Nonlife Actuarial Models

Chapter 8

Bayesian 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

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

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 $f_\Theta(\theta \,|\, \gamma)$ (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 **Bayesian 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.

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

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

- 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 8.2: Let X be a Bernoulli random variable taking value 1 with probability of θ and 0 with probability $1-\theta$. There is a random sample of n observations of X denoted by $\boldsymbol{X} = \{X_1, \cdots, X_n\}$. If Θ follows the beta distribution with parameters α and β, i.e., $\Theta \sim \mathcal{B}(\alpha, \beta)$, 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

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

we conclude that

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

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

$$
\begin{aligned}
f_{\Theta \mid \boldsymbol{X}}(\theta \mid \boldsymbol{x}) &= K(\boldsymbol{x}) f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta) f_{\Theta}(\theta) \\
&\propto f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta) f_{\Theta}(\theta). \qquad (8.6)
\end{aligned}
$$

$K(\boldsymbol{x})$ is free of θ and is a **constant of proportionality**. It scales the posterior pdf so that it integrates to 1.

- The expression $f_{\boldsymbol{X} \mid \Theta}(\boldsymbol{x} \mid \theta) f_{\Theta}(\theta)$ 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

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

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 $\boldsymbol{X}$. $\square$

8.1.2 Loss Function and Bayesian Estimation

- We 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.

- 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

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

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

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

- It is 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 **Bayesian estimator** of $\mu_X(\Theta)$ with respect to the chosen loss function.

- The Bayesian estimator, denoted by $w^*(\boldsymbol{x})$, satisfies

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

for any given $\boldsymbol{x}$.

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

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

- For the squared-error loss function, the Bayesian estimator of $\mu_X(\Theta)$ is the posterior mean, denoted by $\hat{\mu}_X(\boldsymbol{x})$, so that

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

- 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 $\boldsymbol{x}$.

- We first calculate the conditional pdf of X_{n+1} given $\boldsymbol{x}$, which is

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

- 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

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

which 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}.

- 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

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

- Using equation (8.15), we have

$$\begin{aligned}
\mathrm{E}(X_{n+1} \,|\, \boldsymbol{x}) &= \int_0^\infty x_{n+1} \left[\int_{\theta \,\in\, \Omega_\Theta} f_{X_{n+1} \,|\, \boldsymbol{\Theta}}(x_{n+1} \,|\, \theta) f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x}) \, d\theta \right] dx_{n+1} \\
&= \int_{\theta \,\in\, \Omega_\Theta} \left[\int_0^\infty x_{n+1} f_{X_{n+1} \,|\, \boldsymbol{\Theta}}(x_{n+1} \,|\, \theta) \, dx_{n+1} \right] f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x}) \, d\theta \\
&= \int_{\theta \,\in\, \Omega_\Theta} \mathrm{E}(X_{n+1} \,|\, \theta) \, f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x}) \, d\theta \\
&= \int_{\theta \,\in\, \Omega_\Theta} \mu_X(\theta) \, f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x}) \, d\theta \\
&= \mathrm{E}[\mu_X(\Theta) \,|\, \boldsymbol{x}]. \qquad (8.18)
\end{aligned}$$

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

- In summary, the Bayesian 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).

- We shall use the terminologies Bayesian estimate of expected loss and Bayesian predictor of future loss interchangeably.

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

		$\Pr(X = x \mid \theta)$		
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 3 claims with $\boldsymbol{x} = (20, 20, 30)$ is observed. Calculate the posterior mean of X. Compute the conditional pf 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

$$f_{\Theta\boldsymbol{X}}\,(1, \boldsymbol{x}) = f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,1)f_{\Theta}(1) = (0.045)(0.4) = 0.018,$$

$$f_{\Theta\boldsymbol{X}}\,(2, \boldsymbol{x}) = f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,2)f_{\Theta}(2) = (0.032)(0.4) = 0.0128,$$

and

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

Thus, we obtain

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

so that the posterior distribution of Θ is

$$f_{\Theta \,|\, \boldsymbol{X}}(1 \,|\, \boldsymbol{x}) = \frac{f_{\Theta \boldsymbol{X}}(1, \boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})} = \frac{0.018}{0.0308} = 0.5844,$$

$$f_{\Theta \,|\, \boldsymbol{X}}(2 \,|\, \boldsymbol{x}) = \frac{f_{\Theta \boldsymbol{X}}(2, \boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})} = \frac{0.0128}{0.0308} = 0.4156,$$

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

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

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

and

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

Thus, the posterior mean of X is

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

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

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

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

$$f_{X_4 \,|\, \boldsymbol{X}}(10 \,|\, \boldsymbol{x}) = (0.2)(0.5844) + (0.4)(0.4156) = 0.2831,$$

$$f_{X_4 \,|\, \boldsymbol{X}}(20 \,|\, \boldsymbol{x}) = (0.3)(0.5844) + (0.4)(0.4156) = 0.3416,$$

and

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

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. $\square$

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.

- 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.

- 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 $f_{\boldsymbol{X} \,|\, \Theta}(\boldsymbol{x} \,|\, \theta)$ 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.

8.2.1 The gamma-Poisson conjugate distribution

- Let $\{X_i\}$ be iid Poisson random variables with parameter λ. The random variable Λ of the parameter λ is assumed to follow a gamma distribution with hyperparameters α and β, i.e., the prior pdf of Λ is

$$f_\Lambda(\lambda; \alpha, \beta) = \frac{\lambda^{\alpha-1} e^{-\frac{\lambda}{\beta}}}{\Gamma(\alpha)\beta^\alpha},$$

and the likelihood of $\boldsymbol{X} = \{X_1, X_2, \cdots, X_n\}$ is

$$\begin{aligned} f_{\boldsymbol{X}\,|\,\Lambda}(\boldsymbol{x}\,|\,\lambda) &= \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \\ &= \frac{\lambda^{n\bar{x}} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}. \end{aligned}$$

- Thus, the posterior pdf of Λ satisfies

$$\begin{aligned} f_{\Lambda\,|\,\boldsymbol{X}}(\lambda\,|\,\boldsymbol{x}) &\propto f_{\boldsymbol{X}\,|\,\Lambda}(\boldsymbol{x}\,|\,\lambda) f_{\Lambda}(\lambda; \alpha, \beta) \\ &\propto \lambda^{\alpha + n\bar{x} - 1} e^{-\lambda\left(n + \frac{1}{\beta}\right)}. \end{aligned}$$

- We conclude that the posterior pdf of Λ is $f_{\Lambda}(\lambda; \alpha^*, \beta^*)$, where

$$\alpha^* = \alpha + n\bar{x} \tag{8.19}$$

and

$$\beta^* = \left[n + \frac{1}{\beta}\right]^{-1} = \frac{\beta}{n\beta + 1}. \tag{8.20}$$

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

8.2.2 The beta-geometric conjugate distribution

- Let $\{X_i\}$ be iid geometric random variables with parameter θ so that the likelihood of $\boldsymbol{X}$ is

$$f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta) = \prod_{i=1}^{n} \theta(1-\theta)^{x_i} = \theta^n(1-\theta)^{n\bar{x}}.$$

If the prior distribution of Θ is beta with hyperparameters α and β, then the posterior pdf of Θ satisfies

$$\begin{aligned} f_{\Theta\,|\,\boldsymbol{X}}(\theta\,|\,\boldsymbol{x}) &\propto f_{\boldsymbol{X}\,|\,\Theta}(\boldsymbol{x}\,|\,\theta) f_\Theta(\theta;\alpha,\beta) \\ &\propto \theta^{\alpha+n-1}(1-\theta)^{\beta+n\bar{x}-1}. \end{aligned}$$

- We conclude that the posterior distribution of Θ is beta with parameters

$$\alpha^* = \alpha + n \tag{8.21}$$

and

$$\beta^* = \beta + n\bar{x}, \tag{8.22}$$

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

8.2.3 The gamma-exponential conjugate distribution

- Let $\{X_i\}$ be iid exponential random variables with parameter λ so that the likelihood of $\boldsymbol{X}$ is

$$f_{\boldsymbol{X} \mid \Lambda}(\boldsymbol{x} \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda n \bar{x}}.$$

- If the prior distribution of Λ is gamma with hyperparameters α and β, then the posterior pdf of Λ satisfies

$$\begin{aligned} f_{\Lambda \mid \boldsymbol{X}}(\lambda \mid \boldsymbol{x}) &\propto f_{\boldsymbol{X} \mid \Lambda}(\boldsymbol{x} \mid \lambda) f_{\Lambda}(\lambda; \alpha, \beta) \\ &\propto \lambda^{\alpha+n-1} e^{-\lambda\left(\frac{1}{\beta}+n\bar{x}\right)}. \end{aligned}$$

- We conclude that the posterior distribution of Λ is gamma with parameters

$$\alpha^* = \alpha + n \tag{8.23}$$

and

$$\beta^* = \left[\frac{1}{\beta} + n\bar{x}\right]^{-1} = \frac{\beta}{1 + \beta n\bar{x}}. \tag{8.24}$$

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

Table A.3: Some conjugate distributions

Prior pdf and hyperparameters	Likelihood of $\boldsymbol{X}$	Hyperparameters of posterior pdf
$\mathcal{B}(\alpha, \beta)$	Bernoulli	$\alpha + n\bar{x},\ \beta + n - n\bar{x}$
$\mathcal{B}(\alpha, \beta)$	$\mathcal{BN}(m_i, \theta)$	$\alpha + n\bar{x},\ \beta + \sum_{i=1}^{n}(m_i - x_i)$
$\mathcal{G}(\alpha, \beta)$	$\mathcal{PN}(\lambda)$	$\alpha + n\bar{x},\ \dfrac{\beta}{n\beta + 1}$
$\mathcal{B}(\alpha, \beta)$	$\mathcal{GM}(\theta)$	$\alpha + n,\ \beta + n\bar{x}$
$\mathcal{G}(\alpha, \beta)$	$\mathcal{E}(\lambda)$	$\alpha + n,\ \dfrac{\beta}{1 + \beta n\bar{x}}$

8.3 Bayesian versus Bühlmann Credibility

- If the prior distribution is conjugate to the likelihood, the Bayesian estimate is easy to obtain.
- For the conjugate distributions discussed, the Bühlmann credibility estimate is equal to the Bayesian 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 Bayesian estimate.

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

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

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

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

Thus,

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

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

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

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 Bayesian 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 Bayesian estimate of the expected claim frequency is

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

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 Bayesian estimate.

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

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

and

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

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

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

Hence, the expected value of the process variance is

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

and the variance of the hypothetical means is

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

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

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

and the Bühlmann credibility factor is

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

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 Bayesian 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

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

which is the same as the Bühlmann credibility estimate. $\square$

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 Bayesian estimate.

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

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

and

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

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

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

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

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

so that

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

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 Bayesian 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 Bayesian estimate of the expected claim severity is

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

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

- If the conjugate distributions discussed 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 Bayesian estimate.

- In such cases, the Bühlmann credibility estimate is said to have **exact credibility**.

- There is 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 $C(x)$.

- The binomial, geometric, Poisson and exponential distributions all belong to the linear exponential family.

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

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 Bayesian estimate.

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

- When the conditions of Theorem 8.1 hold, the Bühlmann credibility estimate is the same as the Bayesian 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 Bayesian estimator.

- The Bühlmann credibility estimator, however, still has the minimum mean squared error in the class of linear estimators based on the sample.

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 Bayesian estimate of the expected claim frequency. Compare the mean squared errors of these estimate as well as that of the sample mean.

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

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

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

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

and

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

Thus, we have

$$k = \frac{\mu_{\text{PV}}}{\sigma^2_{\text{HM}}} = \frac{1.5}{0.25} = 6,$$

and the Bühlmann credibility factor is

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

As the prior mean of X is

$$M = \text{E}[\text{E}(X \,|\, \Lambda)] = \text{E}(\Lambda) = 1.5,$$

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

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

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

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

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}) &= \sum_{\lambda \in \{1,2\}} f_{\boldsymbol{X} \mid \Lambda}(\boldsymbol{x} \mid \lambda) \Pr(\Lambda = \lambda) \\ &= 0.5 \left[\sum_{\lambda \in \{1,2\}} \left(\prod_{i=1}^{6} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \right) \right] \\ &= 0.5 \left[\left(\frac{1}{e^6} \prod_{i=1}^{6} \frac{1}{x_i!} \right) + \left(\frac{1}{e^{12}} \prod_{i=1}^{6} \frac{2^{x_i}}{x_i!} \right) \right] \\ &= K, \qquad \text{say.} \end{aligned}$$

Thus, the posterior pdf of Λ is

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

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 Bayesian 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

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 6 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 Bayesian estimates computed in Step 3 by $\hat{\lambda}_1, \cdots, \hat{\lambda}_m$. The estimated mean squared error of the Bayesian 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 Bayesian 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).