covered under the policies.

We denote generically the predicted loss based on the manual by M, and the observed value of the loss based on recent data of the experience of the risk group by D. Thus, M and D may refer to the predicted value and observed value, respectively, of N, S, $\bar{X}$, and P. The **classical credibility** approach (also called the **limited-fluctuation credibility** approach) proposes to formulate the updated prediction of the loss measure as a weighted average of D and M. The weight attached to D is called the **credibility factor**, and is denoted by Z, with $0 \leq Z \leq 1$. Thus, the updated prediction, generically denoted by U, is given by

$$U = ZD + (1 - Z)M. \tag{6.1}$$

Example 6.1 The loss per worker insured in a ship-building company was \$230 last year. If the pure premium per worker in a similar industry is \$292 and the credibility factor of the company (the risk group) is 0.46, calculate the updated predicted pure premium for the company's insurance.

Solution We have $D = 230$, $M = 292$, and $Z = 0.46$, so that from equation (6.1) we obtain

$$U = (0.46)(230) + (1 - 0.46)(292) = \$263.48,$$

which will be the pure premium charged per worker of the company next year. □

From equation (6.1) we observe that U is always between the experience measure D and the manual rate M. The closer Z is to 1, the closer the updated predicted value U will be to the observed measure D. The credibility factor Z determines the relative importance of the data in calculating the updated prediction. **Full credibility** is said to be achieved if $Z = 1$, in which case the prediction depends upon the data only but not the manual. When $Z < 1$, the data are said to have **partial credibility**. Intuitively, a larger data set would justify a larger Z.

For the classical frequentist approach in statistics with no extraneous (or prior) information, estimation and prediction are entirely based on the data available. Thus, in the frequentist statistical framework one might say that all data have full credibility. The credibility theory literature, however, attempts to use extraneous (or prior) information (as provided by insurance manuals) to obtain an improved update of the prediction.

6.2 Full credibility

The classical credibility approach determines the minimum data size required for the experience data to be given full credibility (namely, for setting $Z = 1$). The minimum data size is called the **standard for full credibility**, which depends on the loss measure of interest. In this section we derive the formulas for the computation of the full-credibility standards for loss measures such as claim frequency, claim severity, aggregate loss, and pure premium.

6.2.1 Full credibility for claim frequency

Assume that the claim frequency random variable N has mean μ_N and variance σ_N^2. To assess how likely an observed value of N is "representative" of the true mean, we ask the following question: What is the probability of observing claim frequency within $100k\%$ of the mean? This probability is given by

$$\Pr(\mu_N - k\mu_N \le N \le \mu_N + k\mu_N) = \Pr\left(-\frac{k\mu_N}{\sigma_N} \le \frac{N-\mu_N}{\sigma_N} \le \frac{k\mu_N}{\sigma_N}\right). \tag{6.2}$$

If we further assume that N is normally distributed, then $(N-\mu_N)/\sigma_N$ follows a standard normal distribution. Thus, denoting $\Phi(\cdot)$ as the df of the standard normal random variable, expression (6.2) becomes

$$\begin{aligned}\Pr\left(-\frac{k\mu_N}{\sigma_N} \le \frac{N-\mu_N}{\sigma_N} \le \frac{k\mu_N}{\sigma_N}\right) &= \Phi\left(\frac{k\mu_N}{\sigma_N}\right) - \Phi\left(-\frac{k\mu_N}{\sigma_N}\right)\\ &= \Phi\left(\frac{k\mu_N}{\sigma_N}\right) - \left[1 - \Phi\left(\frac{k\mu_N}{\sigma_N}\right)\right]\\ &= 2\Phi\left(\frac{k\mu_N}{\sigma_N}\right) - 1. \end{aligned} \tag{6.3}$$

If

$$\frac{k\mu_N}{\sigma_N} = z_{1-\frac{\alpha}{2}}, \tag{6.4}$$

where z_β is the 100βth percentile of the standard normal, i.e. $\Phi(z_\beta) = \beta$, then the probability in (6.3) is given by $2(1-\alpha/2) - 1 = 1 - \alpha$. Thus, there is a probability of $1-\alpha$ that an observed claim frequency is within $100k\%$ of the true mean, where α satisfies equation (6.4).

Example 6.2 Suppose the claim frequency of a risk group is normally distributed with mean 420 and variance 521, find the probability that the observed number of claims is within 10% of the true mean. Find the symmetric interval about the mean which covers 90% of the observed claim frequency. Express this interval in percentage error of the mean.

Solution We first calculate

$$\frac{k\mu_N}{\sigma_N} = \frac{(0.1)(420)}{\sqrt{521}} = 1.8401.$$

As $\Phi(1.8401) = 0.9671$, the required probability is $2(0.9671) - 1 = 0.9342$. Thus, the probability of observing claim frequency within 10% of the true mean is 93.42%.

Now $z_{0.95} = 1.645$. Using equation (6.4), we have

$$\frac{k\mu_N}{\sigma_N} = 1.645,$$

so that

$$k = \frac{1.645\,\sigma_N}{\mu_N} = \frac{1.645\sqrt{521}}{420} = 0.0894,$$

and there is a probability of 90% that the observed claim frequency is within 8.94% of the true mean. □

Note that the application of equation (6.4) requires knowledge of μ_N and σ_N^2. To simplify the computation, it is often assumed that the claim frequency is distributed as a Poisson variable with mean λ_N, which is large enough so that the normal approximation applies. Due to the Poisson assumption, we have $\mu_N = \sigma_N^2 = \lambda_N$, and equation (6.4) can be written as

$$\frac{k\lambda_N}{\sqrt{\lambda_N}} = k\sqrt{\lambda_N} = z_{1-\frac{\alpha}{2}}. \tag{6.5}$$

Example 6.3 Repeat Example 6.2, assuming that the claim frequency is distributed as a Poisson with mean 850, and that the normal approximation can be used for the Poisson.

Solution From equation (6.5), we have

$$k\sqrt{\lambda_N} = 0.1\sqrt{850} = 2.9155.$$

As $\Phi(2.9155) = 0.9982$, the coverage probability within 10% of the mean is $2(0.9982) - 1 = 99.64\%$. For coverage probability of 90%, we have

$$k = \frac{1.645}{\sqrt{850}} = 0.0564,$$

so that there is a probability of 90% that the observed claim frequency is within 5.64% of the mean. □

In the classical credibility approach, full credibility is attained for claim frequency if there is a probability of at least $1 - \alpha$ that the observed number of claims is within $100k\%$ of the true mean, where α and k are some given values. From equation (6.5) we can see that, under the Poisson assumption, the above probability statement holds if

$$k\sqrt{\lambda_N} \geq z_{1-\frac{\alpha}{2}}. \tag{6.6}$$

Table 6.1. *Selected values of standard for full credibility λ_F for claim frequency*

α	Coverage probability	k: 10%	5%	1%
0.20	80%	165	657	16,424
0.10	90%	271	1,083	27,056
0.05	95%	385	1,537	38,415
0.01	99%	664	2,654	66,349

Hence, subject to the following assumptions concerning the claim frequency distribution:

1 the claim frequency is distributed as a Poisson variable,
2 the mean of the Poisson variable is large enough to justify the normal approximation to the Poisson,

full credibility for claim frequency is attributed to the data if $\lambda_N \geq (z_{1-\frac{\alpha}{2}}/k)^2$. We now define

$$\lambda_F \equiv \left(\frac{z_{1-\frac{\alpha}{2}}}{k}\right)^2, \tag{6.7}$$

which is the standard for full credibility for claim frequency, i.e. full credibility is attained if $\lambda_N \geq \lambda_F$.

As the expected number of claims λ_N is unknown in practice, the implementation of the credibility model is to compare the observed value of N in the recent period against λ_F calculated using equation (6.7). Full credibility is attained if $N \geq \lambda_F$. Table 6.1 presents the values of λ_F for selected values of α and k.

Table 6.1 shows that, given the accuracy parameter k, the standard for full credibility λ_F increases with the required coverage probability $1-\alpha$. Likewise, given the required coverage probability $1-\alpha$, λ_F increases with the required accuracy (i.e. decreases with k).

Example 6.4 If an insurance company requires a coverage probability of 99% for the number of claims to be within 5% of the true expected claim frequency, how many claims in the recent period are required for full credibility? If the insurance company receives 2,890 claims this year from the risk group and the manual list of expected claim is 3,000, what is the updated expected number of claims next year? Assume the claim-frequency distribution is Poisson and the normal approximation applies.

Solution We compute λ_F using equation (6.7) to obtain

$$\lambda_F = \left(\frac{z_{0.995}}{0.05}\right)^2 = \left(\frac{2.576}{0.05}\right)^2 = 2{,}653.96.$$

Hence, 2,654 claims are required for full credibility. As the observed claim frequency of 2,890 is larger than 2,654, full credibility is attributed to the data, i.e. $Z = 1$. Thus, $1 - Z = 0$, and from equation (6.1) the updated estimate of the expected number of claims in the next period is 2,890. Note that as full credibility is attained, the updated prediction does not depend on the manual value of $M = 3{,}000$. □

Example 6.5 If an insurance company decides to assign full credibility for 800 claims or more, what is the required coverage probability for the number of claims to be within 8% of the true value? Assume the claim-frequency distribution is Poisson and the normal approximation applies.

Solution To find the coverage probability when 800 claims are sufficient to acquire full credibility to within 8% of the true mean, we apply equation (6.5) to find α, which satisfies

$$k\sqrt{\lambda_N} = 0.08\sqrt{800} = 2.2627 = z_{1-\frac{\alpha}{2}},$$

so that $\alpha = 0.0237$ and the coverage probability is $1 - 0.0237 = 97.63\%$. □

Standard for full credibility is sometimes expressed in terms of the number of exposure units. The example below illustrates an application.

Example 6.6 Recent experience of a workers compensation insurance has established the mean accident rate as 0.045 and the standard for full credibility of claims as 1,200. For a group with a similar risk profile, what is the minimum number of exposure units (i.e. number of workers in the group) required for full credibility?

Solution As the standard for full credibility has been established for claim frequency, the standard for full credibility based on exposure is

$$\frac{1{,}200}{0.045} = 26{,}667 \text{ workers.}$$

□

6.2.2 Full credibility for claim severity

We now consider the standard for full credibility when the loss measure of interest is the claim severity. Suppose there is a sample of N claims of amounts $X_1, X_2, \cdots, X_N$. We assume $\{X_i\}$ to be iid with mean μ_X and variance σ_X^2, and use the sample mean $\bar{X}$ to estimate μ_X. Full credibility is attributed to $\bar{X}$ if the probability of $\bar{X}$ being within $100k\%$ of the true mean of claim loss μ_X is at least $1-\alpha$, for given values of k and α. We also assume that the sample size N is sufficiently large so that $\bar{X}$ is approximately normally distributed with mean μ_X and variance σ_X^2/N. Hence, the coverage probability is

$$\Pr(\mu_X - k\mu_X \leq \bar{X} \leq \mu_X + k\mu_X) = \Pr\left(-\frac{k\mu_X}{\frac{\sigma_X}{\sqrt{N}}} \leq \frac{\bar{X}-\mu_X}{\frac{\sigma_X}{\sqrt{N}}} \leq \frac{k\mu_X}{\frac{\sigma_X}{\sqrt{N}}}\right)$$
$$\simeq 2\Phi\left(\frac{k\mu_X}{\frac{\sigma_X}{\sqrt{N}}}\right) - 1. \tag{6.8}$$

For the coverage probability to be larger than $1-\alpha$, we must have

$$\frac{k\mu_X}{\frac{\sigma_X}{\sqrt{N}}} \geq z_{1-\frac{\alpha}{2}}, \tag{6.9}$$

so that

$$N \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{k}\right)^2 \left(\frac{\sigma_X}{\mu_X}\right)^2, \tag{6.10}$$

which is the standard for full credibility for severity. Note that the coefficient of variation of X is $C_X = \sigma_X/\mu_X$. Using equation (6.7), expression (6.10) can be written as

$$N \geq \lambda_F C_X^2. \tag{6.11}$$

Hence, $\lambda_F C_X^2$ is the standard for full credibility for claim severity. If the experience claim frequency exceeds $\lambda_F C_X^2$, $\bar{X}$ will be the predictor for the average severity of the next period (i.e. the manual rate will not be relevant). Note that in the derivation above, N is treated as a constant rather than a random variable dependent on the claim experience. To implement the methodology in practice, μ_X and σ_X^2 have to be estimated from the sample.

Example 6.7 What is the standard for full credibility for claim severity with $\alpha = 0.01$ and $k = 0.05$, given that the mean and variance estimates of the severity are 1,000 and 2,000,000, respectively?

Solution From Table 1, we have $\lambda_F = 2{,}654$. Thus, using equation (6.11), the standard for full credibility for severity is

$$2{,}654\left[\frac{2{,}000{,}000}{(1{,}000)(1{,}000)}\right] = 5{,}308.$$

□

In this example the standard for full credibility is higher for the severity estimate than for the claim-frequency estimate. As shown in equation (6.11), the standard for full credibility for severity is higher (lower) than that for claim frequency if the coefficient of variation of X is larger (smaller) than 1.

In deriving the standard for full credibility for severity, we do not make use of the assumption of Poisson distribution for claim frequency. The number of claims, however, must be large enough to justify normal approximation for the average loss per claim $\bar{X}$.

6.2.3 Full credibility for aggregate loss

To derive the standard for full credibility for aggregate loss, we determine the minimum (expected) claim frequency such that the probability of the observed aggregate loss S being within $100k\%$ of the expected aggregate loss is at least $1-\alpha$; that is, denoting μ_S and σ_S^2 as the mean and variance of S, respectively, we need to evaluate

$$\Pr(\mu_S - k\mu_S \leq S \leq \mu_S + k\mu_S) = \Pr\left(-\frac{k\mu_S}{\sigma_S} \leq \frac{S-\mu_S}{\sigma_S} \leq \frac{k\mu_S}{\sigma_S}\right). \quad (6.12)$$

To compute μ_S and σ_S^2, we use the compound distribution formulas derived in Appendix A.12 and applied in Section 1.5.1. Specifically, if N and $X_1, X_2, \cdots, X_N$ are mutually independent, we have $\mu_S = \mu_N \mu_X$ and $\sigma_S^2 = \mu_N \sigma_X^2 + \mu_X^2 \sigma_N^2$. If we further assume that N is distributed as a Poisson variable with mean λ_N, then $\mu_N = \sigma_N^2 = \lambda_N$, and we have $\mu_S = \lambda_N \mu_X$ and $\sigma_S^2 = \lambda_N(\mu_X^2 + \sigma_X^2)$. Thus

$$\frac{\mu_S}{\sigma_S} = \frac{\lambda_N \mu_X}{\sqrt{\lambda_N(\mu_X^2 + \sigma_X^2)}} = \frac{\mu_X \sqrt{\lambda_N}}{\sqrt{\mu_X^2 + \sigma_X^2}}. \quad (6.13)$$

Applying normal approximation to the distribution of the aggregate loss S, equation (6.12) can be written as

$$\begin{aligned}\Pr(\mu_S - k\mu_S \leq S \leq \mu_S + k\mu_S) &\simeq 2\Phi\left(\frac{k\mu_S}{\sigma_S}\right) - 1 \\ &= 2\Phi\left(\frac{k\mu_X\sqrt{\lambda_N}}{\sqrt{\mu_X^2 + \sigma_X^2}}\right) - 1. \quad (6.14)\end{aligned}$$

For the above probability to be at least $1 - \alpha$, we must have

$$\frac{k\mu_X\sqrt{\lambda_N}}{\sqrt{\mu_X^2 + \sigma_X^2}} \geq z_{1-\frac{\alpha}{2}}, \tag{6.15}$$

so that

$$\lambda_N \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{k}\right)^2 \left(\frac{\mu_X^2 + \sigma_X^2}{\mu_X^2}\right). \tag{6.16}$$

Thus, the standard for full credibility for aggregate loss is

$$\left(\frac{z_{1-\frac{\alpha}{2}}}{k}\right)^2 \left(\frac{\mu_X^2 + \sigma_X^2}{\mu_X^2}\right) = \lambda_F\left(1 + C_X^2\right). \tag{6.17}$$

From equations (6.7) and (6.17), it can be seen that the standard for full credibility for aggregate loss is always higher than that for claim frequency. This result is due to the randomness of both the claim frequency and the claim severity in impacting the aggregate loss. Indeed, as

$$\lambda_F\left(1 + C_X^2\right) = \lambda_F + \lambda_F C_X^2, \tag{6.18}$$

we conclude that

Standard for full credibility for aggregate loss

= Standard for full credibility for claim frequency

+ Standard for full credibility for claim severity.

Example 6.8 A block of health insurance policies has estimated mean severity of 25 and variance of severity of 800. For $\alpha = 0.15$ and $k = 0.08$, calculate the standard for full credibility for claim frequency and aggregate loss. Assume the claim frequency follows a Poisson distribution and normal approximation can be used for the claim-frequency and aggregate-loss distributions. If the block has an expected number of claims of 400 for the next period, is full credibility attained?

Solution As $z_{0.925} = \Phi^{-1}(0.925) = 1.4395$, from equation (6.7) we have

$$\lambda_F = \left(\frac{1.4395}{0.08}\right)^2 = 323.78,$$

which is the standard for full credibility for claim frequency. The coefficient of variation of the claim severity is

$$\frac{\sqrt{800}}{25} = 1.1314.$$

Thus, using equation (6.17) the standard for full credibility for aggregate loss is

$$(323.78)[1 + (1.1314)^2] = 738.24,$$

which is 2.28 times that of the standard for full credibility for claim frequency. The expected number of claims for the next period, 400, is larger than 323.78 but smaller than 738.24. Thus, full credibility is attained for the risk group for claim frequency but not for aggregate loss. □

Example 6.9 Data for the claim experience of a risk group in the current period show the following: (a) there are 542 claims and (b) the sample mean and variance of the claim severity are, respectively, 48 and 821. For $\alpha = 0.01$ and $k = 0.1$, do the data justify full credibility for claim frequency and claim severity for the next period?

Solution From Table 1, the standard for full credibility for claim frequency at $\alpha = 0.01$ and $k = 0.1$ is 664, which is larger than the claim frequency of 542. Thus, full credibility is not attained for claim frequency. To calculate the standard for full credibility for severity, we use equation (6.11) to obtain

$$\lambda_F C_X^2 = 664 \left[\frac{821}{(48)^2} \right] = 236.61.$$

As $542 > 236.61$, full credibility is attained for claim severity. □

6.2.4 Full credibility for pure premium

Pure premium, denoted by P, is the premium charged to cover losses before taking account of expenses and profits. Repeating the arguments as before, we evaluate the following probability

$$\Pr(\mu_P - k\mu_P \le P \le \mu_P + k\mu_P) = 2\Phi\left(\frac{k\mu_P}{\sigma_P}\right) - 1, \tag{6.19}$$

where μ_P and σ_P^2 are, respectively, the mean and variance of P. As $P = S/E$, where the number of exposure units E is a constant, we have $\mu_P/\sigma_P = \mu_S/\sigma_S$. Thus, the final expression in equation (6.14) can be used to calculate the probability in equation (6.19), and we conclude that the standard for full credibility for pure premium is the same as that for aggregate loss.

Example 6.10 A block of accident insurance policies has mean claim frequency of 0.03 per policy. Claim-frequency distribution is assumed to be Poisson. If the claim-severity distribution is lognormally distributed with $\mu = 5$ and $\sigma = 1$, calculate the number of policies required to attain full credibility for pure premium, with $\alpha = 0.02$ and $k = 0.05$.

Solution The mean and the variance of the claim severity are

$$\mu_X = \exp\left(\mu + \frac{\sigma^2}{2}\right) = \exp(5.5) = 244.6919$$

and

$$\sigma_X^2 = \left[\exp\left(2\mu + \sigma^2\right)\right]\left[\exp(\sigma^2) - 1\right] = 102{,}880.6497.$$

Thus, the coefficient of variation of claim severity is

$$C_X = \frac{\sqrt{102{,}880.6497}}{244.6919} = 1.3108.$$

Now $z_{0.99} = \Phi^{-1}(0.99) = 2.3263$, so that the standard for full credibility for pure premium requires a minimum expected claim frequency of

$$\lambda_F(1 + C_X^2) = \left(\frac{2.3263}{0.05}\right)^2 \left[1 + (1.3108)^2\right] = 5{,}884.2379.$$

Hence, the minimum number of policies for full credibility for pure premium is

$$\frac{5{,}884.2379}{0.03} = 196{,}142.$$

□

6.3 Partial credibility

When the risk group is not sufficiently large, full credibility cannot be attained. In this case, a value of $Z < 1$ has to be determined. Denoting generically the loss measure of interest by W, the basic assumption in deriving Z is that the probability of ZW lying within the interval $[Z\mu_W - k\mu_W, Z\mu_W + k\mu_W]$ is $1-\alpha$ for a given value of k. For the case where the loss measure of interest is the claim frequency N, we require

$$\Pr\left(Z\mu_N - k\mu_N \leq ZN \leq Z\mu_N + k\mu_N\right) = 1 - \alpha, \tag{6.20}$$

which upon standardization becomes

$$\Pr\left(\frac{-k\mu_N}{Z\sigma_N} \leq \frac{N-\mu_N}{\sigma_N} \leq \frac{k\mu_N}{Z\sigma_N}\right) = 1-\alpha. \tag{6.21}$$

Assuming Poisson claim-frequency distribution with mean λ_N and applying the normal approximation, the left-hand side of the above equation reduces to

$$2\Phi\left(\frac{k\sqrt{\lambda_N}}{Z}\right) - 1. \tag{6.22}$$

Thus, we have

$$\frac{k\sqrt{\lambda_N}}{Z} = z_{1-\frac{\alpha}{2}}, \tag{6.23}$$

so that

$$Z = \left(\frac{k}{z_{1-\frac{\alpha}{2}}}\right)\sqrt{\lambda_N} = \sqrt{\frac{\lambda_N}{\lambda_F}}. \tag{6.24}$$

Equation (6.24) is called the **square-root rule for partial credibility**. For predicting claim frequency, the rule states that the **partial-credibility factor** Z is the square root of the ratio of the expected claim frequency to the standard for full credibility for claim frequency. The principle in deriving the partial credibility factor for claim frequency can be applied to other loss measures as well, and similar results are obtained. In general, the partial credibility factor is the square root of the ratio of the size of the risk group (measured in number of exposure units, number of claims or expected number of claims) to the standard for full credibility.

The partial credibility factors for claim severity, aggregate loss, and pure premium are summarized below

$$\textbf{Claim severity: } Z = \sqrt{\frac{N}{\lambda_F C_X^2}}$$

$$\textbf{Aggregate loss/Pure premium: } Z = \sqrt{\frac{\lambda_N}{\lambda_F(1+C_X^2)}}.$$

Example 6.11 A block of insurance policies had 896 claims this period with mean loss of 45 and variance of loss of 5,067. Full credibility is based on a coverage probability of 98% for a range of within 10% deviation from the true mean. The mean frequency of claims is 0.09 per policy and the block has 18,600 policies. Calculate Z for the claim frequency, claim severity, and aggregate loss for the next period.

Solution The expected claim frequency is $\lambda_N = 18{,}600(0.09) = 1{,}674$. We have $z_{0.99} = \Phi^{-1}(0.99) = 2.3263$, so that the full-credibility standard for claim frequency is

$$\lambda_F = \left(\frac{2.3263}{0.1}\right)^2 = 541.17 < 1{,}674 = \lambda_N.$$

Thus, for claim frequency there is full credibility and $Z = 1$. The estimated coefficient of variation for claim severity is

$$C_X = \frac{\sqrt{5{,}067}}{45} = 1.5818,$$

so that the standard for full credibility for claim severity is

$$\lambda_F C_X^2 = (541.17)(1.5818)^2 = 1{,}354.13,$$

which is larger than the sample size 896. Hence, full credibility is not attained for claim severity. The partial credibility factor is

$$Z = \sqrt{\frac{896}{1{,}354.13}} = 0.8134.$$

For aggregate loss, the standard for full credibility is

$$\lambda_F(1 + C_X^2) = 1{,}895.23 > 1{,}674 = \lambda_N.$$

Thus, full credibility is not attained for aggregate loss, and the partial credibility factor is

$$Z = \sqrt{\frac{1{,}674}{1{,}895.23}} = 0.9398.$$

□

6.4 Variation of assumptions

The classical credibility approach relies heavily on the assumption of Poisson distribution for the number of claims. This assumption can be relaxed, however. We illustrate the case where credibility for claim frequency is considered under an alternative assumption.

Assume the claim frequency N is distributed as a binomial random variable with parameters E and θ, i.e. $N \sim \mathcal{BN}(E, \theta)$. Thus, θ is the probability of a claim and E is the number of exposure units. The mean and variance of N are

$\lambda_N = E\theta$ and $\sigma_N^2 = E\theta(1-\theta)$. The standard for full credibility for claim frequency, as given by equation (6.4), is

$$\frac{k\mu_N}{\sigma_N} = \frac{kE\theta}{\sqrt{E\theta(1-\theta)}} = k\sqrt{\frac{E\theta}{1-\theta}} = z_{1-\frac{\alpha}{2}}, \tag{6.25}$$

which reduces to

$$E = \left(\frac{z_{1-\frac{\alpha}{2}}}{k}\right)^2 \left(\frac{1-\theta}{\theta}\right) = \lambda_F \left(\frac{1-\theta}{\theta}\right). \tag{6.26}$$

This can also be expressed in terms of the expected number of claims, which is

$$E\theta = \lambda_F(1-\theta). \tag{6.27}$$

As $\lambda_F(1-\theta) < \lambda_F$, the standard for full credibility under the binomial assumption is less than that under the Poisson assumption. However, as θ is typically small, $1-\theta$ is close to 1 and the difference between the two models is small.

Example 6.12 Assume full credibility is based on 99% coverage of observed claim frequency within 1% of the true mean. Compare the standard for full credibility for claim frequency based on assumptions of Poisson claim frequency versus binomial claim frequency with the probability of claim per policy being 0.05.

Solution From Table 6.1, full-credibility standard for claim frequency requires an expected claim frequency of 66,349, or exposure of 66,349/0.05 = 1.327 million units, if the Poisson assumption is adopted. Under the binomial assumption, the expected claim number from equation (6.27) is 66,349(1 − 0.05) = 63,031.55. In terms of exposure, we require 63,031.55/0.05 = 1.261 million units. □

6.5 Summary and discussions

Table 6.2 summarizes the formulas for various cases of full- and partial-credibility factors for different loss measures.

The above results for the classical credibility approach assume the approximation of the normal distribution for the loss measure of interest. Also, if predictions of claim frequency and aggregate loss are required, the claim frequency is assumed to be Poisson. The coefficient of variation of the claim severity is assumed to be given or reliable estimates are obtainable from the data.

Table 6.2. *Summary of standards for full-credibility and partial-credibility factor Z*

Loss measure	Standard for full credibility	Partial-credibility factor Z
Claim frequency	$\lambda_F = \left(\frac{z_{1-\frac{\alpha}{2}}}{k}\right)^2$	$\sqrt{\frac{\lambda_N}{\lambda_F}}$
Claim severity	$\lambda_F C_X^2$	$\sqrt{\frac{N}{\lambda_F C_X^2}}$
Aggregate loss/Pure premium	$\lambda_F(1 + C_X^2)$	$\sqrt{\frac{\lambda_N}{\lambda_F(1 + C_X^2)}}$

The computation of the full standards depends on the given level of probability coverage $1 - \alpha$ and the accuracy parameter k. As Table 6.1 shows, the full-credibility standard has large variations over different values of α and k, and it may be difficult to determine the suitable values to adopt.

Although the classical credibility approach is easy to apply, it is not based on a well-adopted statistical principle of prediction. In particular, there are several shortcomings of the approach, such as:

1 This approach emphasizes the role of D. It does not attach any importance to the accuracy of the prior information M.
2 The full-credibility standards depend on some unknown parameter values. The approach does not address the issue of how the calibration of these parameters may affect the credibility.
3 There are some limitations in the assumptions, which are made for the purpose of obtaining tractable analytical results.

Exercises

6.1 Assume the claim severity has a mean of 256 and a standard deviation of 532. A sample of 456 claims are observed. Answer the following questions.
(a) What is the probability that the sample mean is within 10% of the true mean?
(b) What is the coefficient of variation of the claim-severity distribution?
(c) What is the coefficient of variation of the sample mean of the claim severity?

(d) Within what percentage of the true mean will the sample mean be observed with a probability of 92%?

(e) What assumptions have you made in answering the above questions?

6.2 Assume the aggregate-loss distribution follows a compound distribution with the claim frequency distributed as a Poisson with mean 569, and the claim severity distributed with mean 120 and standard deviation 78.

(a) Calculate the mean and the variance of the aggregate loss.

(b) Calculate the probability that the observed aggregate loss is within 6% of the mean aggregate loss.

(c) If the mean of the claim frequency increases to 620, how might the claim-severity distribution be changed so that the probability in (b) remains unchanged (give one possible answer)?

(d) If the standard deviation of the claim-severity distribution reduces to 60, how might the claim-frequency distribution be changed so that the probability in (b) remains unchanged?

6.3 A risk group has 569 claims this period, giving a claim average of 1,290 and standard deviation of 878. Calculate the standard for full credibility for claim frequency and claim severity, where full credibility is based on deviation of up to 6% of the true mean with a coverage probability of 94%. You may assume the claim frequency to be Poisson. Is full credibility attained in each case?

6.4 Assume the variance of the claim-frequency distribution is twice its mean (the distribution is not Poisson). Find the standard for full credibility for claim frequency and aggregate loss.

6.5 Assume Poisson distribution for the claim frequency. Show that the partial-credibility factor for claim severity is $\sqrt{N/(\lambda_F C_X^2)}$. The notations are as defined in the text.

6.6 Assume Poisson distribution for the claim frequency. Show that the partial-credibility factor for aggregate loss is $\sqrt{\lambda_N/[\lambda_F(1 + C_X^2)]}$. The notations are as defined in the text.

6.7 The standard for full credibility for claim frequency of a risk group is 2,156. If the standard is based on a coverage probability of 94%, what is the accuracy parameter k?

(a) If the required accuracy parameter is halved, will the standard for full credibility increase or decrease? What is its new value?

(b) For the standard defined in (a), if the standard for full credibility for claim severity is 4,278 claims, what is the standard for full credibility for aggregate loss?

6.8 Claim severity is uniformly distributed in the interval [2000, 3000]. If claim frequency is distributed as a Poisson, determine the standard for full credibility for aggregate loss.

6.9 Assume claim frequency to be Poisson. If claim severity is distributed exponentially with mean 356, find the standard for full credibility for aggregate loss. If the maximum amount of a claim is capped at 500, calculate the revised standard for full credibility for the aggregate loss.

6.10 A block of health insurance policies has 2,309 claims this year, with mean claim of \$239 and standard deviation of \$457. If full credibility is based on 95% coverage to within 5% of the true mean claim severity, and the prior mean severity is \$250, what is the updated prediction for the mean severity next year based on the limited-fluctuation approach?

6.11 Claim severity is distributed lognormally with $\mu = 5$ and $\sigma^2 = 2$. The classical credibility approach is adopted to predict mean severity, with a minimum coverage probability of 92% to within 5% of the mean severity. An average loss of 354 per claim was calculated for 6,950 claims for the current year. What is the predicted loss for each claim? If the average loss of 354 was actually calculated for 9,650 claims, what is the predicted loss for each claim? State any assumptions you have made in the calculation.

6.12 Claim severity has mean 358 and standard deviation 421. An insurance company has 85,000 insurance policies. Using the classical credibility approach with coverage probability of 96% to within 6% of the aggregate loss, determine the credibility factor Z if the average claim per policy is (a) 3%, and (b) 5%.

6.13 Claim frequency has a binomial distribution with the probability of claim per policy being 0.068. Assume full credibility is based on 98% coverage of observed claim frequency within 4% of the true mean. Determine the credibility factor if there are 68,000 policies.

6.14 Claim severity has mean 26. In Year 1, 1,200 claims were filed with mean severity of 32. Based on the limited-fluctuation approach, severity per policy for Year 2 was then revised to 29.82. In Year 2, 1,500 claims were filed with mean severity of 24.46. What is the revised mean severity prediction for Year 3, if the same actuarial assumptions are used as for the prediction for Year 2?

6.15 Claim frequency N has a Poisson distribution, and claim size X is distributed as $\mathcal{P}(6, 0.5)$, where N and X are independent. For full credibility of the pure premium, the observed pure premium is required to be within 2% of the expected pure premium 90% of the time. Determine the expected number of claims required for full credibility.

Questions adapted from SOA exams

6.16 An insurance company has 2,500 policies. The annual amount of claims for each policy follows a compound distribution. The primary distribution is $\mathcal{NB}(2, 1/1.2)$ and the secondary distribution is $\mathcal{P}(3, 1000)$. Full credibility is attained if the observed aggregate loss is within 5% of the expected aggregate loss 90% of the time. Determine the partial credibility of the annual aggregate loss of the company.

6.17 An insurance company has determined that the limited-fluctuation full credibility standard is 2,000 if (a) the total number of claims is to be within 3% of the expected value with probability $1 - \alpha$, and (b) the number of claims follows a Poisson distribution. The standard is then changed so that the total cost of claims is to be within 5% of the expected value with probability $1 - \alpha$, where claim severity is distributed as $\mathcal{U}(0, 10000)$. Determine the expected number of claims for the limited-fluctuation full credibility standard.

6.18 The number of claims is distributed as $\mathcal{NB}(r, 0.25)$, and the claim severity takes values 1, 10, and 100 with probabilities 0.4, 0.4, and 0.2, respectively. If claim frequency and claim severity are independent, determine the expected number of claims needed for the observed aggregate losses to be within 10% of the expected aggregate losses with 95% probability.

7

Bühlmann credibility

While the classical credibility theory addresses the important problem of combining claim experience and prior information to update the prediction for loss, it does not provide a very satisfactory solution. The method is based on arbitrary selection of the coverage probability and the accuracy parameter. Furthermore, for tractability some restrictive assumptions about the loss distribution have to be imposed.

Bühlmann credibility theory sets the problem in a rigorous statistical framework of optimal prediction, using the least mean squared error criterion. It is flexible enough to incorporate various distributional assumptions of loss variables. The approach is further extended to enable the claim experience of different blocks of policies with different exposures to be combined for improved forecast through the Bühlmann–Straub model.

The Bühlmann and Bühlmann–Straub models recognize the interaction of two sources of variability in the data, namely the variation due to between-group differences and variation due to within-group fluctuations. We begin this chapter with the set-up of the Bühlmann credibility model, and a review of how the variance of the loss variable is decomposed into between-group and within-group variations. We derive the Bühlmann credibility factor and updating formula as the minimum mean squared error predictor. The approach is then extended to the Bühlmann–Straub model, in which the loss random variables have different exposures.

Learning objectives

1 Basic framework of Bühlmann credibility
2 Variance decomposition
3 Expected value of the process variance
4 Variance of the hypothetical mean
5 Bühlmann credibility
6 Bühlmann–Straub credibility

7.1 Framework and notations

Consider a risk group or block of insurance policies with loss measure denoted by X, which may be claim frequency, claim severity, aggregate loss, or pure premium. We assume that the risk profiles of the group are characterized by a parameter θ, which determines the distribution of the loss measure X. We denote the conditional mean and variance of X given θ by

$$\mathrm{E}(X \mid \theta) = \mu_X(\theta), \tag{7.1}$$

and

$$\mathrm{Var}(X \mid \theta) = \sigma_X^2(\theta). \tag{7.2}$$

We assume that the insurance company has similar blocks of policies with different risk profiles. Thus, the parameter θ varies with different risk groups. We treat θ as the realization of a random variable Θ, the distribution of which is called the **prior distribution**. When θ varies over the support of Θ, the conditional mean and variance of X become random variables in Θ, and are denoted by $\mu_X(\Theta) = \mathrm{E}(X \mid \Theta)$ and $\sigma_X^2(\Theta) = \mathrm{Var}(X \mid \Theta)$, respectively.

Example 7.1 An insurance company has blocks of worker compensation policies. The claim frequency is known to be Poisson with parameter λ, where λ is 20 for the low-risk group and 50 for the high-risk group. Suppose 30% of the risk groups are low risk and 70% are high risk. What are the conditional mean and variance of the claim frequency?

Solution The parameter determining the claim frequency X is λ, which is a realization of the random variable Λ. As X is Poisson, the conditional mean and conditional variance of X are equal to λ. Thus, we have the results in Table 7.1, so that

$$\mu_X(\Lambda) = \mathrm{E}(X \mid \Lambda) = \begin{cases} 20, & \text{with probability } 0.30, \\ 50, & \text{with probability } 0.70. \end{cases}$$

Likewise, we have

$$\sigma_X^2(\Lambda) = \mathrm{Var}(X \mid \Lambda) = \begin{cases} 20, & \text{with probability } 0.30, \\ 50, & \text{with probability } 0.70. \end{cases}$$

□

Example 7.2 The claim severity X of a block of health insurance policies is normally distributed with mean θ and variance 10. If θ takes values within the interval [100, 200] and follows a uniform distribution, what are the conditional mean and conditional variance of X?

Table 7.1. *Results for Example 7.1*

λ	$\Pr(\Lambda = \lambda)$	$\mathrm{E}(X \mid \lambda)$	$\mathrm{Var}(X \mid \lambda)$
20	0.3	20	20
50	0.7	50	50

Solution The conditional variance of X is 10, irrespective of θ. Hence, we have $\sigma_X^2(\Theta) = \mathrm{Var}(X \mid \Theta) = 10$ with probability 1. The conditional mean of X is Θ, i.e. $\mu_X(\Theta) = \mathrm{E}(X \mid \Theta) = \Theta$, which is uniformly distributed in [100, 200] with pdf

$$f_\Theta(\theta) = \begin{cases} 0.01, & \text{for } \theta \in [100, 200], \\ 0, & \text{otherwise.} \end{cases}$$

□

The Bühlmann model assumes that there are n observations of losses, denoted by $\boldsymbol{X} = \{X_1, \cdots, X_n\}$. The observations may be losses recorded in n periods and they are assumed to be independently and identically distributed as X, which depends on the parameter θ. The task is to update the prediction of X for the next period, i.e. X_{n+1}, based on $\boldsymbol{X}$. In the Bühlmann approach the solution depends on the variation between the conditional means as well as the average of the conditional variances of the risk groups. In the next section we discuss the calculation of these components, after which we will derive the updating formula proposed by Bühlmann.

7.2 Variance components

The variation of the loss measure X consists of two components: the variation between risk groups and the variation within risk groups. The first component, variation between risk groups, is due to the randomness of the risk profiles of each group and is captured by the parameter Θ. The second component, variation within risk group, is measured by the conditional variance of the risk group.[1]

We first consider the calculation of the overall mean of the loss measure X. The **unconditional mean** (or **overall mean**) of X measures the overall central tendency of X, averaged over all the underlying differences in the risk groups. Applying equation (A.111) of iterative expectation, the unconditional mean

[1] Readers may refer to Appendix A.11 for a review of the calculation of conditional expectation and total variance. To economize on notations, we will use X to denote a general loss measure as well as claim severity.

of X is

$$\mathrm{E}(X) = \mathrm{E}[\mathrm{E}(X \mid \Theta)] = \mathrm{E}[\mu_X(\Theta)]. \tag{7.3}$$

Thus, the unconditional mean is the average of the conditional means taken over the distribution of Θ.[2]

For the **unconditional variance** (or **total variance**), the calculation is more involved. The total variance of X is due to the variation in Θ as well as the variance of X conditional on Θ. We use the results derived in Appendix A.11. Applying equation (A.115), we have

$$\mathrm{Var}(X) = \mathrm{E}[\mathrm{Var}(X \mid \Theta)] + \mathrm{Var}[\mathrm{E}(X \mid \Theta)]. \tag{7.4}$$

Note that $\mathrm{Var}(X \mid \Theta)$ measures the variance of a given risk group. It is a function of the random variable Θ and we call this the **process variance**. Thus, $\mathrm{E}[\mathrm{Var}(X \mid \Theta)]$ is the **expected value of the process variance** (**EPV**). On the other hand, $\mathrm{E}(X \mid \Theta)$ is the mean of a given risk group. We call this conditional mean the **hypothetical mean**. Thus, $\mathrm{Var}[\mathrm{E}(X \mid \Theta)]$ is the **variance of the hypothetical means** (**VHM**), as it measures the variations in the *means* of the risk groups. Verbally, equation (7.4) can be written as

$$\begin{aligned}\textbf{total variance} &= \textbf{expected value of process variance} \\ &\quad + \textbf{variance of hypothetical means},\end{aligned} \tag{7.5}$$

or

$$\textbf{total variance} = \textbf{EPV} + \textbf{VHM}. \tag{7.6}$$

It can also be stated alternatively as

$$\begin{aligned}\textbf{total variance} &= \textbf{mean of conditional variance} \\ &\quad + \textbf{variance of conditional mean}.\end{aligned} \tag{7.7}$$

Symbolically, we use the following notations

$$\mathrm{E}[\mathrm{Var}(X \mid \Theta)] = \mathrm{E}[\sigma_X^2(\Theta)] = \mu_{\mathrm{PV}}, \tag{7.8}$$

and

$$\mathrm{Var}[\mathrm{E}(X \mid \Theta)] = \mathrm{Var}[\mu_X(\Theta)] = \sigma_{\mathrm{HM}}^2, \tag{7.9}$$

[2] Note that the expectation operations in equation (7.3) have different meanings. The operation in $\mathrm{E}(X)$ is taken unconditionally on X. The first (outer) operation in $\mathrm{E}[\mathrm{E}(X \mid \Theta)]$ is taken over Θ, while the second (inner) operation is taken over X *conditional* on Θ. Lastly, the operation in $\mathrm{E}[\mu_X(\Theta)]$ is taken over Θ unconditionally.

so that equation (7.4) can be written as

$$\mathrm{Var}(X) = \mu_{\mathrm{PV}} + \sigma^2_{\mathrm{HM}}. \tag{7.10}$$

Example 7.3 For Examples 7.1 and 7.2, calculate the unconditional mean, the expected value of the process variance, the variance of the hypothetical means, and the total variance.

Solution For Example 7.1, the unconditional mean is

$$\begin{aligned}\mathrm{E}(X) &= \Pr(\Lambda = 20)\mathrm{E}(X \mid \Lambda = 20) + \Pr(\Lambda = 50)\mathrm{E}(X \mid \Lambda = 50)\\ &= (0.3)(20) + (0.7)(50)\\ &= 41.\end{aligned}$$

The expected value of the process variance, EPV, is

$$\begin{aligned}\mathrm{E}[\mathrm{Var}(X \mid \Lambda)] &= \Pr(\Lambda = 20)\mathrm{Var}(X \mid \Lambda = 20) + \Pr(\Lambda = 50)\mathrm{Var}(X \mid \Lambda = 50)\\ &= (0.3)(20) + (0.7)(50)\\ &= 41.\end{aligned}$$

As the mean of the hypothetical means (i.e. the unconditional mean) is 41, the variance of the hypothetical means, VHM, is

$$\mathrm{Var}[\mathrm{E}(X \mid \Lambda)] = (0.3)(20-41)^2 + (0.7)(50-41)^2 = 189.$$

Thus, the total variance of X is

$$\mathrm{Var}(X) = \mathrm{E}[\mathrm{Var}(X \mid \Lambda)] + \mathrm{Var}[\mathrm{E}(X \mid \Lambda)] = 41 + 189 = 230.$$

For Example 7.2, as Θ is uniformly distributed in $[100, 200]$, the unconditional mean of X is

$$\mathrm{E}(X) = \mathrm{E}[\mathrm{E}(X \mid \Theta)] = \mathrm{E}(\Theta) = 150.$$

As X has a constant variance of 10, the expected value of the process variance is

$$\mathrm{E}[\mathrm{Var}(X \mid \Theta)] = \mathrm{E}(10) = 10.$$

The variance of the hypothetical means is[3]

$$\mathrm{Var}[\mathrm{E}(X \mid \Theta)] = \mathrm{Var}(\Theta) = \frac{(200-100)^2}{12} = 833.33,$$

[3] See Appendix A.10.3 for the variance of the uniform distribution.

and the total variance of X is

$$\mathrm{Var}(X) = 10 + 833.33 = 843.33.$$

□

If we divide the variance of the hypothetical means by the total variance, we obtain the proportion of the variation in X that is due to the differences in the means of the risk groups. Thus, for Example 7.1, we have

$$\frac{\sigma^2_{\mathrm{HM}}}{\mathrm{Var}(X)} = \frac{189}{230} = 82.17\%,$$

so that 82.17% of the variation in a randomly observed X is due to the differences in the averages of the risk groups. For Example 7.2, this figure is $833.33/843.33 = 98.81\%$.

Example 7.4 The claim severity X of a block of health insurance policies is normally distributed with mean 100 and variance σ^2. If σ^2 takes values within the interval [50, 100] and follows a uniform distribution, find the conditional mean of claim severity, the expected value of the process variance, the variance of the hypothetical means, and the total variance.

Solution We denote the random variable of the variance of X by Ω . Note that the conditional mean of X does not vary with Ω, and we have $\mathrm{E}(X \mid \Omega) = 100$, so that the unconditional mean of X is

$$\mathrm{E}(X) = \mathrm{E}[\mathrm{E}(X \mid \Omega)] = \mathrm{E}(100) = 100.$$

As Ω is uniformly distributed in [50, 100], the expected value of the process variance is

$$\mathrm{EPV} = \mu_{\mathrm{PV}} = \mathrm{E}[\mathrm{Var}(X \mid \Omega)] = \mathrm{E}(\Omega) \ = 75.$$

For the variance of the hypothetical means, we have

$$\mathrm{VHM} = \sigma^2_{\mathrm{HM}} = \mathrm{Var}[\mathrm{E}(X \mid \Omega)] = \mathrm{Var}(100) = 0.$$

Thus, the total variance of X is 75, which is *entirely* due to the process variance, as there is no variation in the conditional mean. □

Example 7.5 An insurance company sells workers compensation policies, each of which belongs to one of three possible risk groups. The risk groups have claim frequencies N that are Poisson distributed with parameter λ and claim severity X that are gamma distributed with parameters α and β. Claim frequency and claim

Table 7.2. *Data for Example 7.5*

Risk group	Relative frequency	Distribution of N: $\mathcal{PN}(\lambda)$	Distribution of X: $\mathcal{G}(\alpha, \beta)$
1	0.2	$\lambda = 20$	$\alpha = 5, \beta = 2$
2	0.4	$\lambda = 30$	$\alpha = 4, \beta = 3$
3	0.4	$\lambda = 40$	$\alpha = 3, \beta = 2$

severity are independently distributed given a risk group, and the aggregate loss is S. The data of the risk groups are given in Table 7.2. For each of the following loss measures: (a) claim frequency N, (b) claim severity X, and (c) aggregate loss S, calculate EPV, VHM, and the total variance.

Solution (a) Claim frequency We first calculate the conditional mean and conditional variance of N given the risk group, which is characterized by the parameter Λ. As N is Poisson, the mean and variance are equal to Λ, so that we have the results in Table 7.3.

Table 7.3. *Results for Example 7.5 (a)*

Risk group	Probability	$\mathrm{E}(N \mid \Lambda) = \mu_N(\Lambda)$	$\mathrm{Var}(N \mid \Lambda) = \sigma_N^2(\Lambda)$
1	0.2	20	20
2	0.4	30	30
3	0.4	40	40

Thus, the EPV is

$$\mu_{\mathrm{PV}} = \mathrm{E}[\mathrm{Var}(N \mid \Lambda)] = (0.2)(20) + (0.4)(30) + (0.4)(40) = 32,$$

which is also equal to the unconditional mean $\mathrm{E}[\mu_N(\Lambda)]$. For VHM, we first calculate

$$\mathrm{E}\{[\mu_N(\Lambda)]^2\} = (0.2)(20)^2 + (0.4)(30)^2 + (0.4)(40)^2 = 1{,}080,$$

so that

$$\sigma_{\mathrm{HM}}^2 = \mathrm{Var}[\mu_N(\Lambda)] = \mathrm{E}\{[\mu_N(\Lambda)]^2\} - \{\mathrm{E}[\mu_N(\Lambda)]\}^2 = 1{,}080 - (32)^2 = 56.$$

Therefore, the total variance of N is

$$\mathrm{Var}(N) = \mu_{\mathrm{PV}} + \sigma^2_{\mathrm{HM}} = 32 + 56 = 88.$$

(b) Claim severity There are three claim-severity distributions, which are specific to each risk group. Note that the relative frequencies of the risk groups as well as the claim frequencies in the risk groups jointly determine the relative occurrence of each claim-severity distribution. The probabilities of occurrence of the severity distributions, as well as their conditional means and variances are given in Table 7.4, in which Γ denotes the vector random variable representing α and β.

Table 7.4. *Results for Example 7.5 (b)*

Group	Group probability	λ	Col 2 $\times$ Col 3	Probability of severity X	$\mathrm{E}(X \mid \Gamma) = \mu_X(\Gamma)$	$\mathrm{Var}(X \mid \Gamma) = \sigma^2_X(\Gamma)$
1	0.2	20	4	0.125	10	20
2	0.4	30	12	0.375	12	36
3	0.4	40	16	0.500	6	12

Column 4 gives the expected number of claims in each group weighted by the group probability. Column 5 gives the probability of occurrence of each type of claim-severity distribution, which is obtained by dividing the corresponding figure in Column 4 by the sum of Column 4 (e.g. 0.125 = 4/(4 + 12 + 16)). The last two columns give the conditional mean $\alpha\beta$ and conditional variance $\alpha\beta^2$ corresponding to the three different distributions of claim severity. Similar to the calculation in (a), we have

$$\mathrm{E}(X) = \mathrm{E}[\mathrm{E}(X \mid \Gamma)] = (0.125)(10) + (0.375)(12) + (0.5)(6) = 8.75,$$

and

$$\mu_{\mathrm{PV}} = (0.125)(20) + (0.375)(36) + (0.5)(12) = 22.$$

To calculate VHM, we first compute the raw second moment of the conditional mean of X, which is

$$\mathrm{E}\{[\mu_X(\Gamma)]^2\} = (0.125)(10)^2 + (0.375)(12)^2 + (0.5)(6)^2 = 84.50.$$

Hence

$$\sigma^2_{\text{HM}} = \text{Var}[\mu_X(\Gamma)] = \text{E}\{[\mu_X(\Gamma)]^2\} - \{\text{E}[\mu_X(\Gamma)]\}^2$$
$$= 84.50 - (8.75)^2 = 7.9375.$$

Therefore, the total variance of X is

$$\text{Var}(X) = \mu_{\text{PV}} + \sigma^2_{\text{HM}} = 22 + 7.9375 = 29.9375.$$

(c) Aggregate loss The distribution of the aggregate loss S is determined jointly by Λ and Γ, which we shall denote as Θ. For the conditional mean of S, we have

$$\text{E}(S \mid \Theta) = \text{E}(N \mid \Theta)\text{E}(X \mid \Theta) = \lambda\alpha\beta.$$

For the conditional variance of S, we use the result on compound distribution with Poisson claim frequency stated in equation (A.123), and make use of the assumption of gamma severity to obtain

$$\text{Var}(S \mid \Theta) = \lambda[\sigma^2_X(\Gamma) + \mu^2_X(\Gamma)] = \lambda(\alpha\beta^2 + \alpha^2\beta^2).$$

The conditional means and conditional variances of S are summarized in Table 7.5.

Table 7.5. *Results for Example 7.5 (c)*

Group	Group probability	Parameters λ, α, β	$\text{E}(S \mid \Theta) = \mu_S(\Theta)$	$\text{Var}(S \mid \Theta) = \sigma^2_S(\Theta)$
1	0.2	20, 5, 2	200	2,400
2	0.4	30, 4, 3	360	5,400
3	0.4	40, 3, 2	240	1,920

The unconditional mean of S is

$$\text{E}(S) = \text{E}[\text{E}(S \mid \Theta)] = (0.2)(200) + (0.4)(360) + (0.4)(240) = 280,$$

and the EPV is

$$\mu_{\text{PV}} = (0.2)(2{,}400) + (0.4)(5{,}400) + (0.4)(1{,}920) = 3{,}408.$$

Also, the VHM is given by

$$
\begin{aligned}
\sigma^2_{\mathrm{HM}} &= \mathrm{Var}[\mu_S(\Theta)] \\
&= \mathrm{E}\{[\mu_S(\Theta)]^2\} - \{\mathrm{E}[\mu_S(\Theta)]\}^2 \\
&= \left[(0.2)(200)^2 + (0.4)(360)^2 + (0.4)(240)^2\right] - (280)^2 \\
&= 4{,}480.
\end{aligned}
$$

Therefore, the total variance of S is

$$\mathrm{Var}(S) = 3{,}408 + 4{,}480 = 7{,}888.$$

□

EPV and VHM measure two different aspects of the total variance. When a risk group is homogeneous so that the loss claims are similar within the group, the conditional variance is small. If all risk groups have similar loss claims within the group, the expected value of the process variance EPV is small. On the other hand, if the risk groups have very different risk profiles across groups, their hypothetical means will differ more and thus the variance of the hypothetical means VHM will be large. In other words, it will be easier to distinguish between risk groups if the variance of the hypothetical means is large and the average of the process variance is small.

We define k as the ratio of EPV to VHM, i.e.

$$k = \frac{\mu_{\mathrm{PV}}}{\sigma^2_{\mathrm{HM}}} = \frac{\mathrm{EPV}}{\mathrm{VHM}}. \tag{7.11}$$

A small EPV or large VHM will give rise to a small k. The risk groups will be more *distinguishable* in the mean when k is smaller, in which case we may put more weight on the data in updating our revised prediction for future losses. For the cases in Example 7.5, the values of k for claim frequency, claim severity, and aggregate loss are, respectively, 0.5714, 2.7717, and 0.7607. For Example 7.4, as $\sigma^2_{\mathrm{HM}} = 0$, k is infinite.[4]

Example 7.6 Frequency of claim per year, N, is distributed as a binomial random variable $\mathcal{BN}(10, \theta)$, and claim severity, X, is distributed as an exponential random variable with mean $c\theta$, where c is a known constant. Given θ, claim frequency and claim severity are independently distributed. Derive an expression of k for the aggregate loss per year, S, in terms of c and the moments

[4] In this case, the data contain no useful information for updating the *mean* of the risk group separately from the overall mean, although they might be used to update the specific group *variance* if required.

of Θ, and show that it does not depend on c. If Θ is 0.3 or 0.7 with equal probabilities, calculate k.

Solution We first calculate the conditional mean of S as a function of θ. Due to the independence assumption of N and X, the hypothetical mean of S is

$$\mathrm{E}(S \mid \Theta) = \mathrm{E}(N \mid \Theta)\mathrm{E}(X \mid \Theta) = (10\Theta)(c\Theta) = 10c\Theta^2.$$

Using equation (A.122), the process variance is

$$\begin{aligned}
\mathrm{Var}(S \mid \Theta) &= \mu_N(\Theta)\sigma_X^2(\Theta) + \sigma_N^2(\Theta)\mu_X^2(\Theta) \\
&= (10\Theta)(c\Theta)^2 + [10\Theta(1-\Theta)](c\Theta)^2 \\
&= 10c^2\Theta^3 + 10c^2\Theta^3(1-\Theta) \\
&= 10c^2\Theta^3(2-\Theta).
\end{aligned}$$

Hence, the unconditional mean of S is

$$\mathrm{E}(S) = \mathrm{E}[\mathrm{E}(S \mid \Theta)] = \mathrm{E}(10c\Theta^2) = 10c\mathrm{E}(\Theta^2)$$

and the variance of the hypothetical means is

$$\begin{aligned}
\sigma_{\mathrm{HM}}^2 &= \mathrm{Var}[\mathrm{E}(S \mid \Theta)] \\
&= \mathrm{Var}(10c\Theta^2) \\
&= 100c^2\mathrm{Var}(\Theta^2) \\
&= 100c^2\{\mathrm{E}(\Theta^4) - [\mathrm{E}(\Theta^2)]^2\}.
\end{aligned}$$

The expected value of the process variance is

$$\begin{aligned}
\mu_{\mathrm{PV}} &= \mathrm{E}[\mathrm{Var}(S \mid \Theta)] \\
&= \mathrm{E}[10c^2\Theta^3(2-\Theta)] \\
&= 10c^2[2\mathrm{E}(\Theta^3) - \mathrm{E}(\Theta^4)].
\end{aligned}$$

Combining the above results we conclude that

$$k = \frac{\mu_{\mathrm{PV}}}{\sigma_{\mathrm{HM}}^2} = \frac{10c^2[2\mathrm{E}(\Theta^3) - \mathrm{E}(\Theta^4)]}{100c^2\{\mathrm{E}(\Theta^4) - [\mathrm{E}(\Theta^2)]^2\}} = \frac{2\mathrm{E}(\Theta^3) - \mathrm{E}(\Theta^4)}{10\{\mathrm{E}(\Theta^4) - [\mathrm{E}(\Theta^2)]^2\}}.$$

Thus, k does not depend on c. To compute its value for the given distribution of Θ, we present the calculations in Table 7.6.

Table 7.6. *Calculations of Example 7.6*

θ	$\Pr(\Theta = \theta)$	θ^2	θ^3	θ^4
0.3	0.5	0.09	0.027	0.0081
0.7	0.5	0.49	0.343	0.2401

Thus, the required moments of Θ are

$$\mathrm{E}(\Theta) = (0.5)(0.3) + (0.5)(0.7) = 0.5,$$

$$\mathrm{E}(\Theta^2) = (0.5)(0.09) + (0.5)(0.49) = 0.29,$$

$$\mathrm{E}(\Theta^3) = (0.5)(0.027) + (0.5)(0.343) = 0.185$$

and

$$\mathrm{E}(\Theta^4) = (0.5)(0.0081) + (0.5)(0.2401) = 0.1241,$$

so that

$$k = \frac{2(0.185) - 0.1241}{10\left[0.1241 - (0.29)^2\right]} = 0.6148.$$

In this example, note that both EPV and VHM depend on c. However, as the effects of c on these components are the same, the ratio of EPV to VHM is invariant to c. Also, though X and N are independent *given* θ, they are correlated *unconditionally* due to their common dependence on Θ. □

7.3 Bühlmann credibility

Bühlmann's approach of updating the predicted loss measure is based on a linear predictor using past observations. It is also called the **greatest accuracy approach** or the **least squares approach**. Recall that for the classical credibility approach, the updated prediction U is given by (see equation (6.1))

$$U = ZD + (1 - Z)M. \qquad (7.12)$$

The Bühlmann credibility method has a similar basic equation, in which D is the sample mean of the data and M is the overall prior mean $\mathrm{E}(X)$. The Bühlmann credibility factor Z depends on the sample size n and the EPV to VHM ratio k. In particular, Z varies with n and k as follows:

1 Z increases with the sample size n of the data.
2 Z increases with the *distinctiveness* of the risk groups. As argued above, the risk groups are more distinguishable when k is small. Thus, Z increases as k decreases.

We now state formally the assumptions of the Bühlmann model and derive the updating formula as the **least mean squared error (MSE) linear predictor**.

1 $\boldsymbol{X} = \{X_1, \cdots, X_n\}$ are loss measures that are independently and identically distributed as the random variable X. The distribution of X depends on the parameter θ.
2 The parameter θ is a realization of a random variable Θ. Given θ, the conditional mean and variance of X are

$$\mathrm{E}(X \mid \theta) = \mu_X(\theta), \tag{7.13}$$

and

$$\mathrm{Var}(X \mid \theta) = \sigma_X^2(\theta). \tag{7.14}$$

3 The unconditional mean of X is $\mathrm{E}(X) = \mathrm{E}[\mathrm{E}(X \mid \Theta)] = \mu_X$. The mean of the conditional variance of X is

$$\begin{aligned}
\mathrm{E}[\mathrm{Var}(X \mid \Theta)] &= \mathrm{E}[\sigma_X^2(\Theta)] \\
&= \mu_{\mathrm{PV}} \\
&= \text{Expected value of process variance} \\
&= \text{EPV},
\end{aligned} \tag{7.15}$$

and the variance of the conditional mean is

$$\begin{aligned}
\mathrm{Var}[\mathrm{E}(X \mid \Theta)] &= \mathrm{Var}[\mu_X(\Theta)] \\
&= \sigma_{\mathrm{HM}}^2 \\
&= \text{Variance of hypothetical means} \\
&= \text{VHM}.
\end{aligned} \tag{7.16}$$

The unconditional variance (or total variance) of X is

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}[\mathrm{Var}(X \mid \Theta)] + \mathrm{Var}[\mathrm{E}(X \mid \Theta)] \\ &= \mu_{\mathrm{PV}} + \sigma^2_{\mathrm{HM}} \\ &= \mathrm{EPV} + \mathrm{VHM}. \end{aligned} \tag{7.17}$$

4 The Bühlmann approach formulates a predictor of X_{n+1} based on a linear function of $\boldsymbol{X}$, where X_{n+1} is assumed to have the same distribution as X. The predictor minimizes the mean squared error in predicting X_{n+1} over the joint distribution of Θ, X_{n+1}, and $\boldsymbol{X}$. Specifically, the predictor is given by

$$\hat{X}_{n+1} = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n, \tag{7.18}$$

where $\beta_0, \beta_1, \cdots, \beta_n$ are chosen to minimize the mean squared error, MSE, defined as

$$\mathrm{MSE} = \mathrm{E}[(X_{n+1} - \hat{X}_{n+1})^2]. \tag{7.19}$$

To solve the above problem we make use of the least squares regression results in Appendix A.17. We define $\boldsymbol{W}$ as the $(n+1) \times 1$ vector $(1, \boldsymbol{X}')'$, and $\boldsymbol{\beta}$ as the $(n+1) \times 1$ vector $(\beta_0, \beta_1, \cdots, \beta_n)'$. We also write $\boldsymbol{\beta}_S$ as $(\beta_1, \cdots, \beta_n)'$. Thus, the predictor $\hat{X}_{n+1}$ can be written as

$$\hat{X}_{n+1} = \boldsymbol{\beta}'\boldsymbol{W} = \beta_0 + \boldsymbol{\beta}_S'\boldsymbol{X}. \tag{7.20}$$

The MSE is then given by

$$\begin{aligned}\mathrm{MSE} &= \mathrm{E}[(X_{n+1} - \hat{X}_{n+1})^2] \\ &= \mathrm{E}[\left(X_{n+1} - \boldsymbol{\beta}'\boldsymbol{W}\right)^2] \\ &= \mathrm{E}(X_{n+1}^2 + \boldsymbol{\beta}'\boldsymbol{W}\boldsymbol{W}'\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{W}X_{n+1}) \\ &= \mathrm{E}(X_{n+1}^2) + \boldsymbol{\beta}'\mathrm{E}(\boldsymbol{W}\boldsymbol{W}')\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathrm{E}(\boldsymbol{W}X_{n+1}). \end{aligned} \tag{7.21}$$

Thus, the MSE has the same form as RSS in equation (A.167), with the sample moments replaced by the population moments. Hence, the solution of $\boldsymbol{\beta}$ that minimizes MSE is, by virtue of equation (A.168)

$$\hat{\boldsymbol{\beta}} = \left[\mathrm{E}(\boldsymbol{W}\boldsymbol{W}')\right]^{-1} \mathrm{E}(\boldsymbol{W}X_{n+1}). \tag{7.22}$$

Following the results in equations (A.174) and (A.175), we have

$$\hat{\boldsymbol{\beta}}_S = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix}$$

$$= \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_1, X_2) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{Cov}(X_1, X_n) & \mathrm{Cov}(X_2, X_n) & \cdots & \mathrm{Var}(X_n) \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{Cov}(X_1, X_{n+1}) \\ \mathrm{Cov}(X_2, X_{n+1}) \\ \vdots \\ \mathrm{Cov}(X_n, X_{n+1}) \end{bmatrix} \tag{7.23}$$

and

$$\hat{\beta}_0 = \mathrm{E}(X_{n+1}) - \sum_{i=1}^{n} \hat{\beta}_i \mathrm{E}(X_i) = \mu_X - \mu_X \sum_{i=1}^{n} \hat{\beta}_i. \tag{7.24}$$

From equation (7.17), we have

$$\mathrm{Var}(X_i) = \mu_{\mathrm{PV}} + \sigma^2_{\mathrm{HM}}, \qquad \text{for } i = 1, \cdots, n. \tag{7.25}$$

Also, $\mathrm{Cov}(X_i, X_j)$ is given by (for $i \neq j$)

$$\begin{aligned} \mathrm{Cov}(X_i, X_j) &= \mathrm{E}(X_i X_j) - \mathrm{E}(X_i)\mathrm{E}(X_j) \\ &= \mathrm{E}\left[\mathrm{E}(X_i X_j \mid \Theta)\right] - \mu_X^2 \\ &= \mathrm{E}\left[\mathrm{E}(X_i \mid \Theta)\mathrm{E}(X_j \mid \Theta)\right] - \mu_X^2 \\ &= \mathrm{E}\left\{[\mu_X(\Theta)]^2\right\} - \{\mathrm{E}[\mu_X(\Theta)]\}^2 \\ &= \mathrm{Var}[\mu_X(\Theta)] \\ &= \sigma^2_{\mathrm{HM}}. \end{aligned} \tag{7.26}$$

Thus, equation (7.23) can be written as

$$\hat{\boldsymbol{\beta}}_S = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} = \left(\mu_{\mathrm{PV}} \mathbf{I} + \sigma^2_{\mathrm{HM}} \mathbf{1}\mathbf{1}'\right)^{-1} \left(\sigma^2_{\mathrm{HM}} \mathbf{1}\right), \tag{7.27}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix and $\mathbf{1}$ is the $n \times 1$ vector of ones. We now write

$$k = \frac{\mu_{\mathrm{PV}}}{\sigma^2_{\mathrm{HM}}}, \tag{7.28}$$

and evaluate the inverse matrix on the right-hand side of equation (7.27) as follows

$$\begin{aligned}\left(\mu_{\mathrm{PV}}\mathbf{I} + \sigma^2_{\mathrm{HM}}\mathbf{1}\mathbf{1}'\right)^{-1} &= \frac{1}{\mu_{\mathrm{PV}}}\left(\mathbf{I} + \frac{\sigma^2_{\mathrm{HM}}}{\mu_{\mathrm{PV}}}\mathbf{1}\mathbf{1}'\right)^{-1} \\ &= \frac{1}{\mu_{\mathrm{PV}}}\left(\mathbf{I} + \frac{1}{k}\mathbf{1}\mathbf{1}'\right)^{-1}. \end{aligned} \tag{7.29}$$

With $\mathbf{1}'\mathbf{1} = n$, it is easy to verify that

$$\left(\mathbf{I} + \frac{1}{k}\mathbf{1}\mathbf{1}'\right)^{-1} = \mathbf{I} - \frac{1}{n+k}\mathbf{1}\mathbf{1}'. \tag{7.30}$$

Substituting equation (7.30) into equations (7.27) and (7.29), we obtain

$$\begin{aligned}\hat{\boldsymbol{\beta}}_S &= \frac{1}{\mu_{\mathrm{PV}}}\left(\mathbf{I} - \frac{1}{n+k}\mathbf{1}\mathbf{1}'\right)\left(\sigma^2_{\mathrm{HM}}\mathbf{1}\right) \\ &= \frac{1}{k}\left(\mathbf{I} - \frac{1}{n+k}\mathbf{1}\mathbf{1}'\right)\mathbf{1} \\ &= \frac{1}{k}\left(\mathbf{1} - \frac{n}{n+k}\mathbf{1}\right) \\ &= \frac{1}{n+k}\mathbf{1}. \end{aligned} \tag{7.31}$$

Note that equation (7.24) can be written as

$$\hat{\beta}_0 = \mu_X - \mu_X \hat{\boldsymbol{\beta}}'_S\mathbf{1}. \tag{7.32}$$

Thus, the least MSE linear predictor of X_{n+1} is

$$\begin{aligned}\hat{\beta}_0 + \hat{\boldsymbol{\beta}}'_S\boldsymbol{X} &= (\mu_X - \mu_X\hat{\boldsymbol{\beta}}'_S\mathbf{1}) + \hat{\boldsymbol{\beta}}'_S\boldsymbol{X} \\ &= \left(\mu_X - \frac{\mu_X}{n+k}\mathbf{1}'\mathbf{1}\right) + \frac{1}{n+k}\mathbf{1}'\boldsymbol{X} \\ &= \frac{1}{n+k}\mathbf{1}'\boldsymbol{X} + \frac{k\mu_X}{n+k}. \end{aligned} \tag{7.33}$$

Noting that $\mathbf{1}'\boldsymbol{X} = n\bar{X}$, we conclude that

$$\hat{X}_{n+1} = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_S'\boldsymbol{X} = \frac{n\bar{X}}{n+k} + \frac{k\mu_X}{n+k} = Z\bar{X} + (1-Z)\mu_X, \qquad (7.34)$$

where

$$Z = \frac{n}{n+k}. \qquad (7.35)$$

Z defined in equation (7.35) is called the **Bühlmann credibility factor** or simply the **Bühlmann credibility**. It depends on the EPV to VHM ratio k, which is called the **Bühlmann credibility parameter**. The optimal linear forecast $\hat{X}_{n+1}$ given in equation (7.34) is also called the **Bühlmann premium**. Note that k depends only on the parameters of the model,[5] while Z is a function of k and the size n of the *data*. For predicting claim frequency N, the sample size n is the number of periods over which the number of claims is aggregated.[6] For predicting claim severity X, the sample size n is the number of claims. As aggregate loss S refers to the total loss payout per period, the sample size is the number of periods of claim experience.

Example 7.7 Refer to Example 7.5. Suppose the claim experience last year was 26 claims with an average claim size of 12. Calculate the updated prediction of (a) the claim frequency, (b) the average claim size, and (c) the aggregate loss, for next year.

Solution (a) Claim frequency From Example 7.5, we have $k = 0.5714$ and $M = \mathrm{E}(N) = 32$. Now we are given $n = 1$ and $D = 26$. Hence

$$Z = \frac{1}{1 + 0.5714} = 0.6364,$$

so that the updated prediction of the claim frequency of this group is

$$U = (0.6364)(26) + (1 - 0.6364)(32) = 28.1816.$$

(b) Claim severity We have $k = 2.7717$ and $M = \mathrm{E}(X) = 8.75$, with $n = 26$ and $D = 12$. Thus

$$Z = \frac{26}{26 + 2.7717} = 0.9037,$$

[5] Hence, k is fixed, but needs to be estimated in practice. We shall come back to this issue in Chapter 9.

[6] Note that N is the number of claims per period, say year, and n is the number of periods of claim-frequency experience.

so that the updated prediction of the claim severity of this group is

$$U = (0.9037)(12) + (1 - 0.9037)(8.75) = 11.6870.$$

(c) Aggregate loss With $k = 0.7607$, $M = \mathrm{E}(S) = 280$, $n = 1$ and $D = (26)(12) = 312$, we have

$$Z = \frac{1}{1 + 0.7607} = 0.5680,$$

so that the updated prediction of the aggregate loss of this group is

$$U = (0.5680)(312) + (1 - 0.5680)(280) = 298.1760.$$

□

Example 7.8 Refer to Example 7.6. Suppose the numbers of claims in the last three years were 8, 4, and 7, with the corresponding average amount of claim in each of the three years being 12, 19, and 9. Calculate the updated prediction of the aggregate loss for next year for $c = 20$ and 30.

Solution We first calculate the average aggregate loss per year in the last three years, which is

$$\frac{1}{3}\left[(8)(12) + (4)(19) + (7)(9)\right] = 78.3333.$$

As shown in Example 7.6, $k = 0.6148$, which does not vary with c. As there are three observations of S, the Bühlmann credibility factor is

$$Z = \frac{3}{3 + 0.6148} = 0.8299.$$

The unconditional mean of S is

$$\mathrm{E}[\mathrm{E}(S \mid \Theta)] = 10c\mathrm{E}(\Theta^2) = (10)(0.29)c = 2.9c.$$

Hence, using equation (7.34), the updated prediction of S is

$$(0.8299)(78.3333) + (1 - 0.8299)(2.9c),$$

which gives a predicted value of 74.8746 when $c = 20$, and 79.8075 when $c = 30$. □

We have derived the Bühlmann credibility predictor for future loss as the linear predictor (in $\boldsymbol{X}$) that minimizes the mean squared prediction error in equation (7.19). However, we can also consider the problem of a linear *estimator* (in $\boldsymbol{X}$) that minimizes the squared error in estimating the *expected* future loss, i.e. $\mathrm{E}\left[(\mu_{n+1} - \hat{\mu}_{n+1})^2\right]$, where $\mu_{n+1} = \mathrm{E}(X_{n+1}) = \mu_X$ and $\hat{\mu}_{n+1}$ is a linear estimator of μ_{n+1}. Readers are invited to show that the result is the same as the Bühlmann credibility predictor for future loss (see Exercise 7.5). Thus, we shall use the terminologies Bühlmann credibility predictor for future loss and Bühlmann credibility estimator of the expected loss interchangeably.

7.4 Bühlmann–Straub credibility

An important limitation of the Bühlmann credibility theory is that the loss observations X_i are assumed to be *identically* distributed. This assumption is violated if the data are over different periods with different exposures (the definition of exposure will be explained below). The **Bühlmann–Straub credibility model** extends the **Bühlmann theory** to cases where the loss data X_i are not identically distributed. In particular, the process variance of the loss measure is assumed to depend on the exposure. We denote the exposure by m_i, and the *loss per unit of exposure* by X_i. Note that the exposure needs not be the number of insureds, although that may often be the case. We then assume the following for the conditional variance of X_i

$$\mathrm{Var}(X_i \mid \Theta) = \frac{\sigma_X^2(\Theta)}{m_i}, \tag{7.36}$$

for a suitably defined $\sigma_X^2(\Theta)$. The following are some examples:

1 X_i is the average number of claims per insured in year i, $\sigma_X^2(\Theta)$ is the variance of the claim frequency of an insured, and the exposure m_i is the number of insureds covered in year i.[7]
2 X_i is the average aggregate loss per month of the ith block of policies, $\sigma_X^2(\Theta)$ is the variance of the aggregate loss of the block in a month, and the exposure m_i is the number of months of insurance claims for the ith block of policies.
3 X_i is the average loss per unit premium in year i, $\sigma_X^2(\Theta)$ is the variance of the claim amount of an insured per year divided by the premium per insured, and the exposure m_i is the amount of premiums received in year i. To see this, assume there are N_i insured in year i, each paying a premium P. Thus,

[7] Note that $\mathrm{Var}(X_i \mid \Theta)$ is the variance of the *average claim frequency per insured*, while $\sigma_X^2(\Theta)$ is the variance of the *claim frequency of each insured*.

$m_i = N_i P$ and

$$\begin{aligned}
\mathrm{Var}(X_i \mid \Theta) &= \mathrm{Var}\left(\frac{\text{loss per insured}}{P}\right) \\
&= \frac{1}{P^2}\left[\frac{\text{Var(claim amount of an insured)}}{N_i}\right] \\
&= \frac{1}{m_i}\left[\frac{\text{Var(claim amount of an insured)}}{P}\right] \\
&= \frac{\sigma_X^2(\Theta)}{m_i}.
\end{aligned} \tag{7.37}$$

In each of the examples above, the distributions of X_i are not identical. Instead, the conditional variance of X_i varies with m_i. We now formally summarize below the assumptions of the Bühlmann–Straub credibility model.

1 Let m_i be the exposure in period i and X_i be the loss per unit exposure, for $i = 1, \cdots, n$. Suppose $\boldsymbol{X} = \{X_1, \cdots, X_n\}$ are independently (but not identically) distributed and the distribution of X_i depends on the parameter θ.

2 The parameter θ is a realization of a random variable Θ. Given θ, the conditional mean and variance of $\boldsymbol{X}$ are

$$\mathrm{E}(X_i \mid \theta) = \mu_X(\theta), \tag{7.38}$$

and

$$\mathrm{Var}(X_i \mid \theta) = \frac{\sigma_X^2(\theta)}{m_i}, \tag{7.39}$$

for $i \in \{1, \cdots, n\}$, where $\sigma_X^2(\theta)$ is suitably defined as in the examples above.

3 The unconditional mean of X_i is $\mathrm{E}(X_i) = \mathrm{E}[\mathrm{E}(X_i \mid \Theta)] = \mathrm{E}[\mu_X(\Theta)] = \mu_X$. The mean of the conditional variance of X_i is

$$\begin{aligned}
\mathrm{E}[\mathrm{Var}(X_i \mid \Theta)] &= \mathrm{E}\left[\frac{\sigma_X^2(\Theta)}{m_i}\right] \\
&= \frac{\mu_{\mathrm{PV}}}{m_i},
\end{aligned} \tag{7.40}$$

for $i \in \{1, \cdots, n\}$, where $\mu_{\mathrm{PV}} = \mathrm{E}[\sigma_X^2(\Theta)]$, and the variance of its conditional mean is

$$\begin{aligned}
\mathrm{Var}[\mathrm{E}(X_i \mid \Theta)] &= \mathrm{Var}[\mu_X(\Theta)] \\
&= \sigma_{\mathrm{HM}}^2.
\end{aligned} \tag{7.41}$$

4 The Bühlmann–Straub predictor minimizes the MSE of all predictors of X_{n+1} that are linear in $\boldsymbol{X}$ over the joint distribution of Θ, X_{n+1} and $\boldsymbol{X}$. The predictor is given by

$$\hat{X}_{n+1} = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n, \tag{7.42}$$

where $\beta_0, \beta_1, \cdots, \beta_n$ are chosen to minimize the MSE of the predictor.

The steps in the last section can be used to solve for the least MSE predictor of the Bühlmann–Straub model. In particular, equations (7.22), (7.23), and (7.24) hold. The solution differs due to the variance term in equation (7.25). First, for the total variance of X_i, we have

$$\begin{aligned}\mathrm{Var}(X_i) &= \mathrm{E}[\mathrm{Var}(X_i \mid \Theta)] + \mathrm{Var}[\mathrm{E}(X_i \mid \Theta)] \\ &= \frac{\mu_{\mathrm{PV}}}{m_i} + \sigma^2_{\mathrm{HM}}.\end{aligned} \tag{7.43}$$

Second, following the same argument as in equation (7.26), the covariance terms are

$$\begin{aligned}\mathrm{Cov}(X_i, X_j) &= \mathrm{Var}[\mu_X(\Theta)] \\ &= \sigma^2_{\mathrm{HM}},\end{aligned} \tag{7.44}$$

for $i \neq j$. Now equation (7.27) can be written as

$$\hat{\boldsymbol{\beta}}_S = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} = \left(\mathbf{V} + \sigma^2_{\mathrm{HM}} \mathbf{1}\mathbf{1}'\right)^{-1} \left(\sigma^2_{\mathrm{HM}} \mathbf{1}\right), \tag{7.45}$$

where $\mathbf{V}$ is the $n \times n$ diagonal matrix

$$\mathbf{V} = \mu_{\mathrm{PV}} \begin{bmatrix} m_1^{-1} & 0 & 0 & \cdots & 0 \\ 0 & m_2^{-1} & 0 & \cdots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & & \\ 0 & 0 & & \cdots & m_n^{-1} \end{bmatrix}. \tag{7.46}$$

It can be verified that

$$\left(\mathbf{V} + \sigma^2_{\mathrm{HM}} \mathbf{1}\mathbf{1}'\right)^{-1} = \mathbf{V}^{-1} - \frac{\sigma^2_{\mathrm{HM}} (\mathbf{V}^{-1}\mathbf{1})(\mathbf{1}'\mathbf{V}^{-1})}{1 + \sigma^2_{\mathrm{HM}} \mathbf{1}'\mathbf{V}^{-1}\mathbf{1}}. \tag{7.47}$$

Thus, denoting

$$m = \sum_{i=1}^{n} m_i \tag{7.48}$$

and $\boldsymbol{m} = (m_1, \cdots, m_n)'$, we have

$$\mathbf{V}^{-1} - \frac{\sigma^2_{\mathrm{HM}}(\mathbf{V}^{-1}\mathbf{1})(\mathbf{1}'\mathbf{V}^{-1})}{1 + \sigma^2_{\mathrm{HM}}\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}} = \mathbf{V}^{-1} - \frac{1}{\mu_{\mathrm{PV}}}\left(\frac{\sigma^2_{\mathrm{HM}}\boldsymbol{m}\boldsymbol{m}'}{\mu_{\mathrm{PV}} + m\sigma^2_{\mathrm{HM}}}\right), \tag{7.49}$$

so that from equation (7.45) we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_S &= \left[\mathbf{V}^{-1} - \frac{1}{\mu_{\mathrm{PV}}}\left(\frac{\sigma^2_{\mathrm{HM}}\boldsymbol{m}\boldsymbol{m}'}{\mu_{\mathrm{PV}} + m\sigma^2_{\mathrm{HM}}}\right)\right]\left(\sigma^2_{\mathrm{HM}}\mathbf{1}\right) \\
&= \left(\frac{\sigma^2_{\mathrm{HM}}}{\mu_{\mathrm{PV}}}\right)\boldsymbol{m} - \left(\frac{\sigma^2_{\mathrm{HM}}}{\mu_{\mathrm{PV}}}\right)\left(\frac{\sigma^2_{\mathrm{HM}}m\boldsymbol{m}}{\mu_{\mathrm{PV}} + m\sigma^2_{\mathrm{HM}}}\right) \\
&= \left(\frac{\sigma^2_{\mathrm{HM}}}{\mu_{\mathrm{PV}}}\right)\left(1 - \frac{\sigma^2_{\mathrm{HM}}m}{\mu_{\mathrm{PV}} + m\sigma^2_{\mathrm{HM}}}\right)\boldsymbol{m} \\
&= \frac{\sigma^2_{\mathrm{HM}}\boldsymbol{m}}{\mu_{\mathrm{PV}} + m\sigma^2_{\mathrm{HM}}}.
\end{aligned} \tag{7.50}$$

We now define

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{n} m_i X_i = \frac{1}{m}\boldsymbol{m}'\boldsymbol{X} \tag{7.51}$$

and

$$k = \frac{\mu_{\mathrm{PV}}}{\sigma^2_{\mathrm{HM}}} \tag{7.52}$$

to obtain

$$\hat{\boldsymbol{\beta}}'_S\boldsymbol{X} = \frac{\sigma^2_{\mathrm{HM}}\boldsymbol{m}'\boldsymbol{X}}{\mu_{\mathrm{PV}} + m\sigma^2_{\mathrm{HM}}} = \frac{m}{m+k}\bar{X} = Z\bar{X}, \tag{7.53}$$

where

$$Z = \frac{m}{m+k}. \tag{7.54}$$

If we replace $\boldsymbol{X}$ in equation (7.53) by $\mathbf{1}$, we have

$$\hat{\boldsymbol{\beta}}'_S\mathbf{1} = Z, \tag{7.55}$$

so that from equation (7.32) we obtain

$$\hat{\beta}_0 = \mu_X - \mu_X \hat{\boldsymbol{\beta}}_S' \mathbf{1} = (1 - Z)\mu_X. \tag{7.56}$$

Combining the results in equations (7.53) and (7.56), we conclude that

$$\hat{X}_{n+1} = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_S' \boldsymbol{X} = Z\bar{X} + (1 - Z)\mu_X, \tag{7.57}$$

where Z is defined in equation (7.54).

Example 7.9 The number of accident claims incurred per year for each insured is distributed as a binomial random variable $\mathcal{BN}(2, \theta)$, and the claim incidences are independent across insureds. The probability θ of the binomial has a beta distribution with parameters $\alpha = 1$ and $\beta = 10$. The data in Table 7.7 are given for a block of policies.

Table 7.7. *Data for Example 7.9*

Year	Number of insureds	Number of claims
1	100	7
2	200	13
3	250	18
4	280	–

Calculate the Bühlmann–Straub credibility prediction of the number of claims in the fourth year.

Solution Let m_i be the number of insureds in Year i, and X_i be the number of claims per insured in Year i. Define X_{ij} as the number of claims for the jth insured in Year i, which is distributed as $\mathcal{BN}(2, \theta)$. Thus, we have

$$\mathrm{E}(X_i \,|\, \Theta) = \frac{1}{m_i} \sum_{j=1}^{m_i} \mathrm{E}(X_{ij} \,|\, \Theta) = 2\Theta,$$

and

$$\sigma^2_{\mathrm{HM}} = \mathrm{Var}[\mathrm{E}(X_i \,|\, \Theta)] = \mathrm{Var}(2\Theta) = 4\mathrm{Var}(\Theta).$$

As Θ has a beta distribution with parameters $\alpha = 1$ and $\beta = 10$, we have[8]

$$\mathrm{Var}(\Theta) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{10}{(11)^2(12)} = 0.006887.$$

[8] See Appendix A.10.6 for the moments of the beta distribution.

For the conditional variance of X_i, we have

$$\mathrm{Var}(X_i \mid \Theta) = \frac{2\Theta(1-\Theta)}{m_i}.$$

Thus

$$\mu_{\mathrm{PV}} = 2\mathrm{E}[\Theta(1-\Theta)].$$

As

$$\mathrm{E}(\Theta) = \frac{\alpha}{\alpha+\beta} = 0.0909,$$

we have

$$\begin{aligned}
\mu_{\mathrm{PV}} &= 2[\mathrm{E}(\Theta) - \mathrm{E}(\Theta^2)] \\
&= 2\left\{\mathrm{E}(\Theta) - \left(\mathrm{Var}(\Theta) + [\mathrm{E}(\Theta)]^2\right)\right\} \\
&= 2\{0.0909 - [0.006887 + (0.0909)^2]\} = 0.1515.
\end{aligned}$$

Thus

$$k = \frac{\mu_{\mathrm{PV}}}{\sigma^2_{\mathrm{HM}}} = \frac{0.1515}{(4)(0.006887)} = 5.5.$$

As $m = 100 + 200 + 250 = 550$, we have

$$Z = \frac{550}{550+5.5} = 0.9901.$$

Now

$$\mu_X = \mathrm{E}[\mathrm{E}(X_i \mid \Theta)] = (2)(0.0909) = 0.1818$$

and

$$\bar{X} = \frac{7+13+18}{550} = 0.0691.$$

Thus, the predicted number of claims per insured is

$$(0.9901)(0.0691) + (1-0.9901)(0.1818) = 0.0702,$$

and the predicted number of claims in Year 4 is

$$(280)(0.0702) = 19.66.$$

□

Example 7.10 The number of accident claims incurred per year for each insured is a Bernoulli random variable with probability θ, which takes value 0.1 with probability 0.8 and 0.2 with probability 0.2. Each claim may be of amount 20, 30, or 40, with equal probabilities. Claim frequency and claim severity are assumed to be independent for each insured. The data for the total claim amount are given in Table 7.8.

Table 7.8. *Data for Example 7.10*

Year	Number of insureds	Total claim amount
1	100	240
2	200	380
3	250	592
4	280	–

Calculate the Bühlmann–Straub credibility prediction of the pure premium and total loss in the fourth year.

Solution Let X_{ij} be the claim amount for the jth insured in Year i, each of which is distributed as $X = NW$, where

$$N = \begin{cases} 0, & \text{with probability } 1 - \Theta, \\ 1, & \text{with probability } \Theta, \end{cases}$$

and W = 20, 30, and 40, with equal probabilities. We have

$$\mathrm{E}(N \mid \Theta) = \Theta,$$

and

$$\mathrm{Var}(N \mid \Theta) = \Theta(1 - \Theta).$$

We evaluate the moments of Θ to obtain

$$\mathrm{E}(\Theta) = (0.1)(0.8) + (0.2)(0.2) = 0.12,$$

and

$$\mathrm{E}(\Theta^2) = (0.1)^2(0.8) + (0.2)^2(0.2) = 0.016.$$

Also

$$\mathrm{E}(W) = \frac{20 + 30 + 40}{3} = 30,$$

and

$$\mathrm{E}(W^2) = \frac{(20)^2 + (30)^2 + (40)^2}{3} = 966.6667,$$

so that

$$\mathrm{E}(X \mid \Theta) = \mathrm{E}(N \mid \Theta)\mathrm{E}(W) = 30\Theta.$$

Using equation (A.118), we have

$$\begin{aligned}
\mathrm{Var}(X \mid \Theta) &= \mathrm{E}(W^2)\mathrm{Var}(N \mid \Theta) + [\mathrm{E}(N \mid \Theta)]^2\mathrm{Var}(W) \\
&= 966.6667\Theta(1 - \Theta) + \Theta^2[966.6667 - (30)^2] \\
&= 966.6667\Theta - 900\Theta^2.
\end{aligned}$$

Thus, EPV is

$$\mu_{\mathrm{PV}} = \mathrm{E}[\mathrm{Var}(X \mid \Theta)] = (966.6667)(0.12) - (900)(0.016) = 101.60,$$

and VHM is

$$\begin{aligned}
\sigma^2_{\mathrm{HM}} &= \mathrm{Var}[\mathrm{E}(X \mid \Theta)] \\
&= \mathrm{Var}(30\Theta) \\
&= 900 \left\{\mathrm{E}(\Theta^2) - [\mathrm{E}(\Theta)]^2\right\} \\
&= 900[0.016 - (0.12)^2] \\
&= 1.44.
\end{aligned}$$

Thus

$$k = \frac{\mu_{\mathrm{PV}}}{\sigma^2_{\mathrm{HM}}} = \frac{101.60}{1.44} = 70.5556.$$

As $m = 100 + 200 + 250 = 550$

$$Z = \frac{550}{550 + 70.5556} = 0.8863.$$

Now

$$\bar{X} = \frac{240 + 380 + 592}{550} = 2.2036,$$

and

$$\mu_X = \mathrm{E}(X) = 30\mathrm{E}(\Theta) = (30)(0.12) = 3.60,$$

so that the Bühlmann–Straub prediction for the pure premium is

$$(0.8863)(2.2036) + (1 - 0.8863)(3.60) = 2.3624,$$

and the Bühlmann–Straub prediction for the total claim amount in Year 4 is

$$(280)(2.3624) = 661.4638.$$

□

7.5 Summary and discussions

The Bühlmann–Straub credibility model is a generalization of the Bühlmann credibility model, and we summarize its main results again here. Let X_i be the loss measure per unit exposure in period i, with the amount of exposure being m_i, for $i = 1, \cdots, n$. Let $m = \sum_{i=1}^{n} m_i$ and Θ be the parameter determining the distribution of X_i. Assume X_i are independently distributed.

Let

$$\mu_{\mathrm{PV}} = m_i\, \mathrm{E}[\mathrm{Var}(X_i \mid \Theta)], \tag{7.58}$$

and

$$\sigma^2_{\mathrm{HM}} = \mathrm{Var}[\mathrm{E}(X_i \mid \Theta)], \tag{7.59}$$

then the Bühlmann–Straub prediction of X_{n+1} is

$$\hat{X}_{n+1} = Z\bar{X} + (1 - Z)\mu_X, \tag{7.60}$$

where $\mu_X = \mathrm{E}(X_i)$, for $i = 1, \cdots, n$, and

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{n} m_i X_i, \tag{7.61}$$

and

$$Z = \frac{m}{m + k}, \tag{7.62}$$

with

$$k = \frac{\mu_{\mathrm{PV}}}{\sigma^2_{\mathrm{HM}}}. \tag{7.63}$$

In the special case where the exposures of all periods are the same, say $m_i = \bar{m}$ for $i = 1, \cdots, n$, then

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{7.64}$$

and

$$Z = \frac{n\bar{m}}{n\bar{m} + \dfrac{\mu_{\mathrm{PV}}}{\sigma^2_{\mathrm{HM}}}} = \frac{n}{n + \dfrac{\mu_{\mathrm{PV}}}{\bar{m}\sigma^2_{\mathrm{HM}}}} = \frac{n}{n + \dfrac{\mathrm{E}[\mathrm{Var}(X_i \,|\, \Theta)]}{\sigma^2_{\mathrm{HM}}}}. \tag{7.65}$$

Thus, the Bühlmann–Straub credibility predictor can be specialized to the Bühlmann predictor.

For the examples given in this chapter we assume that the variance components are known. In practice, they have to be estimated from the data. In Chapter 9 we shall consider the empirical implementation of the Bühlmann and Bühlmann–Straub credibility models when EPV and VHM are unknown. While we have proved the optimality of the Bühlmann predictor in the class of *linear* predictors, it turns out that its optimality may be more general. We shall see the details of this in the next chapter.

Exercises

7.1 Refer to Example 7.6. Calculate the unconditional covariance between X and N, $\mathrm{Cov}(X, N)$.

7.2 Refer to Example 7.6. Find the Bühlmann credibility parameters for claim frequency N and claim severity X.

7.3 Refer to Example 7.6. If the claim-severity distribution X is gamma with parameters $\alpha = c\,\theta$ and $\beta = 1/c$, derive an expression of the Bühlmann credibility parameter k for the aggregate loss per year S in terms of c.

7.4 Refer to Example 7.8. Calculate the updated prediction for claim frequency and claim severity for next year, for $c = 20$ and 30.

7.5 Following the set-up in Section 7.3, let $\boldsymbol{X} = \{X_1, \cdots, X_n\}$ be a random sample of losses that are independently and identically distributed as the random variable X, the distribution of which depends on a risk parameter θ. Denote $\mu_{n+1}(\Theta) = \mathrm{E}(X_{n+1}|\Theta)$, and consider the problem of estimating $\mu_{n+1}(\Theta)$ by a linear function of $\boldsymbol{X}$, denoted by $\hat{\mu}_{n+1}$, that minimizes the mean squared error $\mathrm{E}\left[(\mu_{n+1}(\Theta) - \hat{\mu}_{n+1})^2\right]$. Show that $\hat{\mu}_{n+1}$ is the same as the Bühlmann credibility predictor for future loss given in equations (7.34) and (7.35).

Questions adapted from SOA exams

7.6 The Bühlmann credibility assigned for estimating X_5 based on $X_1, \cdots, X_4$ is $Z = 0.4$. If the expected value of the process variance is 8, calculate $\mathrm{Cov}(X_i, X_j)$ for $i \neq j$.

7.7 The annual number of claims of a policy is distributed as $\mathcal{GM}(1/(1+\theta))$. If Θ follows the $\mathcal{P}(\alpha, 1)$ distribution, where $\alpha > 2$, and a randomly selected policy has x claims in Year 1, derive the Bühlmann credibility estimate of the expected number of claims of the policy in Year 2.

7.8 For a portfolio of insurance policies the annual claim amount X of a policy has the following pdf

$$f_X(x \,|\, \theta) = \frac{2x}{\theta^2}, \qquad 0 < x < \theta.$$

The prior distribution of Θ has the following pdf

$$f_\Theta(\theta) = 4\theta^3, \qquad 0 < \theta < 1.$$

A randomly selected policy has claim amount 0.1 in Year 1. Determine the Bühlmann credibility estimate of the expected claim amount of the selected policy in Year 2.

7.9 The number of claims in a year of a selected risk group follows the $\mathcal{PN}(\lambda)$ distribution. Claim severity follows the $\mathcal{E}(1/\theta)$ distribution and is independent of the claim frequency. If $\Lambda \sim \mathcal{E}(1)$ and $\Theta \sim \mathcal{PN}(1)$, and Λ and Θ are independent, determine the Bühlmann credibility parameter k for the estimation of the expected annual aggregate loss.

7.10 An insurance company sells two types of policies with the following characteristics:

Type of policy	Proportion of policies	Annual claim frequency, Poisson
1	θ	$\lambda = 0.5$
2	$1 - \theta$	$\lambda = 1.5$

A randomly selected policyholder has one claim in Year 1. Determine the Bühlmann credibility factor Z of this policyholder.

7.11 Claim frequency follows a Poisson distribution with mean λ. Claim size follows an exponential distribution with mean 10λ and is independent of claim frequency. If the distribution of Λ has pdf

$$f_\Lambda(\lambda) = \frac{5}{\lambda^6}, \qquad \lambda > 1,$$

calculate the Bühlmann credibility parameter k for aggregate losses.

7.12 Two risks have the following severity distributions:

Claim amount	Probability of claim amount for Risk 1	Probability of claim amount for Risk 2
250	0.5	0.7
2,500	0.3	0.2
60,000	0.2	0.1

If Risk 1 is twice as likely to be observed as Risk 2 and a claim of 250 is observed, determine the Bühlmann credibility estimate of the expected second claim amount from the same risk.

7.13 Claim frequency in a month is distributed as $\mathcal{PN}(\lambda)$, and the distribution of Λ is $\mathcal{G}(6, 0.01)$. The following data are available:

Month	Number of insureds	Number of claims
1	100	6
2	150	8
3	200	11
4	300	–

Calculate the Bühlmann–Straub credibility estimate of the expected number of claims in Month 4.

7.14 The number of claims made by an individual insured in a year is distributed as $\mathcal{PN}(\lambda)$, where Λ is distributed as $\mathcal{G}(1, 1.2)$. If three claims are observed in Year 1 and no claim is observed in Year 2, calculate the Bühlmann credibility estimate of the expected number of claims in Year 3.

7.15 Annual claim frequency of an individual policyholder has mean λ, which is distributed as $\mathcal{U}(0.5, 1.5)$, and variance σ^2, which is distributed as exponential with mean 1.25. A policyholder is selected randomly and found to have no claim in Year 1. Using Bühlmann credibility, estimate the expected number of claims in Year 2 for the selected policyholder.

7.16 You are given the following joint distribution of X and Θ:

	Θ	
X	0	1
0	0.4	0.1
1	0.1	0.2
2	0.1	0.1

For a given (but unknown) value of Θ and a sample of ten observations of X with a total of 10, determine the Bühlmann credibility premium.

7.17 There are four classes of insureds, each of whom may have zero or one claim, with the following probabilities:

	Number of claims	
Class	0	1
A	0.9	0.1
B	0.8	0.2
C	0.5	0.5
D	0.1	0.9

A class is selected randomly, with probability of one-fourth, and four insureds are selected at random from the class. The total number of claims is two. If five insureds are selected at random from the same class, determine the Bühlmann–Straub credibility estimate of the expected total number of claims.

7.18 An insurance company has a large portfolio of insurance policies. Each insured may file a maximum of one claim per year, and the probability of a claim for each insured is constant over time. A randomly selected insured has a probability 0.1 of filing a claim in a year, and the variance of the claim probability of individual insured is 0.01. A randomly selected individual is found to have filed no claim over the past ten years. Determine the Bühlmann credibility estimate for the expected number of claims the selected insured will file over the next five years.

7.19 A portfolio of insurance policies comprises of 100 insureds. The aggregate loss of each insured in a year follows a compound distribution, where the primary distribution is $\mathcal{NB}(r, 1/1.2)$ and the secondary distribution is $\mathcal{P}(3, 1000)$. If the distribution of r is exponential with mean 2, determine the Bühlmann credibility factor Z of the portfolio.

7.20 An insurance company has a large portfolio of employee compensation policies. The losses of each employee are independently and identically distributed. The overall average loss of each employee is 20, the variance of the hypothetical means is 40, and the expected value of the process variance is 8,000. The following data are available in the last three years for a randomly selected policyholder:

Year	Average loss per employee	Number of employees
1	15	800
2	10	600
3	5	400

Determine the Bühlmann–Straub credibility premium per employee for this policyholder.

7.21 Claim severity has mean μ and variance 500, where the distribution of μ has a mean of 1,000 and a variance of 50. The following three claims were observed: 750, 1,075 and 2,000. Calculate the Bühlmann estimate of the expected claim severity of the next claim.

7.22 Annual losses are distributed as $\mathcal{G}(\alpha, \beta)$, where β does not vary with policyholders. The distribution of α has a mean of 50, and the Bühlmann credibility factor based on two years of experience is 0.25. Calculate the variance of the distribution of α.

7.23 The aggregate losses per year per exposure of a portfolio of insurance risks follow a normal distribution with mean μ and standard deviation 1,000. You are given that μ varies by class of risk as follows:

Class	μ	Probability of class
A	2,000	0.6
B	3,000	0.3
C	4,000	0.1

A randomly selected risk has the following experience over three years:

Year	Number of exposures	Aggregate losses
1	24	24,000
2	30	36,000
3	26	28,000

Calculate the Bühlmann–Straub estimate of the expected aggregate loss per exposure in Year 4 for this risk.

7.24 The annual loss of an individual policy is distributed as $\mathcal{G}(4, \beta)$, where the mean of the distribution of β is 600. A randomly selected policy had losses of 1,400 in Year 1 and 1,900 in Year 2. Loss data for Year 3 were misfiled and the Bühlmann credibility estimate of the expected loss for the selected policy in Year 4 based on the data for Years 1 and 2 was 1,800. The loss for the selected policy in Year 3, however, was found later to be 2,763. Determine the Bühlmann credibility estimate of the expected loss for the selected policy in Year 4 based on the data of Years 1, 2, and 3.

7.25 Claim frequency in a month is distributed as $\mathcal{PN}(\lambda)$, where λ is distributed as $\mathcal{W}(2, 0.1)$. You are given the following data:

Month	Number of insureds	Number of claims
1	100	10
2	150	11
3	250	14

Calculate the Bühlmann–Straub credibility estimate of the expected number of claims in the next 12 months for 300 insureds. [*Hint:* You may use the Excel function GAMMALN to compute the natural logarithm of the gamma function.]

8
Bayesian approach

In this chapter we consider the Bayesian approach in updating the prediction for future losses. We consider the derivation of the posterior distribution of the risk parameters based on the prior distribution of the risk parameters and the likelihood function of the data. The Bayesian estimate of the risk parameter under the squared-error loss function is the mean of the posterior distribution. Likewise, the Bayesian estimate of the mean of the random loss is the posterior mean of the loss conditional on the data.

In general, the Bayesian estimates are difficult to compute, as the posterior distribution may be quite complicated and intractable. There are, however, situations where the computation may be straightforward, as in the case of conjugate distributions. We define conjugate distributions and provide some examples for cases that are of relevance in analyzing loss measures. Under specific classes of conjugate distributions, the Bayesian predictor is the same as the Bühlmann predictor. Specifically, when the likelihood belongs to the linear exponential family and the prior distribution is the natural conjugate, the Bühlmann credibility estimate is equal to the Bayesian estimate. This result provides additional justification for the use of the Bühlmann approach.

Learning objectives

1 Bayesian inference and estimation
2 Prior and posterior pdf
3 Bayesian credibility
4 Conjugate prior distribution
5 Linear exponential distribution
6 Bühlmann credibility versus Bayesian credibility

8.1 Bayesian inference and estimation

The classical and Bühlmann credibility models update the prediction for future losses based on recent claim experience and existing prior information. In these models, the random loss variable X has a distribution that varies with different risk groups. Based on a sample of n observations of random losses, the predicted value of the loss for the next period is updated. The predictor is a weighted average of the sample mean of X and the prior mean, where the weights depend on the distribution of X across different risk groups.

We formulate the aforementioned as a statistical problem suitable for the Bayesian approach of statistical inference and estimation. The set-up is summarized as follows:[1]

1 Let X denote the random loss variable (such as claim frequency, claim severity, and aggregate loss) of a risk group. The distribution of X is dependent on a parameter θ, which varies with different risk groups and is hence treated as the realization of a random variable Θ.
2 Θ has a statistical distribution called the **prior distribution**. The **prior pdf** of Θ is denoted by $f_\Theta(\theta \mid \gamma)$ (or simply $f_\Theta(\theta)$), which depends on the parameter γ, called the **hyperparameter**.
3 The conditional pdf of X given the parameter θ is denoted by $f_{X \mid \Theta}(x \mid \theta)$. Suppose $\boldsymbol{X} = \{X_1, \cdots, X_n\}$ is a random sample of X, and $\boldsymbol{x} = (x_1, \cdots, x_n)$ is a realization of $\boldsymbol{X}$. The conditional pdf of $\boldsymbol{X}$ is

$$f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta) = \prod_{i=1}^{n} f_{X \mid \Theta}(x_i \mid \theta). \tag{8.1}$$

We call $f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta)$ the **likelihood function**.
4 Based on the sample data $\boldsymbol{x}$, the distribution of Θ is updated. The conditional pdf of Θ given $\boldsymbol{x}$ is called the **posterior pdf**, and is denoted by $f_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x})$.
5 An estimate of the mean of the random loss, which is a function of Θ, is computed using the posterior pdf of Θ. This estimate, called the **Bayes estimate**, is also the predictor of future losses.

Bayesian inference differs from classical statistical inference in its treatment of the prior distribution of the parameter θ. Under classical statistical inference, θ is assumed to be *fixed* and *unknown*, and the relevant entity for inference is the likelihood function. For Bayesian inference, the prior distribution has an important role. The likelihood function and the prior pdf jointly determine the posterior pdf, which is then used for statistical inference.

[1] For convenience of exposition, we assume all distributions (both the prior and the likelihood) are continuous. Thus, we use the terminology "pdf" and compute the marginal pdf using integration. If the distribution is discrete, we need to replace "pdf" by "pf", and use summation instead of integration. A brief introduction to Bayesian inference can be found in Appendix A.15.

We now discuss the derivation of the posterior pdf and the Bayesian approach of estimating Θ.

8.1.1 Posterior distribution of parameter

Given the prior pdf of Θ and the likelihood function of $\boldsymbol{X}$, the joint pdf of Θ and X can be obtained as follows

$$f_{\Theta \boldsymbol{X}}(\theta, \boldsymbol{x}) = f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta) f_{\Theta}(\theta). \tag{8.2}$$

Integrating out θ from the joint pdf of Θ and $\boldsymbol{X}$, we obtain the marginal pdf of $\boldsymbol{X}$ as

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{\theta \in \Omega_{\Theta}} f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta) f_{\Theta}(\theta)\, d\theta, \tag{8.3}$$

where Ω_{Θ} is the support of Θ.

Now we can turn the question around and consider the conditional pdf of Θ given the data $\boldsymbol{x}$, i.e. $f_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x})$. Combining equations (8.2) and (8.3), we have

$$\begin{aligned} f_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x}) &= \frac{f_{\Theta \boldsymbol{X}}(\theta, \boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})} \\ &= \frac{f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta) f_{\Theta}(\theta)}{\int_{\theta \in \Omega_{\Theta}} f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta) f_{\Theta}(\theta)\, d\theta}. \end{aligned} \tag{8.4}$$

The posterior pdf describes the distribution of Θ based on prior information about Θ and the sample data $\boldsymbol{x}$. Bayesian inference about the population as described by the risk parameter Θ is then based on the posterior pdf.

Example 8.1 Let X be the Bernoulli random variable which takes value 1 with probability θ and 0 with probability $1 - \theta$. If Θ follows the beta distribution with parameters α and β, i.e. $\Theta \sim \mathcal{B}(\alpha, \beta)$, calculate the posterior pdf of Θ given X.[2]

Solution As X is Bernoulli, the likelihood function of X is

$$f_{X \mid \Theta}(x \mid \theta) = \theta^{x}(1 - \theta)^{1-x}, \qquad \text{for } x = 0, 1.$$

Since Θ is assumed to follow the beta distribution with hyperparameters α and β, the prior pdf of Θ is

$$f_{\Theta}(\theta) = \frac{\theta^{\alpha-1}(1 - \theta)^{\beta-1}}{B(\alpha, \beta)}, \qquad \text{for } \theta \in (0, 1),$$

[2] See Appendix A.10.6 for some properties of the $\mathcal{B}(\alpha, \beta)$ distribution. The support of $\mathcal{B}(\alpha, \beta)$ is the interval (0, 1), so that the distribution is suitable for modeling probability as a random variable.

where $B(\alpha, \beta)$ is the beta function defined in equation (A.102). Thus, the joint pf–pdf of Θ and X is

$$f_{\Theta X}(\theta, x) = f_{X \mid \Theta}(x \mid \theta) f_{\Theta}(\theta) = \frac{\theta^{\alpha+x-1}(1-\theta)^{(\beta-x+1)-1}}{B(\alpha, \beta)},$$

from which we compute the marginal pf of X by integration to obtain

$$\begin{aligned} f_X(x) &= \int_0^1 \frac{\theta^{\alpha+x-1}(1-\theta)^{(\beta-x+1)-1}}{B(\alpha, \beta)} \, d\theta \\ &= \frac{B(\alpha+x, \beta-x+1)}{B(\alpha, \beta)}. \end{aligned}$$

Thus, we conclude

$$\begin{aligned} f_{\Theta \mid X}(\theta \mid x) &= \frac{f_{\Theta X}(\theta, x)}{f_X(x)} \\ &= \frac{\theta^{\alpha+x-1}(1-\theta)^{(\beta-x+1)-1}}{B(\alpha+x, \beta-x+1)}, \end{aligned}$$

which is the pdf of a beta distribution with parameters $\alpha + x$ and $\beta - x + 1$. □

Example 8.2 In Example 8.1, if there is a sample of n observations of X denoted by $\boldsymbol{X} = \{X_1, \cdots, X_n\}$, compute the posterior pdf of Θ.

Solution We first compute the likelihood of $\boldsymbol{X}$ as follows

$$\begin{aligned} f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta) &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{\sum_{i=1}^{n}(1-x_i)}, \end{aligned}$$

and the joint pf–pdf is

$$\begin{aligned} f_{\Theta \boldsymbol{X}}(\theta, \boldsymbol{x}) &= f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta) f_{\Theta}(\theta) \\ &= \left[\theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{\sum_{i=1}^{n}(1-x_i)}\right]\left[\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}\right] \\ &= \frac{\theta^{(\alpha+n\bar{x})-1}(1-\theta)^{(\beta+n-n\bar{x})-1}}{B(\alpha, \beta)}. \end{aligned}$$

As

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \int_0^1 f_{\Theta\boldsymbol{X}}(\theta, \boldsymbol{x})\, d\theta$$
$$= \int_0^1 \frac{\theta^{(\alpha+n\bar{x})-1}(1-\theta)^{(\beta+n-n\bar{x})-1}}{B(\alpha, \beta)}\, d\theta$$
$$= \frac{B(\alpha+n\bar{x}, \beta+n-n\bar{x})}{B(\alpha, \beta)},$$

we conclude that

$$f_{\Theta\,|\,\boldsymbol{X}}(\theta\,|\,\boldsymbol{x}) = \frac{f_{\Theta\boldsymbol{X}}(\theta, \boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})}$$
$$= \frac{\theta^{(\alpha+n\bar{x})-1}(1-\theta)^{(\beta+n-n\bar{x})-1}}{B(\alpha+n\bar{x}, \beta+n-n\bar{x})},$$

and the posterior pdf of Θ follows a beta distribution with parameters $\alpha + n\bar{x}$ and $\beta + n - n\bar{x}$. □

Note that the denominator in equation (8.4) is a function of $\boldsymbol{x}$ but not θ. Denoting

$$K(\boldsymbol{x}) = \frac{1}{\int_{\theta\,\in\,\Omega_\Theta} f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta) f_\Theta(\theta)\, d\theta}, \tag{8.5}$$

we can rewrite the posterior pdf of Θ as

$$f_{\Theta\,|\,\boldsymbol{X}}(\theta\,|\,\boldsymbol{x}) = K(\boldsymbol{x}) f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta) f_\Theta(\theta)$$
$$\propto f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta) f_\Theta(\theta). \tag{8.6}$$

$K(\boldsymbol{x})$ is free of θ and is a **constant of proportionality**. It scales the posterior pdf so that it integrates to 1. The expression $f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta) f_\Theta(\theta)$ enables us to identify the functional form of the posterior pdf in terms of θ without computing the marginal pdf of $\boldsymbol{X}$.

Example 8.3 Let $X \sim \mathcal{BN}(m, \theta)$, and $\boldsymbol{X} = \{X_1, \cdots, X_n\}$ be a random sample of X. If $\Theta \sim \mathcal{B}(\alpha, \beta)$, what is the posterior distribution of Θ?

Solution From equation (8.6), we have

$$f_{\Theta\,|\,\boldsymbol{X}}(\theta\,|\,\boldsymbol{x}) \propto f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta) f_\Theta(\theta)$$
$$\propto \left[\theta^{n\bar{x}}(1-\theta)^{\sum_{i=1}^n (m-x_i)}\right]\left[\theta^{\alpha-1}(1-\theta)^{\beta-1}\right]$$
$$\propto \theta^{(\alpha+n\bar{x})-1}(1-\theta)^{(\beta+mn-n\bar{x})-1}.$$

Comparing the above equation with equation (A.101), we conclude that the posterior pdf belongs to the class of beta distributions. We can further conclude that the hyperparameters of the beta posterior pdf are $\alpha + n\bar{x}$ and $\beta + mn - n\bar{x}$. Note that this is done without computing the expression for the constant of proportionality $K(\boldsymbol{x})$ nor the marginal pdf of $\boldsymbol{X}$. □

8.1.2 Loss function and Bayesian estimation

We now consider the problem of estimating $\mu_X(\Theta) = \mathrm{E}(X \mid \Theta)$ given the observed data $\boldsymbol{x}$. The Bayesian approach of estimation views the estimator as a decision rule, which assigns a value to $\mu_X(\Theta)$ based on the data. Thus, let $w(\boldsymbol{x})$ be an estimator of $\mu_X(\Theta)$. A nonnegative function $L[\mu_X(\Theta), w(\boldsymbol{x})]$, called the **loss function**, is then defined to reflect the penalty in making a wrong decision about $\mu_X(\Theta)$. Typically, the larger the difference between $\mu_X(\Theta)$ and $w(\boldsymbol{x})$, the larger the loss $L[\mu_X(\Theta), w(\boldsymbol{x})]$. A commonly used loss function is the **squared-error loss function** (or **quadratic loss function**) defined by

$$L[\mu_X(\Theta), w(\boldsymbol{x})] = [\mu_X(\Theta) - w(\boldsymbol{x})]^2. \tag{8.7}$$

Other popularly used loss functions include the **absolute-error loss function** and the **zero-one loss function**.[3] We assume, however, that the squared-error loss function is adopted in Bayesian inference.

Given the decision rule and the data, the expected loss in the estimation of $\mu_X(\Theta)$ is

$$\mathrm{E}\{L[\mu_X(\Theta), w(\boldsymbol{x})] \mid \boldsymbol{x}\} = \int_{\theta \in \Omega_\Theta} L[\mu_X(\Theta), w(\boldsymbol{x})] f_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x})\, d\theta. \tag{8.8}$$

It is naturally desirable to have a decision rule that gives as small an expected loss as possible. Thus, for any given $\boldsymbol{x}$, if the decision rule $w(\boldsymbol{x})$ assigns a value to $\mu_X(\Theta)$ that minimizes the expected loss, then the decision rule $w(\boldsymbol{x})$ is called the **Bayes estimator** of $\mu_X(\Theta)$ with respect to the chosen loss function. In other words, the Bayes estimator, denoted by $w^*(\boldsymbol{x})$, satisfies

$$\mathrm{E}\{L[\mu_X(\Theta), w^*(\boldsymbol{x})] \mid \boldsymbol{x}\} = \min_{w(\cdot)} \mathrm{E}\{L[\mu_X(\Theta), w(\boldsymbol{x})] \mid \boldsymbol{x}\}, \tag{8.9}$$

for any given $\boldsymbol{x}$. For the squared-error loss function, the decision rule (estimator) that minimizes the expected loss $\mathrm{E}\{[\mu_X(\Theta) - w(\boldsymbol{x})]^2 \mid \boldsymbol{x}\}$ is[4]

$$w^*(\boldsymbol{x}) = \mathrm{E}[\mu_X(\Theta) \mid \boldsymbol{x}]. \tag{8.10}$$

[3] For estimating θ with the estimator $\hat{\theta}$, the absolute-error loss function is defined by $L[\theta, \hat{\theta}] = |\,\theta - \hat{\theta}\,|$ and the zero-one loss function is defined by $L[\theta, \hat{\theta}] = 0$ if $\hat{\theta} = \theta$ and 1 otherwise.

[4] See DeGroot and Schervish (2002, p. 348) for a proof of this result.

Thus, for the squared-error loss function, the Bayes estimator of $\mu_X(\Theta)$ is the posterior mean, denoted by $\hat{\mu}_X(\boldsymbol{x})$, so that[5]

$$\hat{\mu}_X(\boldsymbol{x}) = \mathrm{E}[\mu_X(\Theta) \,|\, \boldsymbol{x}] = \int_{\theta \,\in\, \Omega_\Theta} \mu_X(\theta) f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x})\, d\theta. \tag{8.11}$$

In the credibility literature (where X is a loss random variable), $\hat{\mu}_X(x)$ is called the **Bayesian premium**.

An alternative way to interpret the Bayesian premium is to consider the prediction of the loss in the next period, namely, X_{n+1}, given the data $\boldsymbol{x}$. To this effect, we first calculate the conditional pdf of X_{n+1} given $\boldsymbol{x}$, which is

$$\begin{aligned} f_{X_{n+1} \,|\, \boldsymbol{X}}(x_{n+1} \,|\, \boldsymbol{x}) &= \frac{f_{X_{n+1}\boldsymbol{X}}(x_{n+1}, \boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})} \\ &= \frac{\int_{\theta \,\in\, \Omega_\Theta} f_{X_{n+1}\boldsymbol{X} \,|\, \Theta}(x_{n+1}, \boldsymbol{x} \,|\, \theta) f_\Theta(\theta)\, d\theta}{f_{\boldsymbol{X}}(\boldsymbol{x})} \\ &= \frac{\int_{\theta \,\in\, \Omega_\Theta} \left[\prod_{i=1}^{n+1} f_{X_i \,|\, \Theta}(x_i \,|\, \theta)\right] f_\Theta(\theta)\, d\theta}{f_{\boldsymbol{X}}(\boldsymbol{x})}. \end{aligned} \tag{8.12}$$

As the posterior pdf of Θ given $\boldsymbol{X}$ is

$$f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x}) = \frac{f_{\Theta\boldsymbol{X}}(\theta, \boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})} = \frac{\left[\prod_{i=1}^{n} f_{X_i \,|\, \Theta}(x_i \,|\, \theta)\right] f_\Theta(\theta)}{f_{\boldsymbol{X}}(\boldsymbol{x})}, \tag{8.13}$$

we conclude

$$\left[\prod_{i=1}^{n} f_{X_i \,|\, \Theta}(x_i \,|\, \theta)\right] f_\Theta(\theta) = f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}). \tag{8.14}$$

Substituting (8.14) into (8.12), we obtain

$$f_{X_{n+1} \,|\, \boldsymbol{X}}(x_{n+1} \,|\, \boldsymbol{x}) = \int_{\theta \,\in\, \Omega_\Theta} f_{X_{n+1} \,|\, \boldsymbol{\Theta}}(x_{n+1} \,|\, \theta) f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x})\, d\theta. \tag{8.15}$$

Equation (8.15) shows that the conditional pdf of X_{n+1} given $\boldsymbol{X}$ can be interpreted as a mixture of the conditional pdf of X_{n+1}, where the mixing density is the posterior pdf of Θ.

[5] The Bayes estimator based on the absolute-error loss function is the posterior median, and the Bayes estimator based on the zero-one loss function is the posterior mode. See DeGroot and Schervish (2002, p. 349) for more discussions.

We now consider the prediction of X_{n+1} given $\boldsymbol{X}$. A natural predictor is the conditional expected value of X_{n+1} given $\boldsymbol{X}$, i.e. $\mathrm{E}(X_{n+1} \,|\, \boldsymbol{x})$, which is given by

$$\mathrm{E}(X_{n+1} \,|\, \boldsymbol{x}) = \int_0^{\infty} x_{n+1} f_{X_{n+1} \,|\, \boldsymbol{X}}(x_{n+1} \,|\, \boldsymbol{x})\, dx_{n+1}. \tag{8.16}$$

Using equation (8.15), we have

$$\begin{aligned}
\mathrm{E}(X_{n+1} \,|\, \boldsymbol{x}) &= \int_0^{\infty} x_{n+1} \left[\int_{\theta \,\in\, \Omega_\Theta} f_{X_{n+1} \,|\, \boldsymbol{\Theta}}(x_{n+1} \,|\, \theta) f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x})\, d\theta \right] dx_{n+1} \\
&= \int_{\theta \,\in\, \Omega_\Theta} \left[\int_0^{\infty} x_{n+1} f_{X_{n+1} \,|\, \boldsymbol{\Theta}}(x_{n+1} \,|\, \theta)\, dx_{n+1} \right] f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x})\, d\theta \\
&= \int_{\theta \,\in\, \Omega_\Theta} \mathrm{E}(X_{n+1} \,|\, \theta) f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x})\, d\theta \\
&= \int_{\theta \,\in\, \Omega_\Theta} \mu_X(\theta) f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x})\, d\theta \\
&= \mathrm{E}[\mu_X(\Theta) \,|\, \boldsymbol{x}].
\end{aligned} \tag{8.17}$$

Thus, the Bayesian premium can also be interpreted as the conditional expectation of X_{n+1} given $\boldsymbol{X}$.

In summary, the Bayes estimate of the mean of the random loss X, called the Bayesian premium, is the posterior mean of X conditional on the data $\boldsymbol{x}$, as given in equation (8.11). It is also equal to the conditional expectation of future loss given the data $\boldsymbol{x}$, as shown in equation (8.17). Thus, we shall use the terminologies Bayesian estimate of expected loss and Bayesian predictor of future loss interchangeably.

8.1.3 Some examples of Bayesian credibility

We now re-visit Examples 8.2 and 8.3 to illustrate the calculation of the Bayesian estimate of the expected loss.

Example 8.4 Let X be the Bernoulli random variable which takes value 1 with probability θ and 0 with probability $1-\theta$, and $\boldsymbol{X} = \{X_1, \cdots, X_n\}$ be a random sample of X. If $\Theta \sim \mathcal{B}(\alpha, \beta)$, calculate the posterior mean of $\mu_X(\Theta)$ and the expected value of a future observation X_{n+1} given the sample data.

Solution From Example 8.2, we know that the posterior distribution of Θ given $\boldsymbol{x}$ is beta with parameters $\alpha^* = \alpha + n\bar{x}$ and $\beta^* = \beta + n - n\bar{x}$. As X is a Bernoulli random variable, $\mu_X(\Theta) = \mathrm{E}(X \mid \Theta) = \Theta$. Hence, the posterior mean of $\mu_X(\Theta)$

is $E(\Theta \mid \boldsymbol{x}) = \alpha^*/(\alpha^* + \beta^*)$. Now the conditional pdf of X_{n+1} given Θ is

$$f_{X_{n+1} \mid \Theta}(x_{n+1} \mid \theta) = \begin{cases} \theta, & \text{for } x_{n+1} = 1, \\ 1 - \theta, & \text{for } x_{n+1} = 0. \end{cases}$$

From equation (8.15), the conditional pdf of X_{n+1} given $\boldsymbol{x}$ is

$$f_{X_{n+1} \mid \boldsymbol{X}}(x_{n+1} \mid \boldsymbol{x}) = \begin{cases} E(\Theta \mid \boldsymbol{x}) = \dfrac{\alpha^*}{\alpha^* + \beta^*}, & \text{for } x_{n+1} = 1, \\ 1 - E(\Theta \mid \boldsymbol{x}) = 1 - \dfrac{\alpha^*}{\alpha^* + \beta^*}, & \text{for } x_{n+1} = 0. \end{cases}$$

Now we apply the above results to equation (8.16) to obtain the conditional mean of X_{n+1} given $\boldsymbol{x}$ as[6]

$$\begin{aligned} E(X_{n+1} \mid \boldsymbol{x}) &= (1) \left[f_{X_{n+1} \mid \boldsymbol{X}}(1 \mid \boldsymbol{x}) \right] + (0) \left[f_{X_{n+1} \mid \boldsymbol{X}}(0 \mid \boldsymbol{x}) \right] \\ &= \frac{\alpha^*}{\alpha^* + \beta^*}, \end{aligned}$$

which is equal to the posterior mean of Θ, $E(\Theta \mid \boldsymbol{x})$. □

Example 8.5 Let $X \sim \mathcal{BN}(2, \theta)$, and $\boldsymbol{X} = \{X_1, \cdots, X_n\}$ be a random sample of X. If $\Theta \sim \mathcal{B}(\alpha, \beta)$, calculate the posterior mean of $\mu_X(\Theta)$ and the expected value of a future observation X_{n+1} given the sample data.

Solution From Example 8.3, we know that the posterior distribution of Θ is beta with parameters $\alpha^* = \alpha + n\bar{x}$ and $\beta^* = \beta + 2n - n\bar{x}$. As X is a binomial random variable, $\mu_X(\Theta) = E(X \mid \Theta) = 2\Theta$. Hence, the posterior mean of $\mu_X(\Theta)$ is $E(2\Theta \mid \boldsymbol{x}) = 2\alpha^*/(\alpha^* + \beta^*)$. Now the conditional pdf of X_{n+1} given Θ is

$$f_{X_{n+1} \mid \Theta}(x_{n+1} \mid \theta) = \binom{2}{x_{n+1}} \theta^{x_{n+1}} (1 - \theta)^{2 - x_{n+1}}, \qquad x_{n+1} \in \{0, 1, 2\}.$$

From equation (8.15), the conditional pdf of X_{n+1} given $\boldsymbol{x}$ is

$$f_{X_{n+1} \mid \boldsymbol{X}}(x_{n+1} \mid \boldsymbol{x}) = \begin{cases} E[(1 - \Theta)^2 \mid \boldsymbol{x}], & \text{for } x_{n+1} = 0, \\ 2E[\Theta(1 - \Theta) \mid \boldsymbol{x}], & \text{for } x_{n+1} = 1, \\ E[\Theta^2 \mid \boldsymbol{x}], & \text{for } x_{n+1} = 2. \end{cases}$$

[6] Note that X_{n+1} is a discrete random variable and we have to replace the integration in equation (8.16) by summation.

Now we apply the above results to equation (8.16) to obtain the conditional mean of X_{n+1} given $\boldsymbol{x}$ as

$$\begin{aligned}
\mathrm{E}(X_{n+1} \,|\, \boldsymbol{x}) &= (1)\left[f_{X_{n+1}\,|\,\boldsymbol{X}}(1 \,|\, \boldsymbol{x})\right] + (2)\left[f_{X_{n+1}\,|\,\boldsymbol{X}}(2 \,|\, \boldsymbol{x})\right] \\
&= 2\mathrm{E}[\Theta(1-\Theta) \,|\, \boldsymbol{x}] + 2\mathrm{E}[\Theta^2 \,|\, \boldsymbol{x}] \\
&= 2\mathrm{E}[\Theta \,|\, \boldsymbol{x}] \\
&= \frac{2\alpha^*}{\alpha^* + \beta^*},
\end{aligned}$$

which is equal to the posterior mean of $\mu_X(\Theta)$. □

Examples 8.4 and 8.5 illustrate the equivalence of equations (8.11) and (8.16). The results can be generalized to the case when X is a binomial random variable with parameters m and θ, where m is any positive integer. Readers may wish to prove this result as an exercise (see Exercise 8.1).

Example 8.6 X is the claim-severity random variable that can take values 10, 20, or 30. The distribution of X depends on the risk group defined by parameter Θ, which are labeled 1, 2, and 3. The relative frequencies of risk groups with Θ equal to 1, 2, and 3 are, respectively, 0.4, 0.4, and 0.2. The conditional distribution of X given the risk parameter Θ is given in Table 8.1.

Table 8.1. *Data for Example 8.6*

		$\Pr(X = x \,|\, \theta)$		
θ	$\Pr(\Theta = \theta)$	$x = 10$	$x = 20$	$x = 30$
1	0.4	0.2	0.3	0.5
2	0.4	0.4	0.4	0.2
3	0.2	0.5	0.5	0.0

A sample of three claims with $\boldsymbol{x} = (20, 20, 30)$ is observed. Calculate the posterior mean of X. Compute the conditional pdf of X_4 given $\boldsymbol{x}$, and calculate the expected value of X_4 given $\boldsymbol{x}$.

Solution We first calculate the conditional probability of $\boldsymbol{x}$ given Θ as follows

$$f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x} \,|\, 1) = (0.3)(0.3)(0.5) = 0.045,$$

$$f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x} \,|\, 2) = (0.4)(0.4)(0.2) = 0.032,$$

and

$$f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x} \,|\, 3) = (0.5)(0.5)(0) = 0.$$

Thus, the joint pf of $\boldsymbol{x}$ and Θ is

$$f_{\Theta X}(1, \boldsymbol{x}) = f_{X \mid \Theta}(\boldsymbol{x} \mid 1) f_{\Theta}(1) = (0.045)(0.4) = 0.018,$$
$$f_{\Theta X}(2, \boldsymbol{x}) = f_{X \mid \Theta}(\boldsymbol{x} \mid 2) f_{\Theta}(2) = (0.032)(0.4) = 0.0128,$$

and

$$f_{\Theta X}(3, \boldsymbol{x}) = f_{X \mid \Theta}(\boldsymbol{x} \mid 3) f_{\Theta}(3) = 0(0.2) = 0.$$

Thus, we obtain

$$f_X(\boldsymbol{x}) = 0.018 + 0.0128 + 0 = 0.0308,$$

so that the posterior distribution of Θ is

$$f_{\Theta \mid X}(1 \mid \boldsymbol{x}) = \frac{f_{\Theta X}(1, \boldsymbol{x})}{f_X(\boldsymbol{x})} = \frac{0.018}{0.0308} = 0.5844,$$
$$f_{\Theta \mid X}(2 \mid \boldsymbol{x}) = \frac{f_{\Theta X}(2, \boldsymbol{x})}{f_X(\boldsymbol{x})} = \frac{0.0128}{0.0308} = 0.4156,$$

and $f_{\Theta \mid X}(3 \mid \boldsymbol{x}) = 0$. The conditional means of X are

$$\mathrm{E}(X \mid \Theta = 1) = (10)(0.2) + (20)(0.3) + (30)(0.5) = 23,$$
$$\mathrm{E}(X \mid \Theta = 2) = (10)(0.4) + (20)(0.4) + (30)(0.2) = 18,$$

and

$$\mathrm{E}(X \mid \Theta = 3) = (10)(0.5) + (20)(0.5) + (30)(0) = 15.$$

Thus, the posterior mean of X is

$$\begin{aligned}\mathrm{E}\left[\mathrm{E}(X \mid \Theta) \mid \boldsymbol{x}\right] &= \sum_{\theta=1}^{3} \left[\mathrm{E}(X \mid \theta)\right] f_{\Theta \mid X}(\theta \mid \boldsymbol{x}) \\ &= (23)(0.5844) + (18)(0.4156) + (15)(0) = 20.92.\end{aligned}$$

Now we compute the conditional distribution of X_4 given $\boldsymbol{x}$. We note that

$$f_{X_4}(x_4 \mid \boldsymbol{x}) = \sum_{\theta=1}^{3} f_{X_4 \mid \Theta}(x_4 \mid \theta) f_{\Theta \mid X}(\theta \mid \boldsymbol{x}).$$

As $f_{\Theta \mid X}(3 \mid \boldsymbol{x}) = 0$, we have

$$f_{X_4}(10 \mid \boldsymbol{x}) = (0.2)(0.5844) + (0.4)(0.4156) = 0.2831,$$
$$f_{X_4}(20 \mid \boldsymbol{x}) = (0.3)(0.5844) + (0.4)(0.4156) = 0.3416,$$

and

$$f_{X_4}(30 \,|\, \boldsymbol{x}) = (0.5)(0.5844) + (0.2)(0.4156) = 0.3753.$$

Thus, the conditional mean of X_4 given $\boldsymbol{x}$ is

$$\mathrm{E}(X_4 \,|\, \boldsymbol{x}) = (10)(0.2831) + (20)(0.3416) + (30)(0.3753) = 20.92,$$

and the result

$$\mathrm{E}\left[\mu_X(\Theta) \,|\, \boldsymbol{x}\right] = \mathrm{E}(X_4 \,|\, \boldsymbol{x})$$

is verified. □

8.2 Conjugate distributions

A difficulty in applying the Bayes approach of statistical inference is the computation of the posterior pdf, which requires the computation of the marginal pdf of the data. However, as the Bayes estimate under squared-error loss is the mean of the posterior distribution, the estimate cannot be calculated unless the posterior pdf is known.

It turns out that there are classes of prior pdfs, which, together with specific likelihood functions, give rise to posterior pdfs that belong to the same class as the prior pdf. Such prior pdf and likelihood are said to be a **conjugate** pair. In Example 8.1, we see that if the prior pdf is beta and the likelihood is Bernoulli, the posterior pdf also follows a beta distribution, albeit with hyperparameters different from those of the prior pdf. In Example 8.3, we see that if the prior pdf is beta and the likelihood is binomial, then the posterior pdf is also beta, though with hyperparameters different from those of the prior. Thus, in these cases, the observed data $\boldsymbol{x}$ do not change the class of the prior, they only change the parameters of the prior.

A formal definition of **conjugate prior distribution** is as follows. Let the prior pdf of Θ be $f_\Theta(\theta \,|\, \gamma)$, where γ is the hyperparameter. The prior pdf $f_\Theta(\theta \,|\, \gamma)$ is conjugate to the likelihood function $f_{X \,|\, \Theta}(\boldsymbol{x} \,|\, \theta)$ if the posterior pdf is equal to $f_\Theta(\theta \,|\, \gamma^*)$, which has the same functional form as the prior pdf but, generally, a different hyperparameter γ^*. In other words, the prior and posterior belong to the same family of distributions.

We adopt the convention of "prior–likelihood" to describe the conjugate distribution. Thus, as shown in Examples 8.1 and 8.3, beta–Bernoulli and beta–binomial are conjugate distributions. We now present further examples of conjugate distributions which may be relevant for analyzing random losses. More conjugate distributions can be found in Table A.3 in the Appendix.

8.2.1 The gamma–Poisson conjugate distribution

Let $\boldsymbol{X} = \{X_1, X_2, \cdots, X_n\}$ be iid $\mathcal{PN}(\lambda)$. We assume $\Lambda \sim \mathcal{G}(\alpha, \beta)$. As shown in Appendix A.16.3, the posterior distribution of Λ is $\mathcal{G}(\alpha^*, \beta^*)$, where

$$\alpha^* = \alpha + n\bar{x} \tag{8.18}$$

and

$$\beta^* = \left[n + \frac{1}{\beta}\right]^{-1} = \frac{\beta}{n\beta + 1}. \tag{8.19}$$

Hence, the gamma prior pdf is conjugate to the Poisson likelihood.

8.2.2 The beta–geometric conjugate distribution

Let $\boldsymbol{X} = \{X_1, X_2, \cdots, X_n\}$ be iid $\mathcal{GM}(\theta)$. If the prior pdf of Θ is $\mathcal{B}(\alpha, \beta)$, then, as shown in Appendix A.16.4, the posterior distribution of Θ is $\mathcal{B}(\alpha^*, \beta^*)$, with

$$\alpha^* = \alpha + n \tag{8.20}$$

and

$$\beta^* = \beta + n\bar{x}, \tag{8.21}$$

so that the beta prior is conjugate to the geometric likelihood.

8.2.3 The gamma–exponential conjugate distribution

Let $\boldsymbol{X} = \{X_1, X_2, \cdots, X_n\}$ be iid $\mathcal{E}(\lambda)$. If the prior distribution of Λ is $\mathcal{G}(\alpha, \beta)$, then, as shown in Appendix A.16.5, the posterior distribution of Λ is $\mathcal{G}(\alpha^*, \beta^*)$, with

$$\alpha^* = \alpha + n \tag{8.22}$$

and

$$\beta^* = \left[\frac{1}{\beta} + n\bar{x}\right]^{-1} = \frac{\beta}{1 + \beta n\bar{x}}. \tag{8.23}$$

Thus, the gamma prior is conjugate to the exponential likelihood.

8.3 Bayesian versus Bühlmann credibility

If the prior distribution is conjugate to the likelihood, the Bayes estimate is easy to obtain. It turns out that for the conjugate distributions discussed in the last section, the Bühlmann credibility estimate is equal to the Bayes estimate. The examples below give the details of these results.

Example 8.7 (gamma–Poisson case) The claim-frequency random variable X is assumed to be distributed as $\mathcal{PN}(\lambda)$, and the prior distribution of Λ is $\mathcal{G}(\alpha, \beta)$. If a random sample of n observations of $\boldsymbol{X} = \{X_1, X_2, \cdots, X_n\}$ is available, derive the Bühlmann credibility estimate of the future claim frequency, and show that this is the same as the Bayes estimate.

Solution As $X_i \sim \text{iid } \mathcal{P}(\lambda)$, we have

$$\mu_{\text{PV}} = \text{E}[\sigma_X^2(\Lambda)] = \text{E}(\Lambda).$$

Since $\Lambda \sim \mathcal{G}(\alpha, \beta)$, we conclude that $\mu_{\text{PV}} = \alpha\beta$. Also, $\mu_X(\Lambda) = \text{E}(X \mid \Lambda) = \Lambda$, so that

$$\sigma_{\text{HM}}^2 = \text{Var}[\mu_X(\Lambda)] = \text{Var}(\Lambda) = \alpha\beta^2.$$

Thus

$$k = \frac{\mu_{\text{PV}}}{\sigma_{\text{HM}}^2} = \frac{1}{\beta},$$

and the Bühlmann credibility factor is

$$Z = \frac{n}{n+k} = \frac{n\beta}{n\beta + 1}.$$

The prior mean of the claim frequency is

$$M = \text{E}[\text{E}(X \mid \Lambda)] = \text{E}(\Lambda) = \alpha\beta.$$

Hence, we obtain the Bühlmann credibility estimate of future claim frequency as

$$\begin{aligned} U &= Z\bar{X} + (1 - Z)M \\ &= \frac{n\beta\bar{X}}{n\beta + 1} + \frac{\alpha\beta}{n\beta + 1} \\ &= \frac{\beta(n\bar{X} + \alpha)}{n\beta + 1}. \end{aligned}$$

The Bayes estimate of the expected claim frequency is the posterior mean of Λ. From Section 8.2.1, the posterior distribution of Λ is $\mathcal{G}(\alpha^*, \beta^*)$, where α^* and β^* are given in equations (8.18) and (8.19), respectively. Thus, the Bayes estimate of the expected claim frequency is

$$\begin{aligned}
\mathrm{E}(X_{n+1} \,|\, \boldsymbol{x}) &= \mathrm{E}[\mathrm{E}(X_{n+1} \mid \Lambda) \,|\, \boldsymbol{x}] \\
&= \mathrm{E}(\Lambda \,|\, \boldsymbol{x}) \\
&= \alpha^* \beta^* \\
&= (\alpha + n\bar{X}) \left[\frac{\beta}{n\beta + 1} \right] \\
&= U,
\end{aligned}$$

which is the Bühlmann credibility estimate. □

Example 8.8 (beta–geometric case) The claim-frequency random variable X is assumed to be distributed as $\mathcal{GM}(\theta)$, and the prior distribution of Θ is $\mathcal{B}(\alpha, \beta)$, where $\alpha > 2$. If a random sample of n observations of $\boldsymbol{X} = \{X_1, X_2, \cdots, X_n\}$ is available, derive the Bühlmann credibility estimate of the future claim frequency, and show that this is the same as the Bayes estimate.

Solution As $X_i \sim$ iid $\mathcal{G}(\theta)$, we have

$$\mu_X(\Theta) = \mathrm{E}(X \mid \Theta) = \frac{1 - \Theta}{\Theta},$$

and

$$\sigma_X^2(\Theta) = \mathrm{Var}(X \mid \Theta) = \frac{1 - \Theta}{\Theta^2}.$$

Assuming $\Theta \sim \mathcal{B}(\alpha, \beta)$, we first compute the following moments

$$\begin{aligned}
\mathrm{E}\left(\frac{1}{\Theta}\right) &= \int_0^1 \frac{1}{\theta} \left[\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)} \right] d\theta \\
&= \frac{B(\alpha - 1, \beta)}{B(\alpha, \beta)} \\
&= \frac{\alpha + \beta - 1}{\alpha - 1},
\end{aligned}$$

and

$$\mathrm{E}\left(\frac{1}{\Theta^2}\right) = \int_0^1 \frac{1}{\theta^2}\left[\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}\right] d\theta$$
$$= \frac{B(\alpha-2,\beta)}{B(\alpha,\beta)}$$
$$= \frac{(\alpha+\beta-1)(\alpha+\beta-2)}{(\alpha-1)(\alpha-2)}.$$

Hence, the expected value of the process variance is

$$\begin{aligned}
\mu_{\mathrm{PV}} &= \mathrm{E}[\sigma_X^2(\Theta)] \\
&= \mathrm{E}\left(\frac{1-\Theta}{\Theta^2}\right) \\
&= \mathrm{E}\left(\frac{1}{\Theta^2}\right) - \mathrm{E}\left(\frac{1}{\Theta}\right) \\
&= \frac{(\alpha+\beta-1)(\alpha+\beta-2)}{(\alpha-1)(\alpha-2)} - \frac{\alpha+\beta-1}{\alpha-1} \\
&= \frac{(\alpha+\beta-1)\beta}{(\alpha-1)(\alpha-2)},
\end{aligned}$$

and the variance of the hypothetical means is

$$\begin{aligned}
\sigma_{\mathrm{HM}}^2 &= \mathrm{Var}[\mu_X(\Theta)] \\
&= \mathrm{Var}\left(\frac{1-\Theta}{\Theta}\right) \\
&= \mathrm{Var}\left(\frac{1}{\Theta}\right) \\
&= \mathrm{E}\left(\frac{1}{\Theta^2}\right) - \left[\mathrm{E}\left(\frac{1}{\Theta}\right)\right]^2 \\
&= \frac{(\alpha+\beta-1)(\alpha+\beta-2)}{(\alpha-1)(\alpha-2)} - \left(\frac{\alpha+\beta-1}{\alpha-1}\right)^2 \\
&= \frac{(\alpha+\beta-1)\beta}{(\alpha-1)^2(\alpha-2)}.
\end{aligned}$$

Thus, the ratio of μ_{PV} to σ_{HM}^2 is

$$k = \frac{\mu_{\mathrm{PV}}}{\sigma_{\mathrm{HM}}^2} = \alpha - 1,$$

and the Bühlmann credibility factor is

$$Z = \frac{n}{n+k} = \frac{n}{n+\alpha-1}.$$

As the prior mean of X is

$$M = \mathrm{E}(X) = \mathrm{E}[\mathrm{E}(X \mid \Theta)] = \mathrm{E}\left(\frac{1-\Theta}{\Theta}\right) = \frac{\alpha+\beta-1}{\alpha-1} - 1 = \frac{\beta}{\alpha-1},$$

the Bühlmann credibility prediction of future claim frequency is

$$\begin{aligned} U &= Z\bar{X} + (1-Z)M \\ &= \frac{n\bar{X}}{n+\alpha-1} + \frac{\alpha-1}{n+\alpha-1}\left(\frac{\beta}{\alpha-1}\right) \\ &= \frac{n\bar{X}+\beta}{n+\alpha-1}. \end{aligned}$$

To compute the Bayes estimate of future claim frequency we note, from Section 8.2.2, that the posterior distribution of Θ is $\mathcal{B}(\alpha^*, \beta^*)$, where α^* and β^* are given in equations (8.20) and (8.21), respectively. Thus, we have

$$\begin{aligned} \mathrm{E}(X_{n+1} \mid \boldsymbol{x}) &= \mathrm{E}[\mathrm{E}(X_{n+1} \mid \Theta) \mid \boldsymbol{x}] \\ &= \mathrm{E}\left(\frac{1-\Theta}{\Theta} \mid \boldsymbol{x}\right) \\ &= \frac{\alpha^*+\beta^*-1}{\alpha^*-1} - 1 \\ &= \frac{\beta^*}{\alpha^*-1} \\ &= \frac{n\bar{X}+\beta}{n+\alpha-1}, \end{aligned}$$

which is the same as the Bühlmann credibility estimate. □

Example 8.9 (gamma–exponential case) The claim-severity random variable X is assumed to be distributed as $\mathcal{E}(\lambda)$, and the prior distribution of Λ is $\mathcal{G}(\alpha, \beta)$, where $\alpha > 2$. If a random sample of n observations of $\boldsymbol{X} = \{X_1, X_2, \cdots, X_n\}$ is available, derive the Bühlmann credibility estimate of the future claim severity, and show that this is the same as the Bayes estimate.

Solution As $X_i \sim$ iid $\mathcal{E}(\lambda)$, we have

$$\mu_X(\Lambda) = \mathrm{E}(X \mid \Lambda) = \frac{1}{\Lambda},$$

and

$$\sigma_X^2(\Lambda) = \mathrm{Var}(X \mid \Lambda) = \frac{1}{\Lambda^2}.$$

Since $\Lambda \sim \mathcal{G}(\alpha, \beta)$, the expected value of the process variance is

$$\begin{aligned}
\mu_{\mathrm{PV}} &= \mathrm{E}[\sigma_X^2(\Lambda)] \\
&= \mathrm{E}\left(\frac{1}{\Lambda^2}\right) \\
&= \int_0^\infty \frac{1}{\lambda^2} \left[\frac{\lambda^{\alpha-1} e^{-\frac{\lambda}{\beta}}}{\Gamma(\alpha)\beta^\alpha}\right] d\lambda \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty \lambda^{\alpha-3} e^{-\frac{\lambda}{\beta}}\, d\lambda \\
&= \frac{\Gamma(\alpha-2)\beta^{\alpha-2}}{\Gamma(\alpha)\beta^\alpha} \\
&= \frac{1}{(\alpha-1)(\alpha-2)\beta^2}.
\end{aligned}$$

The variance of the hypothetical means is

$$\sigma_{\mathrm{HM}}^2 = \mathrm{Var}[\mu_X(\Lambda)] = \mathrm{Var}\left(\frac{1}{\Lambda}\right) = \mathrm{E}\left(\frac{1}{\Lambda^2}\right) - \left[\mathrm{E}\left(\frac{1}{\Lambda}\right)\right]^2.$$

Now

$$\begin{aligned}
\mathrm{E}\left(\frac{1}{\Lambda}\right) &= \int_0^\infty \frac{1}{\lambda} \left[\frac{\lambda^{\alpha-1} e^{-\frac{\lambda}{\beta}}}{\Gamma(\alpha)\beta^\alpha}\right] d\lambda \\
&= \frac{\Gamma(\alpha-1)\beta^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \\
&= \frac{1}{(\alpha-1)\beta},
\end{aligned}$$

so that

$$\begin{aligned}
\sigma_{\mathrm{HM}}^2 &= \frac{1}{(\alpha-1)(\alpha-2)\beta^2} - \left[\frac{1}{(\alpha-1)\beta}\right]^2 \\
&= \frac{1}{(\alpha-1)^2(\alpha-2)\beta^2}.
\end{aligned}$$

Thus, we have

$$k = \frac{\mu_{\mathrm{PV}}}{\sigma^2_{\mathrm{HM}}} = \alpha - 1,$$

and the Bühlmann credibility factor is

$$Z = \frac{n}{n+k} = \frac{n}{n+\alpha-1}.$$

The prior mean of X is

$$M = \mathrm{E}\left[\mathrm{E}(X \mid \Lambda)\right] = \mathrm{E}\left(\frac{1}{\Lambda}\right) = \frac{1}{(\alpha-1)\beta}.$$

Hence we obtain the Bühlmann credibility estimate as

$$\begin{aligned} U &= Z\bar{X} + (1-Z)M \\ &= \frac{n\bar{X}}{n+\alpha-1} + \frac{\alpha-1}{n+\alpha-1}\left[\frac{1}{(\alpha-1)\beta}\right] \\ &= \frac{\beta n\bar{X}+1}{(n+\alpha-1)\beta}. \end{aligned}$$

To calculate the Bayes estimate, we note, from Section 8.2.3, that the posterior pdf of Λ is $\mathcal{G}(\alpha^*, \beta^*)$, where α^* and β^* are given in equations (8.22) and (8.23), respectively. Thus, the Bayes estimate of the expected claim severity is

$$\begin{aligned} \mathrm{E}(X_{n+1} \mid \boldsymbol{x}) &= \mathrm{E}\left(\frac{1}{\Lambda} \mid \boldsymbol{x}\right) \\ &= \frac{1}{(\alpha^*-1)\beta^*} \\ &= \frac{1+\beta n\bar{X}}{(\alpha+n-1)\beta} \\ &= U, \end{aligned}$$

and the equality of the Bühlmann estimate and the Bayes estimate is proven. □

We have shown that if the conjugate distributions discussed in the last section are used to model loss variables, where the distribution of the loss variable follows the likelihood function and the distribution of the risk parameters follows the conjugate prior, then the Bühlmann credibility estimate of the expected loss is equal to the Bayes estimate. In such cases, the Bühlmann credibility estimate is said to have **exact credibility**. Indeed, there are other conjugate distributions for which the Bühlmann credibility is *exact*. For example, the Bühlmann estimate for the case of normal–normal conjugate has exact credibility. In the next section, we discuss a general result for which the Bühlmann credibility estimate is exact.

8.4 Linear exponential family and exact credibility

Consider a random variable X with pdf or pf $f_{X \mid \Theta}(x \mid \theta)$, where θ is the parameter of the distribution. X is said to have a **linear exponential distribution** if $f_{X \mid \Theta}(x \mid \theta)$ can be written as

$$f_{X \mid \Theta}(x \mid \theta) = \exp\left[A(\theta)x + B(\theta) + C(x)\right], \tag{8.24}$$

for some functions $A(\theta)$, $B(\theta)$, and $C(x)$. By identifying these respective functions it is easy to show that some of the commonly used distributions in the actuarial science literature belong to the linear exponential family. Table 8.2 summarizes some of these distributions.[7]

Table 8.2. *Some linear exponential distributions*

Distribution	$\log f_{X \mid \Theta}(x \mid \theta)$	$A(\theta)$	$B(\theta)$	$C(x)$
Binomial, $\mathcal{BN}(m, \theta)$	$\log(C_x^m) + x \log \theta$ $+(m-x)\log(1-\theta)$	$\log \theta - \log(1-\theta)$	$m \log(1-\theta)$	$\log(C_x^m)$
Geometric, $\mathcal{GM}(\theta)$	$\log \theta + x \log(1-\theta)$	$\log(1-\theta)$	$\log \theta$	0
Poisson, $\mathcal{PN}(\theta)$	$x \log \theta - \theta - \log(x!)$	$\log \theta$	$-\theta$	$-\log(x!)$
Exponential, $\mathcal{E}(\theta)$	$-\theta x + \log \theta$	$-\theta$	$\log \theta$	0

If the likelihood function belongs to the linear exponential family, we can identify the prior distribution that is conjugate to the likelihood. Suppose the distribution of Θ has two hyperparameters, denoted by α and β. The **natural conjugate** of the likelihood given in equation (8.24) is

$$f_{\Theta}(\theta \mid \alpha, \beta) = \exp\left[A(\theta)a(\alpha, \beta) + B(\theta)b(\alpha, \beta) + D(\alpha, \beta)\right], \tag{8.25}$$

for some functions $a(\alpha, \beta)$, $b(\alpha, \beta)$, and $D(\alpha, \beta)$. To see this, we combine equations (8.24) and (8.25) to obtain the posterior pdf of Θ conditional on the

[7] For convenience we use θ to denote generically the parameters of the distributions instead of the usual notations (e.g. θ replaces λ for the parameter of the Poisson and exponential distributions). C_x^m in Table 8.2 denotes the combinatorial function.

sample $\boldsymbol{X} = \{X_1, X_2, \cdots, X_n\}$ as

$$f_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x}) \propto \exp\left\{A(\theta)\left[a(\alpha,\beta)+\sum_{i=1}^{n} x_i\right] + B(\theta)\left[b(\alpha,\beta)+n\right]\right\}. \quad (8.26)$$

Hence, the posterior pdf belongs to the same family as the prior pdf with parameters α^* and β^*, assuming they can be solved uniquely from the following equations

$$a(\alpha^*, \beta^*) = a(\alpha, \beta) + n\bar{x}, \quad (8.27)$$

and

$$b(\alpha^*, \beta^*) = b(\alpha, \beta) + n. \quad (8.28)$$

For the Poisson, geometric, and exponential likelihoods, we identify the functions $a(\alpha, \beta)$ and $b(\alpha, \beta)$, and summarize them in Table 8.3. The natural conjugate priors are then obtainable using equation (8.25).[8]

Table 8.3. *Examples of natural conjugate priors*

Conjugate distribution	$\log f_{\Theta}(\theta \mid \alpha, \beta)$	$a(\alpha, \beta)$	$b(\alpha, \beta)$
gamma–Poisson	$(\alpha - 1)\log\theta - \dfrac{\theta}{\beta} - \log[\Gamma(\alpha)\beta^{\alpha}]$	$\alpha - 1$	$\dfrac{1}{\beta}$
beta–geometric	$(\alpha - 1)\log\theta + (\beta - 1)\log(1-\theta)$ $- \log B(\alpha, \beta)$	$\beta - 1$	$\alpha - 1$
gamma–exponential	$(\alpha - 1)\log\theta - \dfrac{\theta}{\beta} - \log[\Gamma(\alpha)\beta^{\alpha}]$	$\dfrac{1}{\beta}$	$\alpha - 1$

The results above enable us to compute the prior pdf (up to a scaling factor) using equation (8.25). To illustrate the calculation of the posterior pdf, we consider the gamma–Poisson conjugate distribution. As $a(\alpha, \beta) = \alpha - 1$, we have, from equation (8.27)

$$\alpha^* - 1 = \alpha - 1 + n\bar{x}, \quad (8.29)$$

which implies

$$\alpha^* = \alpha + n\bar{x}. \quad (8.30)$$

[8] In Table 8.3, the function $D(\alpha, \beta)$, which defines the scaling factor, is ignored. The case of the binomial likelihood and its natural conjugate is left as an exercise (see Exercise 8.2).

Also, from equation (8.28), we have

$$\frac{1}{\beta^*} = \frac{1}{\beta} + n, \tag{8.31}$$

which implies

$$\beta^* = \left[\frac{1}{\beta} + n\right]^{-1}. \tag{8.32}$$

Thus, the results are the same as in equations (8.18) and (8.19).

Having seen some examples of linear exponential likelihoods and their natural conjugates, we now state a theorem which relates the Bühlmann credibility estimate to the Bayes estimate.

Theorem 8.1 *Let X be a random loss variable. If the likelihood of X belongs to the linear exponential family with parameter θ, and the prior distribution of Θ is the natural conjugate of the likelihood of X, then the Bühlmann credibility estimate of the mean of X is the same as the Bayes estimate.*

Proof See Klugman *et al.* (2004, Section 16.4.6), or Jewell (1974), for a proof of this theorem.[9] □

When the conditions of Theorem 8.1 hold, the Bühlmann credibility estimate is the same as the Bayes estimate and is said to have exact credibility. When the conditions of the theorem do not hold, the Bühlmann credibility estimator generally has a larger mean squared error than the Bayes estimator. The Bühlmann credibility estimator, however, still has the minimum mean squared error in the class of linear estimators based on the sample. The example below illustrates a comparison between the two methods, as well as the sample mean.

Example 8.10 Assume the claim frequency X over different periods are iid as $\mathcal{PN}(\lambda)$, and the prior pf of Λ is

$$\Lambda = \begin{cases} 1, & \text{with probability 0.5,} \\ 2, & \text{with probability 0.5.} \end{cases}$$

A random sample of $n = 6$ observations of X is available. Calculate the Bühlmann credibility estimate and the Bayes estimate of the expected claim frequency. Compare the mean squared errors of these estimates as well as that of the sample mean.

[9] Note that the proofs in these references require a transformation of the parameters of the prior pdf, so that the prior pdf is expressed in a form different from equation (8.24).

Solution The expected claim frequency is $\mathrm{E}(X) = \Lambda$. Thus, the mean squared error of the sample mean as an estimate of the expected claim frequency is

$$\begin{aligned}
\mathrm{E}\left[(\bar{X} - \Lambda)^2\right] &= \mathrm{E}\left\{\mathrm{E}\left[(\bar{X} - \Lambda)^2 \mid \Lambda\right]\right\} \\
&= \mathrm{E}\left\{\left[\mathrm{Var}(\bar{X} \mid \Lambda)\right]\right\} \\
&= \mathrm{E}\left[\frac{\mathrm{Var}(X \mid \Lambda)}{n}\right] \\
&= \frac{\mathrm{E}(\Lambda)}{n} \\
&= \frac{1.5}{6} \\
&= 0.25.
\end{aligned}$$

We now derive the Bühlmann credibility estimator. As $\mu_X(\Lambda) = \mathrm{E}(X \mid \Lambda) = \Lambda$ and $\sigma_X^2(\Lambda) = \mathrm{Var}(X \mid \Lambda) = \Lambda$, we have

$$\mu_{\mathrm{PV}} = \mathrm{E}[\sigma_X^2(\Lambda)] = \mathrm{E}(\Lambda) = 1.5,$$

and

$$\sigma_{\mathrm{HM}}^2 = \mathrm{Var}[\mu_X(\Lambda)] = \mathrm{Var}(\Lambda) = (0.5)(1 - 1.5)^2 + (0.5)(2 - 1.5)^2 = 0.25.$$

Thus, we have

$$k = \frac{\mu_{\mathrm{PV}}}{\sigma_{\mathrm{HM}}^2} = \frac{1.5}{0.25} = 6,$$

and the Bühlmann credibility factor is

$$Z = \frac{n}{n + 6} = \frac{6}{6 + 6} = 0.5.$$

As the prior mean of X is

$$M = \mathrm{E}[\mathrm{E}(X \mid \Lambda)] = \mathrm{E}(\Lambda) = 1.5,$$

the Bühlmann credibility estimator is

$$U = Z\bar{X} + (1 - Z)M = 0.5\bar{X} + (0.5)(1.5) = 0.5\bar{X} + 0.75.$$

Given $\Lambda = \lambda$, the expected values of the sample mean and the Bühlmann credibility estimator are, respectively, λ and $0.5\lambda + 0.75$. Thus, the sample mean is an unbiased estimator of λ, while the Bühlmann credibility estimator is generally not. However, when λ varies as a random variable the expected value of the Bühlmann credibility estimator is equal to 1.5, which is the prior mean of X, and so is the expected value of the sample mean.

The mean squared error of the Bühlmann credibility estimate of the expected value of X is computed as follows

$$
\begin{aligned}
\mathrm{E}\left\{[U - \mathrm{E}(X)]^2\right\} &= \mathrm{E}\left[(0.5\bar{X} + 0.75 - \Lambda)^2\right] \\
&= \mathrm{E}\left\{\mathrm{E}\left[(0.5\bar{X} + 0.75 - \Lambda)^2 \mid \Lambda\right]\right\} \\
&= \mathrm{E}\Big\{\mathrm{E}\Big[0.25\bar{X}^2 + (0.75)^2 + \Lambda^2 + 0.75\bar{X} \\
&\quad - 1.5\Lambda - \Lambda\bar{X} \mid \Lambda\Big]\Big\} \\
&= \mathrm{E}\Big[0.25(\mathrm{Var}(\bar{X} \mid \Lambda) + \left[\mathrm{E}(\bar{X} \mid \Lambda)\right]^2) \\
&\quad + \mathrm{E}\left\{(0.75)^2 + \Lambda^2 + 0.75\bar{X} - 1.5\Lambda - \Lambda\bar{X} \mid \Lambda\right\}\Big] \\
&= \mathrm{E}\Big[0.25\left(\frac{\Lambda}{6} + \Lambda^2\right) + (0.75)^2 + \Lambda^2 + 0.75\Lambda \\
&\quad - 1.5\Lambda - \Lambda^2\Big] \\
&= \mathrm{E}\left[0.25\left(\frac{\Lambda}{6} + \Lambda^2\right) + (0.75)^2 - 0.75\Lambda\right] \\
&= \mathrm{E}\left(0.25\Lambda^2 - 0.7083\Lambda + 0.5625\right) \\
&= 0.25\left[(1)(0.5) + (2)^2(0.5)\right] - (0.7083)(1.5) + 0.5625 \\
&= 0.1251.
\end{aligned}
$$

Hence, the mean squared error of the Bühlmann credibility estimator is about half of that of the sample mean.

As the Bayes estimate is the posterior mean, we first derive the posterior pf of Λ. The marginal pf of $\boldsymbol{X}$ is

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}) &= \sum_{\lambda \in \{1, 2\}} f_{\boldsymbol{X} \mid \Lambda}(\boldsymbol{x} \mid \lambda) \Pr(\Lambda = \lambda) \\
&= 0.5\left[\sum_{\lambda \in \{1, 2\}} \left(\prod_{i=1}^{6} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= 0.5\left[\left(\frac{1}{e^6}\prod_{i=1}^{6}\frac{1}{x_i!}\right) + \left(\frac{1}{e^{12}}\prod_{i=1}^{6}\frac{2^{x_i}}{x_i!}\right)\right] \\
&= K, \qquad \text{say.}
\end{aligned}
$$

Thus, the posterior pdf of Λ is

$$
f_{\Lambda\,|\,\boldsymbol{X}}(\lambda\,|\,\boldsymbol{x}) = \begin{cases} \dfrac{0.5}{e^6 K}\left(\displaystyle\prod_{i=1}^{6}\frac{1}{x_i!}\right), & \text{for } \lambda = 1, \\ \dfrac{0.5}{e^{12} K}\left(\displaystyle\prod_{i=1}^{6}\frac{2^{x_i}}{x_i!}\right), & \text{for } \lambda = 2. \end{cases}
$$

The posterior mean of Λ is

$$
\mathrm{E}(\Lambda\,|\,\boldsymbol{x}) = \frac{0.5}{e^6 K}\left(\prod_{i=1}^{6}\frac{1}{x_i!}\right) + \left(\frac{1}{e^{12} K}\prod_{i=1}^{6}\frac{2^{x_i}}{x_i!}\right).
$$

Thus, the Bayes estimate is a highly nonlinear function of the data, and the computation of its mean squared error is intractable. We estimate the mean squared error using simulation as follows:[10]

1 Generate λ with value of 1 or 2 with probability of 0.5 each.
2 Using the value of λ generated in Step 1, generate six observations of X, $x_1, \cdots, x_6$, from the distribution $\mathcal{PN}(\lambda)$.
3 Compute the posterior mean of Λ of this sample using the expression

$$
\frac{0.5}{e^6 K}\left(\prod_{i=1}^{6}\frac{1}{x_i!}\right) + \left(\frac{1}{e^{12} K}\prod_{i=1}^{6}\frac{2^{x_i}}{x_i!}\right).
$$

4 Repeat Steps 1 through 3 m times. Denote the values of λ generated in Step 1 by $\lambda_1, \cdots, \lambda_m$, and the corresponding Bayes estimates computed in Step 3 by $\hat{\lambda}_1, \cdots, \hat{\lambda}_m$. The estimated mean squared error of the Bayes estimate is

$$
\frac{1}{m}\sum_{i=1}^{m}(\hat{\lambda}_i - \lambda_i)^2.
$$

We perform a simulation with $m = 100{,}000$ runs. The estimated mean squared error is 0.1103. Thus, the mean squared error of the Bayes estimate is lower than that of the Bühlmann credibility estimate (0.1251), which is in turn lower than that of the sample mean (0.25). □

[10] The methodology of simulation is covered in detail in Chapters 14 and 15.

8.5 Summary and discussions

The prediction of future random losses can be usefully formulated under the Bayesian framework. Suppose the random loss variable X has a mean $\mathrm{E}(X \mid \Theta) = \mu_X(\Theta)$, and a random sample of $\boldsymbol{X} = \{X_1, X_2, \cdots, X_n\}$ is available. The Bayesian premium is equal to $\mathrm{E}[\mu_X(\Theta) \mid \boldsymbol{x}]$, which is also equal to $\mathrm{E}[X_{n+1} \mid \boldsymbol{x}]$. The former is the Bayes estimate of the expected loss, and the latter is the Bayes predictor of future loss. The Bayes estimate (prediction) is the posterior mean of the expected loss (posterior mean of the future loss), and it has the minimum mean squared error among all estimators of the expected loss.

The Bühlmann credibility estimate is the minimum mean squared error estimate in the class of estimators that are linear in $\boldsymbol{X}$. When the likelihood belongs to the linear exponential family and the prior distribution is the natural conjugate, the Bühlmann credibility estimate is equal to the Bayes estimate. However, in other situations the performance of the Bühlmann credibility estimate is generally inferior to the Bayes estimate.

While the Bayes estimate has optimal properties, its computation is generally complicated. In practical applications, the posterior mean of the expected loss may not be analytically available and has to be computed numerically.

Exercises

8.1 Let $X \sim \mathcal{BN}(m, \theta)$ and $\boldsymbol{X} = \{X_1, \cdots, X_n\}$ be a random sample of X. Assume Θ follows a beta distribution with hyperparameters α and β.
(a) Calculate the posterior mean of $\mu_X(\Theta) = \mathrm{E}(X \mid \Theta)$.
(b) What is the conditional pf of X_{n+1} given the sample data $\boldsymbol{x}$?

8.2 Let $X \sim \mathcal{BN}(m, \theta)$ and $\Theta \sim \mathcal{B}(\alpha, \beta)$. Given the functions $A(\theta)$ and $B(\theta)$ for the $\mathcal{BN}(m, \theta)$ distribution in Table 8.2, identify the functions $a(\alpha, \beta)$ and $b(\alpha, \beta)$ as defined in equation (8.25) so that the $\mathcal{B}(\alpha, \beta)$ distribution is a natural conjugate of $\mathcal{BN}(m, \theta)$. Hence, derive the hyperparameters of the posterior distribution of Θ using equations (8.27) and (8.28).

8.3 In Example 8.10, the mean squared error of the expected loss is analytically derived for the sample mean and the Bühlmann credibility estimate, and numerically estimated for the Bayes estimate. Derive the mean squared errors of the sample mean and the Bühlmann premium as predictors for future loss X_{n+1}. Suggest a simulation procedure for the estimation of the mean squared error of the Bayes predictor for future loss.

8.4 Given $\Theta = \theta$, $X \sim \mathcal{NB}(r, \theta)$. If the prior distribution of Θ is $\mathcal{B}(\alpha, \beta)$, determine the unconditional pf of X.

8.5 Show that the negative binomial distribution $\mathcal{NB}(r, \theta)$ belongs to the linear exponential family, where r is known and θ is the unknown parameter. Identify the functions $A(\theta)$, $B(\theta)$, and $C(x)$ in equation (8.24).

8.6 Given N, X is distributed as $\mathcal{BN}(N, \theta)$. Derive the unconditional distribution of X assuming N is distributed as (a) $\mathcal{PN}(\lambda)$, and (b) $\mathcal{BN}(m, \beta)$, where m is known.

Questions adapted from SOA exams

8.7 The pf of the annual number of claims N of a particular insurance policy is: $f_N(0) = 2\theta$, $f_N(1) = \theta$, and $f_N(2) = 1 - 3\theta$. Over different policies, the pf of Θ is: $f_\Theta(0.1) = 0.8$ and $f_\Theta(0.3) = 0.2$. If there is one claim in Year 1, calculate the Bayes estimate of the expected number of claims in Year 2.

8.8 In a portfolio of insurance policies, the number of claims for each policyholder in each year, denoted by N, may be 0, 1, or 2, with the following pf: $f_N(0) = 0.1$, $f_N(1) = 0.9 - \theta$, and $f_N(2) = \theta$. The prior pdf of Θ is

$$f_\Theta(\theta) = \frac{\theta^2}{0.039}, \qquad 0.2 < \theta < 0.5.$$

A randomly selected policyholder has two claims in Year 1 and two claims in Year 2. Determine the Bayes estimate of the expected number of claims in Year 3 of this policyholder.

8.9 The number of claims N of each policy is distributed as $\mathcal{BN}(8, \theta)$, and the prior distribution of Θ is $\mathcal{B}(\alpha, 9)$. A randomly selected policyholder is found to have made two claims in Year 1 and k claims in Year 2. The Bayesian credibility estimate of the expected number of claims in Year 2 based on the experience of Year 1 is 2.54545, and the Bayesian credibility estimate of the expected number of claims in Year 3 based on the experience of Years 1 and 2 is 3.73333. Determine k.

8.10 Claim severity is distributed as $\mathcal{E}(1/\theta)$. The prior distribution of Θ is inverse gamma with pdf

$$f_\Theta(\theta) = \frac{c^2}{\theta^3} \exp\left(-\frac{c}{\theta}\right), \qquad 0 < \theta < \infty, \; 0 < c.$$

Given an observed loss is x, calculate the mean of the posterior distribution of Θ.

8.11 Consider two random variables D and G, where

$$\Pr(D = d \,|\, G = g) = g^{1-d}(1 - g)^d, \qquad d = 0, 1,$$

and

$$\Pr\left(G = \frac{1}{5}\right) = \frac{3}{5} \qquad \text{and} \qquad \Pr\left(G = \frac{1}{3}\right) = \frac{2}{5}.$$

Calculate

$$\Pr\left(G = \frac{1}{3} \,|\, D = 0\right).$$

8.12 A portfolio has 100 independently and identically distributed risks. The number of claims of each risk follows a $\mathcal{PN}(\lambda)$ distribution. The prior pdf of Λ is $\mathcal{G}(4, 0.02)$. In Year 1, the following loss experience is observed

Number of claims	Number of risks
0	90
1	7
2	2
3	1
Total	100

Determine the Bayesian expected number of claims of the portfolio in Year 2.

8.13 Claim severity X is distributed as $\mathcal{E}(1/\theta)$. It is known that 80% of the policies have $\theta = 8$ and the other 20% have $\theta = 2$. A randomly selected policy has one claim of size 5. Calculate the Bayes expected size of the next claim of this policy.

8.14 The claim frequency N in a period is distributed as $\mathcal{PN}(\lambda)$, where the prior distribution of Λ is $\mathcal{E}(1)$. If a policyholder makes no claim in a period, determine the posterior pdf of Λ for this policyholder.

8.15 Annual claim frequencies follow a Poisson distribution with mean λ. The prior distribution of Λ has pdf

$$f_\Lambda(\lambda) = 0.4\left(\frac{1}{6}e^{-\frac{\lambda}{6}}\right) + 0.6\left(\frac{1}{12}e^{-\frac{\lambda}{12}},\right), \qquad \lambda > 0.$$

Ten claims are observed for an insured in Year 1. Determine the Bayesian expected number of claims for the insured in Year 2.

8.16 The annual number of claims for a policyholder follows a Poisson distribution with mean λ. The prior distribution of Λ is $\mathcal{G}(5, 0.5)$.

A randomly selected insured has five claims in Year 1 and three claims in Year 2. Determine the posterior mean of Λ.

8.17 The annual number of claims of a given policy is distributed as $\mathcal{GM}(\theta)$. One third of the policies have $\theta = 1/3$ and the remaining two-thirds have $\theta = 1/6$. A randomly selected policy had two claims in Year 1. Calculate the Bayes expected number of claims for the selected policy in Year 2.

8.18 An insurance company sells three types of policies with the following characteristics

Type of policy	Proportion of total policies	Distribution of annual claim frequency
A	5%	$\mathcal{PN}(0.25)$
B	20%	$\mathcal{PN}(0.50)$
C	75%	$\mathcal{PN}(1.00)$

A randomly selected policy is observed to have one claim in each of Years 1 through 4. Determine the Bayes estimate of the expected number of claims of this policyholder in Year 5.

8.19 The annual number of claims for a policyholder is distributed as $\mathcal{BN}(2, \theta)$. The prior distribution of Θ has pdf $f_\Theta(\theta) = 4\theta^3$ for $0 < \theta < 1$. This policyholder had one claim in each of Years 1 and 2. Determine the Bayes estimate of the expected number of claims in Year 3.

8.20 Claim sizes follow the $\mathcal{P}(1, \gamma)$ distribution. Half of the policies have $\gamma = 1$, while the other half have $\gamma = 3$. For a randomly selected policy, the claim in Year 1 was 5. Determine the posterior probability that the claim amount of the policy in Year 2 will exceed 8.

8.21 The probability that an insured will have at least one loss during any year is θ. The prior distribution of Θ is $\mathcal{U}(0, 0.5)$. An insured had at least one loss every year in the last eight years. Determine the posterior probability that the insured will have at least one loss in Year 9.

8.22 The probability that an insured will have exactly one claim is θ. The prior distribution of Θ has pdf

$$f_\Theta(\theta) = \frac{3\sqrt{\theta}}{2}, \qquad 0 < \theta < 1.$$

A randomly selected insured is found to have exactly one claim. Determine the posterior probability of $\theta > 0.6$.

8.23 For a group of insureds, the claim size is distributed as $\mathcal{U}(0, \theta)$, where $\theta > 0$. The prior distribution of Θ has pdf

$$f_\Theta(\theta) = \frac{500}{\theta^2}, \qquad \theta > 500.$$

If two independent claims of amounts 400 and 600 are observed, calculate the probability that the next claim will exceed 550.

8.24 The annual number of claims of each policyholder is distributed as $\mathcal{PN}(\lambda)$. The prior distribution of Λ is $\mathcal{G}(2, 1)$. If a randomly selected policyholder had at least one claim last year, determine the posterior probability that this policyholder will have at least one claim this year.

9

Empirical implementation of credibility

We have discussed the limited-fluctuation credibility method, the Bühlmann and Bühlmann–Straub credibility methods, as well as the Bayesian method for future loss prediction. The implementation of these methods requires the knowledge or assumptions of some unknown parameters of the model. For the limited-fluctuation credibility method, Poisson distribution is usually assumed for claim frequency. In addition, we need to know the coefficient of variation of claim severity if predictions of claim severity or aggregate loss/pure premium are required. For the Bühlmann and Bühlmann–Straub methods, the key quantities required are the expected value of the process variance, μ_{PV}, and the variance of the hypothetical means, σ^2_{HM}. These quantities depend on the assumptions of the prior distribution of the risk parameters and the conditional distribution of the random loss variable. For the Bayesian method, the predicted loss can be obtained relatively easily if the prior distribution is conjugate to the likelihood. Yet the posterior mean, which is the Bayesian predictor of the future loss, depends on the hyperparameters of the posterior distribution. Thus, for the empirical implementation of the Bayesian method, the hyperparameters have to be estimated.

In this chapter, we discuss the estimation of the required parameters for the implementation of the credibility estimates. We introduce the empirical Bayes method, which may be nonparametric, semiparametric, or parametric, depending on the assumptions concerning the prior distribution and the likelihood. Our main focus is on the Bühlmann and Bühlmann–Straub credibility models, the nonparametric implementation of which is relatively straightforward.

Learning objectives

1 Empirical Bayes method
2 Nonparametric estimation

3 Semiparametric estimation
4 Parametric estimation

9.1 Empirical Bayes method

Implementation of the credibility estimates requires the knowledge of some unknown parameters in the model. For the limited-fluctuation method, depending on the loss variable of interest, the mean and/or the variance of the loss variable are required. For example, to determine whether full credibility is attained for the prediction of claim frequency, we need to know λ_N, which can be estimated by the sample mean of the claim frequency.[1] For predicting claim severity and aggregate loss/pure premium, the coefficient of variation of the loss variable, C_X, is also required, which may be estimated by

$$\hat{C}_X = \frac{s_X}{\bar{X}}, \tag{9.1}$$

where s_X and $\bar{X}$ are the sample standard deviation and sample mean of X, respectively.

In the Bühlmann and Bühlmann–Straub frameworks, the key quantities of interest are the expected value of the process variance, μ_{PV}, and the variance of the hypothetical means, σ^2_{HM}. These quantities can be derived from the Bayesian framework and depend on both the prior distribution and the likelihood. In a strictly Bayesian approach, the prior distribution is given and inference is drawn based on the given prior. For practical applications when researchers are not in a position to state the prior, empirical methods may be applied to estimate the hyperparameters. This is called the **empirical Bayes method**. Depending on the assumptions about the prior distribution and the likelihood, empirical Bayes estimation may adopt one of the following approaches:[2]

1 **Nonparametric approach:** In this approach, no assumptions are made about the particular forms of the prior density of the risk parameters $f_\Theta(\theta)$ and the conditional density of the loss variable $f_{X \mid \Theta}(x \mid \theta)$. The method is very general and applies to a wide range of models.
2 **Semiparametric approach:** In some practical applications, prior experience may suggest a particular distribution for the loss variable X, while the specification of the prior distribution remains elusive. In such cases, parametric assumptions concerning $f_{X \mid \Theta}(x \mid \theta)$ may be made, while the prior distribution of the risk parameters $f_\Theta(\theta)$ remains unspecified.

[1] Refer to the assumptions for full credibility and its derivation in the limited-fluctuation approach in Section 6.2.1. Recall that λ_N is the mean of the claim frequency, which is assumed to be Poisson.
[2] Further discussions of parametric versus nonparametric estimation can be found in Chapter 10.

3 **Parametric approach:** When the researcher makes specific assumptions about $f_{X \mid \Theta}(x \mid \theta)$ and $f_{\Theta}(\theta)$, the estimation of the parameters in the model may be carried out using the maximum likelihood estimation (MLE) method. The properties of these estimators follow the classical results of MLE, as discussed in Appendix A.19 and Chapter 12. While in some cases the MLE can be derived analytically, in many situations they have to be computed numerically.

9.2 Nonparametric estimation

To implement the limited-fluctuation credibility prediction for claim severity and aggregate loss/pure premium, an estimate of the coefficient of variation C_X is required. $\hat{C}_X$ as defined in equation (9.1) is an example of a nonparametric estimator. Note that under the assumption of a random sample, s_X and $\bar{X}$ are consistent estimators for the population standard deviation and the population mean, respectively, irrespective of the actual distribution of the random loss variable X. Thus, $\hat{C}_X$ is a consistent estimator for C_X, although it is generally not unbiased.[3]

For the implementation of the Bühlmann and Bühlmann–Straub credibility models, the key quantities required are the expected value of the process variance, μ_{PV}, and the variance of the hypothetical means, σ^2_{HM}, which together determine the Bühlmann credibility parameter k. We present below unbiased estimates of these quantities. To the extent that the unbiasedness holds under the mild assumption that the loss observations are statistically independent, and that no specific assumption is made about the likelihood of the loss random variables and the prior distribution of the risk parameters, the estimates are nonparametric.

In Section 7.4 we set up the Bühlmann–Straub credibility model with a sample of loss observations from a risk group. We shall extend this set-up to consider multiple risk groups, each with multiple samples of loss observations over possibly different periods. The results in this set-up will then be specialized to derive results for the situations discussed in Chapter 7. We now formally state the assumptions of the extended set-up as follows:

1 Let X_{ij} denote the loss per unit of exposure and m_{ij} denote the amount of exposure. The index i denotes the ith risk group, for $i = 1, \cdots, r$, with $r > 1$. Given i, the index j denotes the jth loss observation in the ith group, for $j = 1, \cdots, n_i$, where $n_i > 1$ for $i = 1, \cdots, r$. The number of loss observations n_i in each risk group may differ. We may think of j as indexing

[3] Properties of estimators will be discussed in Chapter 10.

an individual within the risk group or a period of the risk group. Thus, for the ith risk group we have loss observations of n_i individuals or periods.

2 X_{ij} are assumed to be independently distributed. The risk parameter of the ith group is denoted by θ_i, which is a realization of the random variable Θ_i. We assume Θ_i to be independently and identically distributed as Θ.

3 The following assumptions are made for the hypothetical means and the process variance

$$\mathrm{E}(X_{ij} \mid \Theta = \theta_i) = \mu_X(\theta_i), \qquad \text{for } i = 1, \cdots, r;\ j = 1, \cdots, n_i, \tag{9.2}$$

and

$$\mathrm{Var}(X_{ij} \mid \theta_i) = \frac{\sigma_X^2(\theta_i)}{m_{ij}}, \qquad \text{for } i = 1, \cdots, r;\ j = 1, \cdots, n_i. \tag{9.3}$$

We define the overall mean of the loss variable as

$$\mu_X = \mathrm{E}[\mu_X(\Theta_i)] = \mathrm{E}[\mu_X(\Theta)], \tag{9.4}$$

the mean of the process variance as

$$\mu_{\mathrm{PV}} = \mathrm{E}[\sigma_X^2(\Theta_i)] = \mathrm{E}[\sigma_X^2(\Theta)], \tag{9.5}$$

and the variance of the hypothetical means as

$$\sigma_{\mathrm{HM}}^2 = \mathrm{Var}[\mu_X(\Theta_i)] = \mathrm{Var}[\mu_X(\Theta)]. \tag{9.6}$$

For future reference, we also define the following quantities

$$m_i = \sum_{j=1}^{n_i} m_{ij}, \qquad \text{for } i = 1, \cdots, r, \tag{9.7}$$

which is the total exposure for the ith risk group; and

$$m = \sum_{i=1}^{r} m_i, \tag{9.8}$$

which is the total exposure over all risk groups. Also, we define

$$\bar{X}_i = \frac{1}{m_i} \sum_{j=1}^{n_i} m_{ij} X_{ij}, \qquad \text{for } i = 1, \cdots, r, \tag{9.9}$$

as the exposure-weighted mean of the ith risk group; and

$$\bar{X} = \frac{1}{m} \sum_{i=1}^{r} m_i \bar{X}_i \tag{9.10}$$

as the overall weighted mean.

The Bühlmann–Straub credibility predictor of the loss in the next period or a random individual of the ith risk group is

$$Z_i\bar{X}_i + (1 - Z_i)\mu_X, \tag{9.11}$$

where

$$Z_i = \frac{m_i}{m_i + k}, \tag{9.12}$$

with

$$k = \frac{\mu_{\text{PV}}}{\sigma^2_{\text{HM}}}. \tag{9.13}$$

To implement the credibility prediction, we need to estimate μ_X, μ_{PV}, and σ^2_{HM}. It is natural to estimate μ_X by $\bar{X}$. To show that $\bar{X}$ is an unbiased estimator of μ_X, we first note that $\text{E}(X_{ij}) = \text{E}[\text{E}(X_{ij} \mid \Theta_i)] = \text{E}[\mu_X(\Theta_i)] = \mu_X$, so that

$$\begin{aligned}
\text{E}(\bar{X}_i) &= \frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij}\text{E}(X_{ij}) \\
&= \frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij}\mu_X \\
&= \mu_X, \qquad \text{for } i = 1, \cdots, r.
\end{aligned} \tag{9.14}$$

Thus, we have

$$\begin{aligned}
\text{E}(\bar{X}) &= \frac{1}{m}\sum_{i=1}^{r} m_i\text{E}(\bar{X}_i) \\
&= \frac{1}{m}\sum_{i=1}^{r} m_i\mu_X \\
&= \mu_X,
\end{aligned} \tag{9.15}$$

so that $\bar{X}$ is an unbiased estimator of μ_X.

We now present an unbiased estimator of μ_{PV} in the following theorem.

Theorem 9.1 *The following quantity is an unbiased estimator of* μ_{PV}

$$\hat{\mu}_{\text{PV}} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}(X_{ij} - \bar{X}_i)^2}{\sum_{i=1}^{r}(n_i - 1)}. \tag{9.16}$$

Proof We re-arrange the inner summation term in the numerator of equation (9.16) to obtain

$$\begin{aligned}\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2 &= \sum_{j=1}^{n_i} m_{ij}\left\{[X_{ij}-\mu_X(\theta_i)]-[\bar{X}_i-\mu_X(\theta_i)]\right\}^2 \\ &= \sum_{j=1}^{n_i} m_{ij}[X_{ij}-\mu_X(\theta_i)]^2 + \sum_{j=1}^{n_i} m_{ij}[\bar{X}_i-\mu_X(\theta_i)]^2 \\ &\quad - 2\sum_{j=1}^{n_i} m_{ij}[X_{ij}-\mu_X(\theta_i)][\bar{X}_i-\mu_X(\theta_i)]. \end{aligned} \tag{9.17}$$

Simplifying the last two terms on the right-hand side of the above equation, we have

$$\begin{aligned}&\sum_{j=1}^{n_i} m_{ij}[\bar{X}_i-\mu_X(\theta_i)]^2 - 2\sum_{j=1}^{n_i} m_{ij}[X_{ij}-\mu_X(\theta_i)][\bar{X}_i-\mu_X(\theta_i)] \\ &= m_i[\bar{X}_i-\mu_X(\theta_i)]^2 - 2[\bar{X}_i-\mu_X(\theta_i)]\sum_{j=1}^{n_i} m_{ij}[X_{ij}-\mu_X(\theta_i)] \\ &= m_i[\bar{X}_i-\mu_X(\theta_i)]^2 - 2m_i[\bar{X}_i-\mu_X(\theta_i)]^2 \\ &= -m_i[\bar{X}_i-\mu_X(\theta_i)]^2. \end{aligned} \tag{9.18}$$

Combining equations (9.17) and (9.18), we obtain

$$\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2 = \left[\sum_{j=1}^{n_i} m_{ij}[X_{ij}-\mu_X(\theta_i)]^2\right] - m_i[\bar{X}_i-\mu_X(\theta_i)]^2. \tag{9.19}$$

We now take expectations of the two terms on the right-hand side of the above. First, we have

$$
\begin{aligned}
\mathrm{E}\left[\sum_{j=1}^{n_i} m_{ij}[X_{ij} - \mu_X(\Theta_i)]^2\right] &= \mathrm{E}\left[\mathrm{E}\left(\sum_{j=1}^{n_i} m_{ij}[X_{ij} - \mu_X(\Theta_i)]^2 \mid \Theta_i\right)\right] \\
&= \mathrm{E}\left[\sum_{j=1}^{n_i} m_{ij}\mathrm{Var}(X_{ij} \mid \Theta_i)\right] \\
&= \mathrm{E}\left[\sum_{j=1}^{n_i} m_{ij}\left[\frac{\sigma_X^2(\Theta_i)}{m_{ij}}\right]\right] \\
&= \sum_{j=1}^{n_i} \mathrm{E}[\sigma_X^2(\Theta_i)] \\
&= n_i\, \mu_{\mathrm{PV}},
\end{aligned}
\tag{9.20}
$$

and, noting that $\mathrm{E}(\bar{X}_i \mid \Theta_i) = \mu_X(\Theta_i)$, we have

$$
\begin{aligned}
\mathrm{E}\left\{m_i[\bar{X}_i - \mu_X(\Theta_i)]^2\right\} &= m_i\,\mathrm{E}\left[\mathrm{E}\left\{[\bar{X}_i - \mu_X(\Theta_i)]^2 \mid \Theta_i\right\}\right] \\
&= m_i\,\mathrm{E}\left[\mathrm{Var}(\bar{X}_i \mid \Theta_i)\right] \\
&= m_i\,\mathrm{E}\left[\mathrm{Var}\left(\frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij}X_{ij} \mid \Theta_i\right)\right] \\
&= m_i\,\mathrm{E}\left[\frac{1}{m_i^2}\sum_{j=1}^{n_i} m_{ij}^2\,\mathrm{Var}(X_{ij} \mid \Theta_i)\right] \\
&= m_i\,\mathrm{E}\left[\frac{1}{m_i^2}\sum_{j=1}^{n_i} m_{ij}^2\left(\frac{\sigma_X^2(\Theta_i)}{m_{ij}}\right)\right] \\
&= \frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij}\,\mathrm{E}[\sigma_X^2(\Theta_i)] \\
&= \mathrm{E}\left[\sigma_X^2(\Theta_i)\right] \\
&= \mu_{\mathrm{PV}}.
\end{aligned}
\tag{9.21}
$$

Combining equations (9.19), (9.20), and (9.21), we conclude that

$$\mathrm{E}\left[\sum_{j=1}^{n_i} m_{ij}(X_{ij} - \bar{X}_i)^2\right] = n_i\mu_{\mathrm{PV}} - \mu_{\mathrm{PV}} = (n_i - 1)\mu_{\mathrm{PV}}. \tag{9.22}$$

Thus, taking expectation of equation (9.16), we have

$$\begin{aligned} \mathrm{E}\left(\hat{\mu}_{\mathrm{PV}}\right) &= \frac{\sum_{i=1}^{r} \mathrm{E}\left[\sum_{j=1}^{n_i} m_{ij}(X_{ij} - \bar{X}_i)^2\right]}{\sum_{i=1}^{r}(n_i - 1)} \\ &= \frac{\sum_{i=1}^{r}(n_i - 1)\mu_{\mathrm{PV}}}{\sum_{i=1}^{r}(n_i - 1)} \\ &= \mu_{\mathrm{PV}}, \end{aligned} \tag{9.23}$$

so that $\hat{\mu}_{\mathrm{PV}}$ is an unbiased estimator of μ_{PV}. □

Note that equation (9.22) shows that

$$\hat{\sigma}_i^2 = \frac{1}{(n_i - 1)}\left[\sum_{j=1}^{n_i} m_{ij}(X_{ij} - \bar{X}_i)^2\right] \tag{9.24}$$

is also an unbiased estimator of μ_{PV}, for $i = 1, \cdots, r$. These estimators, however, make use of data in the ith risk group only, and are thus not as efficient as $\hat{\mu}_{\mathrm{PV}}$. In contrast, $\hat{\mu}_{\mathrm{PV}}$ is a weighted average of $\hat{\sigma}_i^2$, as it can be written as

$$\hat{\mu}_{\mathrm{PV}} = \sum_{i=1}^{r} w_i\, \hat{\sigma}_i^2, \tag{9.25}$$

where

$$w_i = \frac{n_i - 1}{\sum_{i=1}^{r}(n_i - 1)}, \tag{9.26}$$

so that the weights are proportional to the degrees of freedom of the risk groups.

We now turn to the estimation of σ_{HM}^2 and present an unbiased estimator of σ_{HM}^2 in the following theorem.

Theorem 9.2 *The following quantity is an unbiased estimator of* σ_{HM}^2

$$\hat{\sigma}_{\mathrm{HM}}^2 = \frac{\left[\sum_{i=1}^{r} m_i(\bar{X}_i - \bar{X})^2\right] - (r-1)\hat{\mu}_{\mathrm{PV}}}{m - \dfrac{1}{m}\sum_{i=1}^{r} m_i^2}, \tag{9.27}$$

where $\hat{\mu}_{\mathrm{PV}}$ is defined in equation (9.16).

Proof We begin our proof by expanding the term $\sum_{i=1}^{r} m_i(\bar{X}_i - \bar{X})^2$ in the numerator of equation (9.27) as follows

$$\begin{aligned}
\sum_{i=1}^{r} m_i(\bar{X}_i - \bar{X})^2 &= \sum_{i=1}^{r} m_i \left[(\bar{X}_i - \mu_X) - (\bar{X} - \mu_X)\right]^2 \\
&= \sum_{i=1}^{r} m_i(\bar{X}_i - \mu_X)^2 + \sum_{i=1}^{r} m_i(\bar{X} - \mu_X)^2 \\
&\quad - 2\sum_{i=1}^{r} m_i(\bar{X}_i - \mu_X)(\bar{X} - \mu_X) \\
&= \left[\sum_{i=1}^{r} m_i(\bar{X}_i - \mu_X)^2\right] + m(\bar{X} - \mu_X)^2 \\
&\quad - 2(\bar{X} - \mu_X)\sum_{i=1}^{r} m_i(\bar{X}_i - \mu_X) \\
&= \left[\sum_{i=1}^{r} m_i(\bar{X}_i - \mu_X)^2\right] + m(\bar{X} - \mu_X)^2 \\
&\quad - 2m(\bar{X} - \mu_X)^2 \\
&= \left[\sum_{i=1}^{r} m_i(\bar{X}_i - \mu_X)^2\right] - m(\bar{X} - \mu_X)^2. \qquad (9.28)
\end{aligned}$$

We then take expectations on both sides of equation (9.28) to obtain

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{r} m_i(\bar{X}_i - \bar{X})^2\right] &= \left[\sum_{i=1}^{r} m_i \,\mathrm{E}\left[(\bar{X}_i - \mu_X)^2\right]\right] - m\,\mathrm{E}\left[(\bar{X} - \mu_X)^2\right] \\
&= \left[\sum_{i=1}^{r} m_i \mathrm{Var}(\bar{X}_i)\right] - m\mathrm{Var}(\bar{X}). \qquad (9.29)
\end{aligned}$$

Applying the result in equation (A.115) to $\mathrm{Var}(\bar{X}_i)$, we have

$$\mathrm{Var}(\bar{X}_i) = \mathrm{Var}\left[\mathrm{E}(\bar{X}_i \,|\, \Theta_i)\right] + \mathrm{E}\left[\mathrm{Var}(\bar{X}_i \,|\, \Theta_i)\right]. \qquad (9.30)$$

From equation (9.21) we conclude

$$\mathrm{Var}(\bar{X}_i \,|\, \Theta_i) = \frac{\sigma_X^2(\Theta_i)}{m_i}. \qquad (9.31)$$

Also, as $\mathrm{E}(\bar{X}_i \mid \Theta_i) = \mu_X(\Theta_i)$, equation (9.30) becomes

$$\mathrm{Var}(\bar{X}_i) = \mathrm{Var}\left[\mu_X(\Theta_i)\right] + \frac{\mathrm{E}\left[\sigma_X^2(\Theta_i)\right]}{m_i} = \sigma_{\mathrm{HM}}^2 + \frac{\mu_{\mathrm{PV}}}{m_i}. \tag{9.32}$$

Next, for $\mathrm{Var}(\bar{X})$ in equation (9.29), we have

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\frac{1}{m}\sum_{i=1}^{r} m_i \bar{X}_i\right) \\
&= \frac{1}{m^2}\sum_{i=1}^{r} m_i^2 \mathrm{Var}(\bar{X}_i) \\
&= \frac{1}{m^2}\sum_{i=1}^{r} m_i^2 \left(\sigma_{\mathrm{HM}}^2 + \frac{\mu_{\mathrm{PV}}}{m_i}\right) \\
&= \left[\sum_{i=1}^{r} \frac{m_i^2}{m^2}\right]\sigma_{\mathrm{HM}}^2 + \frac{\mu_{\mathrm{PV}}}{m}.
\end{aligned} \tag{9.33}$$

Substituting equations (9.32) and (9.33) into (9.29), we obtain

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{r} m_i(\bar{X}_i - \bar{X})^2\right] &= \left[\sum_{i=1}^{r} m_i\left(\sigma_{\mathrm{HM}}^2 + \frac{\mu_{\mathrm{PV}}}{m_i}\right)\right] \\
&\quad - \left[\left(\sum_{i=1}^{r} \frac{m_i^2}{m}\right)\sigma_{\mathrm{HM}}^2 + \mu_{\mathrm{PV}}\right] \\
&= \left[m - \frac{1}{m}\sum_{i=1}^{r} m_i^2\right]\sigma_{\mathrm{HM}}^2 + (r-1)\mu_{\mathrm{PV}}.
\end{aligned} \tag{9.34}$$

Thus, taking expectation of $\hat{\sigma}_{\mathrm{HM}}^2$, we can see that

$$\begin{aligned}
\mathrm{E}\left(\hat{\sigma}_{\mathrm{HM}}^2\right) &= \frac{\mathrm{E}\left[\sum_{i=1}^{r} m_i(\bar{X}_i - \bar{X})^2\right] - (r-1)\mathrm{E}(\hat{\mu}_{\mathrm{PV}})}{m - \frac{1}{m}\sum_{i=1}^{r} m_i^2} \\
&= \frac{\left[m - \frac{1}{m}\sum_{i=1}^{r} m_i^2\right]\sigma_{\mathrm{HM}}^2 + (r-1)\mu_{\mathrm{PV}} - (r-1)\mu_{\mathrm{PV}}}{m - \frac{1}{m}\sum_{i=1}^{r} m_i^2} \\
&= \sigma_{\mathrm{HM}}^2.
\end{aligned} \tag{9.35}$$

□

From equation (9.16), we can see that an unbiased estimate of μ_{PV} can be obtained with a single risk group, i.e. $r = 1$. However, as can be seen from equation (9.27), $\hat{\sigma}^2_{\text{HM}}$ cannot be computed unless $r > 1$. This is due to the fact that σ^2_{HM} measures the variations in the hypothetical means and requires at least two risk groups for a well-defined estimate.

As the Bühlmann model is a special case of the Bühlmann–Straub model, the results in Theorems 9.1 and 9.2 can be used to derive unbiased estimators of μ_{PV} and σ^2_{HM} for the Bühlmann model. This is summarized in the following corollary.

Corollary 9.1 *In the Bühlmann model with r risk groups, denote the loss observations by X_{ij}, for $i = 1, \cdots, r$, with $r > 1$, and $j = 1, \cdots, n_i$. Let the exposures of X_{ij} be the same, so that without loss of generality we set $m_{ij} \equiv 1$. The following quantity is then an unbiased estimator of μ_{PV}*

$$\tilde{\mu}_{\text{PV}} = \frac{\sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{\sum_{i=1}^{r} (n_i - 1)}, \tag{9.36}$$

and the following quantity is an unbiased estimator of σ^2_{HM}

$$\tilde{\sigma}^2_{\text{HM}} = \frac{\left[\sum_{i=1}^{r} n_i (\bar{X}_i - \bar{X})^2\right] - (r-1)\tilde{\mu}_{\text{PV}}}{n - \dfrac{1}{n} \sum_{i=1}^{r} n_i^2}, \tag{9.37}$$

where $n = \sum_{i=1}^{r} n_i$. In particular, if all risk groups have the same sample size, so that $n_i = n^*$ for $i = 1, \cdots, r$, then we have

$$\begin{aligned} \tilde{\mu}_{\text{PV}} &= \frac{1}{r(n^* - 1)} \left[\sum_{i=1}^{r} \sum_{j=1}^{n^*} (X_{ij} - \bar{X}_i)^2\right] \\ &= \frac{1}{r} \left[\sum_{i=1}^{r} s_i^2\right], \end{aligned} \tag{9.38}$$

where s_i^2 is the sample variance of the losses of the ith *group, and*

$$\begin{aligned} \tilde{\sigma}^2_{\text{HM}} &= \frac{1}{r-1} \left[\sum_{i=1}^{r} (\bar{X}_i - \bar{X})^2\right] - \frac{\tilde{\mu}_{PV}}{n^*} \\ &= S^2 - \frac{\tilde{\mu}_{PV}}{n^*}, \end{aligned} \tag{9.39}$$

where S^2 is the between-group sample variance.

Proof The proof is a straightforward application of Theorems 9.1 and 9.2, and is left as an exercise. □

With estimated values of the model parameters, the Bühlmann–Straub credibility predictor of the ith risk group can be calculated as

$$\hat{Z}_i \bar{X}_i + (1 - \hat{Z}_i)\bar{X}, \tag{9.40}$$

where

$$\hat{Z}_i = \frac{m_i}{m_i + \hat{k}}, \tag{9.41}$$

with

$$\hat{k} = \frac{\hat{\mu}_{\text{PV}}}{\hat{\sigma}^2_{\text{HM}}}. \tag{9.42}$$

For the Bühlmann credibility predictor, equations (9.41) and (9.42) are replaced by

$$\tilde{Z}_i = \frac{n_i}{n_i + \tilde{k}} \tag{9.43}$$

and

$$\tilde{k} = \frac{\tilde{\mu}_{\text{PV}}}{\tilde{\sigma}^2_{\text{HM}}}. \tag{9.44}$$

While $\hat{\mu}_{\text{PV}}$ and $\hat{\sigma}^2_{\text{HM}}$ are unbiased estimators of μ_{PV} and σ^2_{HM}, respectively, $\hat{k}$ is not unbiased for k, due to the fact that k is a nonlinear function of μ_{PV} and σ^2_{HM}. Likewise, $\tilde{k}$ is not unbiased for k.

Note that $\hat{\sigma}^2_{\text{HM}}$ and $\tilde{\sigma}^2_{\text{HM}}$ may be negative in empirical applications. In such circumstances, they may be set to zero, which implies that $\hat{k}$ and $\tilde{k}$ will be infinite, and that $\hat{Z}_i$ and $\tilde{Z}_i$ will be zero for all risk groups. Indeed, if the hypothetical means have no variation, the risk groups are homogeneous and there should be no differential weighting. In sum, the predicted loss is the overall average.

From equation (9.10), the total loss experienced is $m\bar{X} = \sum_{i=1}^{r} m_i \bar{X}_i$. Now if future losses are predicted according to equation (9.40), the total loss predicted will in general be different from the total loss experienced. If it is desired to equate the total loss predicted to the total loss experienced, some re-adjustment is needed. This may be done by using an alternative estimate of the average loss, denoted by $\hat{\mu}_X$, in place of $\bar{X}$ in equation (9.40). Now the total loss predicted becomes

$$\sum_{i=1}^{r} m_i[\hat{Z}_i \bar{X}_i + (1 - \hat{Z}_i)\hat{\mu}_X] = \sum_{i=1}^{r} m_i \left\{[1 - (1 - \hat{Z}_i)]\bar{X}_i + (1 - \hat{Z}_i)\hat{\mu}_X\right\}. \tag{9.45}$$

As $m_i(1-\hat{Z}_i) = \hat{Z}_i\hat{k}$, the above equation can be written as

$$\begin{aligned}
\text{Total loss predicted} &= \sum_{i=1}^{r} m_i \left\{[1-(1-\hat{Z}_i)]\bar{X}_i + (1-\hat{Z}_i)\hat{\mu}_X\right\} \\
&= \sum_{i=1}^{r} m_i\bar{X}_i + \sum_{i=1}^{r} m_i(1-\hat{Z}_i)(\hat{\mu}_X - \bar{X}_i) \\
&= \text{Total loss experienced} \\
&\quad + \hat{k}\sum_{i=1}^{r} \hat{Z}_i(\hat{\mu}_X - \bar{X}_i). \qquad (9.46)
\end{aligned}$$

Thus, to balance the total loss predicted and the total loss experienced, we must have $\hat{k}\sum_{i=1}^{r}\hat{Z}_i\left(\hat{\mu}_X - \bar{X}_i\right) = 0$, which implies

$$\hat{\mu}_X = \frac{\sum_{i=1}^{r}\hat{Z}_i\bar{X}_i}{\sum_{i=1}^{r}\hat{Z}_i}, \qquad (9.47)$$

and the loss predicted for the ith group is $\hat{Z}_i\bar{X}_i + (1-\hat{Z}_i)\hat{\mu}_X$.

Example 9.1 An analyst has data of the claim frequencies of workers' compensations of three insured companies. Table 9.1 gives the data of company A for the last three years and companies B and C for the last four years. The numbers of workers (in hundreds) and the numbers of claims each year per hundred workers are given.

Table 9.1. *Data for Example 9.1*

		Years			
Company		1	2	3	4
A	Claims per hundred workers	–	1.2	0.9	1.8
	Workers (in hundreds)	–	10	11	12
B	Claims per hundred workers	0.6	0.8	1.2	1.0
	Workers (in hundreds)	5	5	6	6
C	Claims per hundred workers	0.7	0.9	1.3	1.1
	Workers (in hundreds)	8	8	9	10

Calculate the Bühlmann–Straub credibility predictions of the numbers of claims per hundred workers for the three companies next year, without and with corrections for balancing the total loss with the predicted loss.

Solution The total exposures of each company are

$$m_A = 10 + 11 + 12 = 33,$$

$$m_B = 5 + 5 + 6 + 6 = 22,$$

and

$$m_C = 8 + 8 + 9 + 10 = 35,$$

which give the total exposures of all companies as $m = 33 + 22 + 35 = 90$. The exposure-weighted means of the claim frequency of the companies are

$$\bar{X}_A = \frac{(10)(1.2) + (11)(0.9) + (12)(1.8)}{33} = 1.3182,$$

$$\bar{X}_B = \frac{(5)(0.6) + (5)(0.8) + (6)(1.2) + (6)(1.0)}{22} = 0.9182,$$

and

$$\bar{X}_C = \frac{(8)(0.7) + (8)(0.9) + (9)(1.3) + (10)(1.1)}{35} = 1.0143.$$

The numerator of $\hat{\mu}_{\mathrm{PV}}$ in equation (9.16) is

$$\begin{aligned}
&(10)(1.2 - 1.3182)^2 + (11)(0.9 - 1.3182)^2 + (12)(1.8 - 1.3182)^2 \\
&+ (5)(0.6 - 0.9182)^2 + (5)(0.8 - 0.9182)^2 + (6)(1.2 - 0.9182)^2 \\
&+ (6)(1.0 - 0.9182)^2 + (8)(0.7 - 1.0143)^2 + (8)(0.9 - 1.0143)^2 \\
&+ (9)(1.3 - 1.0143)^2 + (10)(1.1 - 1.0143)^2 \\
&\quad = 7.6448.
\end{aligned}$$

Hence, we have

$$\hat{\mu}_{\mathrm{PV}} = \frac{7.6448}{2 + 3 + 3} = 0.9556.$$

The overall mean is

$$\bar{X} = \frac{(1.3182)(33) + (0.9182)(22) + (1.0143)(35)}{90} = 1.1022.$$

The first term in the numerator of $\hat{\sigma}^2_{\mathrm{HM}}$ in equation (9.27) is

$$(33)(1.3182 - 1.1022)^2 + (22)(0.9182 - 1.1022)^2$$

$$+ (35)(1.0143 - 1.1022)^2 = 2.5549,$$

and the denominator is

$$90 - \frac{1}{90}[(33)^2 + (22)^2 + (35)^2] = 58.9111,$$

so that

$$\hat{\sigma}^2_{\mathrm{HM}} = \frac{2.5549 - (2)(0.9556)}{58.9111} = \frac{0.6437}{58.9111} = 0.0109.$$

Thus, the Bühlmann–Straub credibility parameter estimate is

$$\hat{k} = \frac{\hat{\mu}_{\mathrm{PV}}}{\hat{\sigma}^2_{\mathrm{HM}}} = \frac{0.9556}{0.0109} = 87.6697,$$

and the Bühlmann–Straub credibility factor estimates of the companies are

$$\hat{Z}_A = \frac{33}{33 + 87.6697} = 0.2735,$$

$$\hat{Z}_B = \frac{22}{22 + 87.6697} = 0.2006,$$

and

$$\hat{Z}_C = \frac{35}{35 + 87.6697} = 0.2853.$$

We then compute the Bühlmann–Straub credibility predictors of the claim frequencies per hundred workers for company A as

$$(0.2735)(1.3182) + (1 - 0.2735)(1.1022) = 1.1613,$$

for company B as

$$(0.2006)(0.9182) + (1 - 0.2006)(1.1022) = 1.0653,$$

and for company C as

$$(0.2853)(1.0143) + (1 - 0.2853)(1.1022) = 1.0771.$$

Note that the total claim frequency predicted based on the historical exposure is

$$(33)(1.1613) + (22)(1.0653) + (35)(1.0771) = 99.4580,$$

which is not equal to the total recorded claim frequency of $(90)(1.1022) = 99.20$. To balance the two figures, we use equation (9.47) to obtain

$$\hat{\mu}_X = \frac{(0.2735)(1.3182) + (0.2006)(0.9182) + (0.2853)(1.0143)}{0.2735 + 0.2006 + 0.2853} = 1.0984.$$

Using this as the credibility complement, we obtain the updated predictors as

$$\begin{aligned}
\mathrm{A}: &\ (0.2735)(1.3182) + (1 - 0.2735)(1.0984) = 1.1585,\\
\mathrm{B}: &\ (0.2006)(0.9182) + (1 - 0.2006)(1.0984) = 1.0623,\\
\mathrm{C}: &\ (0.2853)(1.0143) + (1 - 0.2853)(1.0984) = 1.0744.
\end{aligned}$$

It can be checked that the total claim frequency predicted based on the historical exposure is

$$(33)(1.1585) + (22)(1.0623) + (35)(1.0744) = 99.20,$$

which balances with the total claim frequency recorded. □

Example 9.2 An insurer sold health policies to three companies. The claim experience of these companies in the last period is summarized in Table 9.2.

Table 9.2. *Data for Example 9.2*

Company	Number of employees	Mean claim amount per employee	Standard deviation of claim amount
A	350	467.20	116.48
B	673	328.45	137.80
C	979	390.23	86.50

Suppose company A has 380 employees in the new period, calculate the Bühlmann credibility predictor of its aggregate claim amount.

Solution Assuming the claim amounts of the employees within each company are independently and identically distributed, we employ the Bühlmann model. From equation (9.36), we have

$$\begin{aligned}\tilde{\mu}_{\mathrm{PV}} &= \frac{(349)(116.48)^2 + (672)(137.80)^2 + (978)(86.50)^2}{349 + 672 + 978} \\ &= \frac{24{,}813{,}230.04}{1{,}999} \\ &= 12{,}412.84.\end{aligned}$$

The overall mean of the claim amounts is

$$\begin{aligned}\bar{X} &= \frac{(350)(467.20) + (673)(328.45) + (979)(390.23)}{350 + 673 + 979} \\ &= \frac{766{,}602.02}{2{,}002} \\ &= 382.92.\end{aligned}$$

We compute the denominator of equation (9.37) to obtain

$$2{,}002 - \frac{1}{2{,}002}[(350)^2 + (673)^2 + (979)^2] = 1{,}235.83.$$

Thus, from equation (9.37), we have

$$\begin{aligned}
\tilde{\sigma}^2_{\mathrm{HM}} &= [(350)(467.20 - 382.92)^2 + (673)(328.45 - 382.92)^2 \\
&\quad + (979)(390.23 - 382.92)^2 - (2)(12{,}412.84)]/1{,}235.83 \\
&= 3{,}649.66,
\end{aligned}$$

and the Bühlmann credibility parameter estimate is

$$\tilde{k} = \frac{12{,}412.84}{3{,}649.66} = 3.4011.$$

For company A, its Bühlmann credibility factor is

$$\tilde{Z}_A = \frac{350}{350 + 3.4011} = 0.99,$$

so that the Bühlmann credibility predictor for the claim amount of the current period is

$$(380)\,[(0.99)(467.20) + (1 - 0.99)(382.92)] = 177{,}215.36. \qquad \square$$

Example 9.3 An insurer insures two rental car companies with similar sizes and operations. The aggregate-loss (in thousand dollars) experience in the last three years is summarized in Table 9.3. Assume the companies have stable business and operations in this period, calculate the predicted aggregate loss of company B next year.

Table 9.3. *Data for Example 9.3*

Company	Mean annual aggregate loss over 3 years	Standard deviation of annual aggregate loss
A	235.35	48.42
B	354.52	76.34

Solution In this problem, the numbers of observations of each risk group are $n^* = 3$. We calculate $\tilde{\mu}_{\mathrm{PV}}$ using equation (9.38) to obtain

$$\tilde{\mu}_{\mathrm{PV}} = \frac{(48.42)^2 + (76.34)^2}{2} = 4{,}086.1460.$$

As the overall mean is

$$\bar{X} = \frac{235.35 + 354.52}{2} = 294.94,$$

using equation (9.39) we obtain $\tilde{\sigma}^2_{\text{HM}}$ as

$$\begin{aligned}\tilde{\sigma}^2_{\text{HM}} &= (235.35 - 294.94)^2 + (354.52 - 294.94)^2 - \frac{4{,}086.1460}{3} \\ &= 5{,}738.6960.\end{aligned}$$

Thus, the Bühlmann credibility parameter estimate is

$$\tilde{k} = \frac{4{,}086.1460}{5{,}738.6960} = 0.7120,$$

so that the estimate of the Bühlmann credibility factor of company B is

$$\tilde{Z}_B = \frac{3}{3 + 0.7120} = 0.8082.$$

Therefore, the Bühlmann credibility prediction of the aggregate loss of company B next year is

$$(0.8082)(354.52) + (1 - 0.8082)(294.94) = 343.09. \qquad \square$$

9.3 Semiparametric estimation

The unbiasedness of $\hat{\mu}_{\text{PV}}$ and $\hat{\sigma}^2_{\text{HM}}$ holds under very mild conditions that the loss random variables X_{ij} are statistically independent of each other and are identically distributed within each risk group (under the same risk parameters). Other than this, no particular assumptions are necessary for the prior distribution of the risk parameters and the conditional distribution of the loss variables. In some applications, however, researchers may have information about the possible conditional distribution $f_{X_{ij} \mid \Theta_i}(x \mid \theta_i)$ of the loss variables. For example, claim frequency per exposure may be assumed to be Poisson distributed. In contrast, the prior distribution of the risk parameters, which are not observable, are usually best assumed to be unknown. Under such circumstances, estimates of the parameters of the Bühlmann–Straub model can be estimated using the semiparametric method.

Suppose X_{ij} are the claim frequencies per exposure and $X_{ij} \sim \mathcal{P}(\lambda_i)$, for $i = 1, \cdots, r$ and $j = 1, \cdots, n_i$. As $\sigma^2_X(\lambda_i) = \lambda_i$, we have

$$\mu_{\text{PV}} = \text{E}[\sigma^2_X(\Lambda_i)] = \text{E}(\Lambda_i) = \text{E}[\text{E}(X \mid \Lambda_i)] = \text{E}(X). \qquad (9.48)$$

Thus, μ_{PV} can be estimated using the overall sample mean of X, $\bar{X}$. From (9.27) an alternative estimate of σ^2_{HM} can then be obtained by substituting $\hat{\mu}_{\text{PV}}$ with $\bar{X}$.

Example 9.4 In Example 9.1, if the claim frequencies are assumed to be Poisson distributed, estimate the Bühlmann–Straub credibility parameter k using the semiparametric method.

Solution We estimate μ_{PV} using $\bar{X} = 1.1022$. Thus, the estimate of σ^2_{HM} is

$$\hat{\sigma}^2_{\mathrm{HM}} = \frac{2.5549 - (2)(1.1022)}{58.9111} = 0.005950,$$

so that the semiparametric estimate of the Bühlmann–Straub credibility parameter k is

$$\hat{k} = \frac{1.1022}{0.005950} = 185.24.$$ □

9.4 Parametric estimation

If the prior distribution of Θ and the conditional distribution of X_{ij} given Θ_i, for $i = 1, \cdots, r$ and $j = 1, \cdots, n_i$ are of known functional forms, then the hyperparameter of Θ, γ, can be estimated using the maximum likelihood estimation (MLE) method.[4] The quantities μ_{PV} and σ^2_{HM} are functions of γ, and we denote them by $\mu_{\mathrm{PV}} = \mu_{\mathrm{PV}}(\gamma)$ and $\sigma^2_{\mathrm{HM}} = \sigma^2_{\mathrm{HM}}(\gamma)$. As k is a function of μ_{PV} and σ^2_{HM}, the MLE of k can be obtained by replacing γ in $\mu_{\mathrm{PV}} = \mu_{\mathrm{PV}}(\gamma)$ and $\sigma^2_{\mathrm{HM}} = \sigma^2_{\mathrm{HM}}(\gamma)$ by the MLE of γ, $\hat{\gamma}$. Specifically, the MLE of k is

$$\hat{k} = \frac{\mu_{\mathrm{PV}}(\hat{\gamma})}{\sigma^2_{\mathrm{HM}}(\hat{\gamma})}. \tag{9.49}$$

We now consider the estimation of γ. For simplicity, we assume $m_{ij} \equiv 1$. The marginal pdf of X_{ij} is given by

$$f_{X_{ij}}(x_{ij} \,|\, \gamma) = \int_{\theta_i \in \Omega_\Theta} f_{X_{ij} \,|\, \Theta_i}(x_{ij} \,|\, \theta_i) f_{\Theta_i}(\theta_i \,|\, \gamma)\, d\theta_i. \tag{9.50}$$

Given the data X_{ij}, for $i = 1, \cdots, r$ and $j = 1, \cdots, n_i$, the likelihood function $L(\gamma)$ is

$$L(\gamma) = \prod_{i=1}^{r} \prod_{j=1}^{n_i} f_{X_{ij}}(x_{ij} \,|\, \gamma), \tag{9.51}$$

and the log-likelihood function is

$$\log[L(\gamma)] = \sum_{i=1}^{r} \sum_{j=1}^{n_i} \log f_{X_{ij}}(x_{ij} \,|\, \gamma). \tag{9.52}$$

[4] See Appendix A.19 for a review of the maximum likelihood estimation method. The result in equation (9.49) is justified by the invariance principle of the MLE (see Section 12.3).

The MLE of γ, $\hat{\gamma}$, is obtained by maximizing $L(\gamma)$ in equation (9.51) or $\log[L(\gamma)]$ in equation (9.52) with respect to γ.

Example 9.5 The claim frequencies X_{ij} are assumed to be Poisson distributed with parameter λ_i, i.e. $X_{ij} \sim \mathcal{PN}(\lambda_i)$. The prior distribution of Λ_i is gamma with hyperparameters α and β, where α is a known constant. Derive the MLE of β and k.

Solution As α is a known constant, the only hyperparameter of the prior is β. The marginal pf of X_{ij} is

$$\begin{aligned} f_{X_{ij}}(x_{ij} \mid \beta) &= \int_0^\infty \left[\frac{\lambda_i^{x_{ij}} \exp(-\lambda_i)}{x_{ij}!}\right]\left[\frac{\lambda_i^{\alpha-1} \exp\left(-\frac{\lambda_i}{\beta}\right)}{\Gamma(\alpha)\beta^\alpha}\right] d\lambda_i \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha x_{ij}!}\int_0^\infty \lambda_i^{x_{ij}+\alpha-1} \exp\left[-\lambda_i\left(\frac{1}{\beta}+1\right)\right] d\lambda_i \\ &= \frac{\Gamma(x_{ij}+\alpha)}{\Gamma(\alpha)\beta^\alpha x_{ij}!}\left(\frac{1}{\beta}+1\right)^{-(x_{ij}+\alpha)} \\ &= \frac{c_{ij}\beta^{x_{ij}}}{(1+\beta)^{x_{ij}+\alpha}}, \end{aligned}$$

where c_{ij} does not involve β. Thus, the likelihood function is

$$L(\beta) = \prod_{i=1}^{r}\prod_{j=1}^{n_i} \frac{c_{ij}\beta^{x_{ij}}}{(1+\beta)^{x_{ij}+\alpha}},$$

and ignoring the term that does not involve β, the log-likelihood function is

$$\log[L(\beta)] = (\log \beta)\left(\sum_{i=1}^{r}\sum_{j=1}^{n_i} x_{ij}\right) - [\log(1+\beta)]\left[n\alpha + \sum_{i=1}^{r}\sum_{j=1}^{n_i} x_{ij}\right],$$

where $n = \sum_{i=1}^{r} n_i$. The derivative of $\log[L(\beta)]$ with respect to β is

$$\frac{n\bar{x}}{\beta} - \frac{n(\alpha+\bar{x})}{1+\beta},$$

where

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{r}\sum_{j=1}^{n_i} x_{ij}\right).$$

The MLE of β, $\hat{\beta}$, is obtained by solving for β when the first derivative of $\log[L(\beta)]$ is set to zero. Hence, we obtain[5]

$$\hat{\beta} = \frac{\bar{x}}{\alpha}.$$

As $X_{ij} \sim \mathcal{PN}(\lambda_i)$ and $\Lambda_i \sim \mathcal{G}(\alpha, \beta)$, $\mu_{\text{PV}} = \text{E}[\sigma_X^2(\Lambda_i)] = \text{E}(\Lambda_i) = \alpha\beta$. Also, $\sigma_{\text{HM}}^2 = \text{Var}[\mu_X(\Lambda_i)] = \text{Var}(\Lambda_i) = \alpha\beta^2$, so that

$$k = \frac{\alpha\beta}{\alpha\beta^2} = \frac{1}{\beta}.$$

Thus, the MLE of k is

$$\hat{k} = \frac{1}{\hat{\beta}} = \frac{\alpha}{\bar{x}}. \qquad \square$$

9.5 Summary and discussions

While the hyperparameters of the prior distribution are assumed to be known values in the Bayesian model, they are typically unknown in practical applications. The empirical Bayes method adopts the Bayesian approach of analysis, but treats the hyperparameters as quantities to be obtained from the data. In this chapter, we discuss some empirical Bayes approaches for the estimation of the quantities necessary for the implementation of the credibility prediction.

The nonparametric approach makes no assumption about the prior pdf (or pf) of the risk parameters and the conditional pdf (or pf) of the loss variables. For the Bühlmann–Straub and Bühlmann models, these estimates are easy to calculate. The semiparametric approach assumes knowledge of the conditional pdf (or pf) of the loss variable but not the prior distribution of the risk parameters. We illustrate an application of this approach when the loss variable is distributed as Poisson. When assumptions are made for both the prior and conditional distributions, the likelihood function of the hyperparameters can be derived, at least in principle. We can then estimate the hyperparameters using the MLE method, which may require numerical methods.

[5] It can be verified that the second derivative of $\log[L(\beta)]$ evaluated at $\hat{\beta}$ is negative, so that $\log[L(\beta)]$ is maximized. Note that α is a known constant in the computation of $\hat{\beta}$.

Exercises

9.1 The claim experience of three policyholders in three years is given as follows:

Policyholder		Year 1	Year 2	Year 3
1	Total claims	–	2,200	2,700
	Number in group	–	100	110
2	Total claims	2,100	2,000	1,900
	Number in group	90	80	85
3	Total claims	2,400	2,800	3,000
	Number in group	120	130	140

Determine the Bühlmann–Straub credibility premium for each group in Year 4.

9.2 An actuary is making credibility estimates for rating factors using the Bühlmann–Straub nonparametric empirical Bayes method. Let X_{it} denote the rating for Group i in Year t, for $i = 1$, 2, and 3, and $t = 1, \cdots, T$, and m_{it} denote the exposure. Define $m_i = \sum_{t=1}^{T} m_{it}$ and $\bar{X}_i = (\sum_{t=1}^{T} m_{it} X_{it})/m_i$. The following data are available:

Group	m_i	$\bar{X}_i$
1	50	15.02
2	300	13.26
3	150	11.63

The actuary computed the empirical Bayes estimate of the Bühlmann–Straub credibility factor of Group 1 to be 0.6791.

(a) What are the Bühlmann–Straub credibility estimates of the rating factors for the three groups using the overall mean of X_{it} as the manual rating?

(b) If it is desired to set the aggregate estimated credibility rating equal to the aggregate experienced rating, estimate the rating factors of the three groups.

Questions adapted from SOA exams

9.3 You are given the following data:

	Year 1	Year 2
Total losses	12,000	14,000
Number of policyholders	25	30

If the estimate of the variance of the hypothetical means is 254, determine the credibility factor for Year 3 using the nonparametric empirical Bayes method.

9.4 The claim experience of three territories in the region in a year is as follows:

Territory	Number of insureds	Number of claims
A	10	4
B	20	5
C	30	3

The numbers of claims for each insured each year are independently Poisson distributed. Each insured in a territory has the same number of expected claim frequencies, and the number of insured is constant over time for each territory. Determine the empirical Bayes estimate of the Bühlmann–Straub credibility factor for Territory A.

9.5 You are given the following data:

Group		Year 1	Year 2	Year 3	Total
A	Total claims		10,000	15,000	25,000
	Number in group		50	60	110
	Average		200	250	227.27
B	Total claims	16,000	18,000		34,000
	Number in group	100	90		190
	Average	160	200		178.95
	Total claims				59,000
	Number in group				300
	Average				196.67

If the estimate of the variance of the hypothetical means is 651.03, determine the nonparametric empirical Bayes estimate of the Bühlmann–Straub credibility factor of Group A.

9.6 During a two-year period, 100 policies had the following claims experience:

Total claims in Years 1 and 2	Number of policies
0	50
1	30
2	15
3	4
4	1

You are given that the number of claims per year follows a Poisson distribution, and that each policyholder was insured for the entire two-year period. A randomly selected policyholder had one claim over the two-year period. Using the semiparametric empirical Bayes method, estimate the number of claims in Year 3 for the selected policyholder.

9.7 The number of claims of each policyholder in a block of auto-insurance policies is Poisson distributed. In Year 1, the following data are observed for 8,000 policyholders:

Number of claims	Number of policyholders
0	5,000
1	2,100
2	750
3	100
4	50
5 or more	0

A randomly selected policyholder had one claim in Year 1. Determine the semiparametric Bayes estimate of the number of claims in Year 2 of this policyholder.

9.8 Three policyholders have the following claims experience over three months:

Policyholder	Month 1	Month 2	Month 3	Mean	Variance
A	4	6	5	5	1
B	8	11	8	9	3
C	5	7	6	6	1

Calculate the nonparametric empirical Bayes estimate of the credibility factor for Month 4.

9.9 Over a three-year period, the following claims experience was observed for two insureds who own delivery vans:

		Year		
Insured		1	2	3
A	Number of vehicles	2	2	1
	Number of claims	1	1	0
B	Number of vehicles	–	3	2
	Number of claims	–	2	3

The number of claims of each insured each year follows a Poisson distribution. Determine the semiparametric empirical Bayes estimate of the claim frequency per vehicle for Insured A in Year 4.

9.10 Three individual policyholders have the following claim amounts over four years:

Policyholder	Year 1	Year 2	Year 3	Year 4
A	2	3	3	4
B	5	5	4	6
C	5	5	3	3

Using the nonparametric empirical Bayes method, calculate the estimated variance of the hypothetical means.

9.11 Two policyholders A and B had the following claim experience in the last four years:

Policyholder	Year 1	Year 2	Year 3	Year 4
A	730	800	650	700
B	655	650	625	750

Determine the credibility premium for Policyholder B using the nonparametric empirical Bayes method.

9.12 The number of claims of each driver is Poisson distributed. The experience of 100 drivers in a year is as follows:

Number of claims	Number of drivers
0	54
1	33
2	10
3	2
4	1

Determine the credibility factor of a single driver using the semiparametric empirical Bayes method.

9.13 During a five-year period, 100 policies had the following claim experience:

Number of claims in Years 1 through 5	Number of policies
0	46
1	34
2	13
3	5
4	2

The number of claims of each policyholder each year follows a Poisson distribution, and each policyholder was insured for the entire period. A randomly selected policyholder had three claims over the five-year period. Using the semiparametric empirical Bayes method, determine the estimate for the number of claims in Year 6 for the same policyholder.

9.14 Denoting X_{ij} as the loss of the ith policyholder in Year j, the following data of four policyholders in seven years are known

$$\sum_{i=1}^{4}\sum_{j=1}^{7}(X_{ij}-\bar{X}_i)^2 = 33.60, \qquad \sum_{i=1}^{4}(\bar{X}_i-\bar{X})^2 = 3.30.$$

Using nonparametric empirical Bayes method, calculate the credibility factor for an individual policyholder.