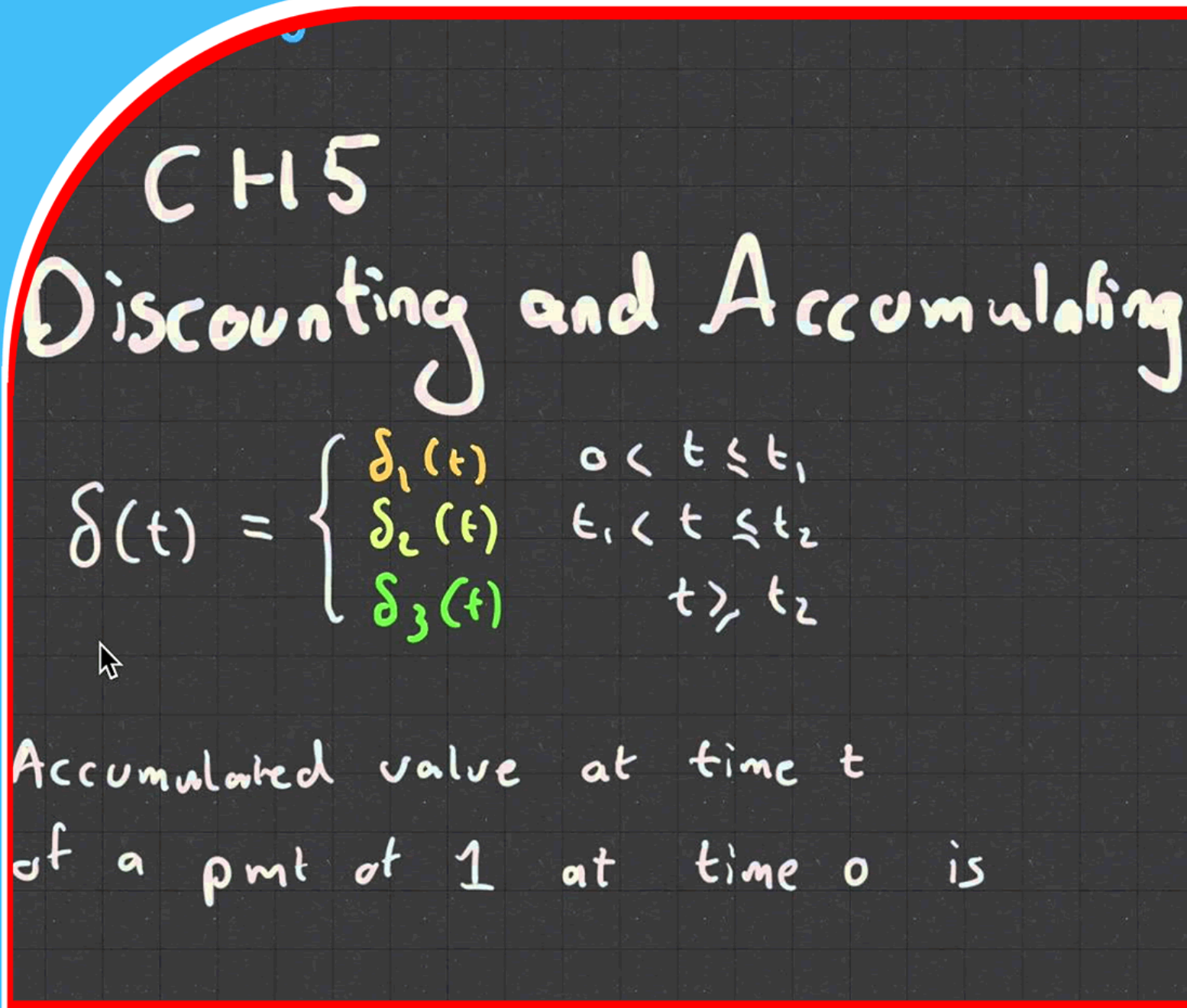
American Journal of
Statistics and Actuarial Science
(AJSAS)



The Exponential Premium Principle and Its Link to Credibility Theory

Maram AlOmari

American Journal of Statistics and Actuarial Science
ISSN 2958-5244(Online)
Vol.6, Issue 1, pp 1–34, 2025

www.ajpojournals.org

The Exponential Premium Principle and Its Link to Credibility Theory

Maram AlOmari
Department of Mathematics King Saud University, Riyadh, Saudi Arabia

Article history
Submitted 07.02.2025 Revised Version Received 06.03.2025 Accepted 03.04.2025

Abstract

Purpose: The main object of discussing this study is to introduce and derive a new alternative of the equivalence premium principle called the exponential premium principle in actuarial science. The research elaborates on how past-experience data can be incorporated into the calculation of premiums based on the exponential premium so that the premiums will be more precise and fair. Moreover, through the application of Bayesian tools and of risk theory, the connection between the exponential premium principle and the theory of credibility is established, leading in turn to the derivation of the exponential credibility premium.

Materials and Methods: The study begins with the establishment of the mathematical principles for the derivation of different methods of loss functions. The exponential loss function is one way of presenting the formulation of exponential premium principles, which makes it different from the classical equivalence and expectation principles. Bayesian inference provides individual experience with input into each of the pricing levels. Concepts of credibility theory have also been used to demonstrate the relation between the exponential premium and the credibility premium.

Findings: The results demonstrate that the exponential premium provides an even more flexible and theoretically justified framework for premium calculations compared to traditional ways. The derivation of the exponential credibility premium shows that it gives an appropriate balance between collective risk assessment through the collective risk premium and individual experience through the credibility premium, allowing for more personalized and fair insurance pricing. The study also makes note of the exponential principle's ability to encompass risk variations and shield insurers against financial instability.

Unique Contribution to Theory, Practice and Policy: Theoretical contribution: The study will enhance the body of knowledge on premium calculation principles by integrating Bayesian learning with credibility theory. Practical application: The exponential credibility premium could improve risk assessment and hence enhance equity in insurance pricing. Policy Implications*:* The findings may be utilized by regulators and insurers to improve the process of premium-setting so that at least they reflect a balanced consideration of collective and individual risk in a more definitive way.

Keywords: *Exponential premium, Credibility theory, Bayesian inference, Insurance pricing, Risk assessment.*

JEL Classification: *G22, C13, C58, D81, G32*

INTRODUCTION

When an insurance policy is issued, the insurer agrees to bear at least some risk in exchange for a premium payment. An insurance company sells multiple insurance policies at once or throughout the year1. Actuaries group risks that share some similarity to facilitate processing such a large amount of information. To expand upon the idea of similarity, it is a spectrum, at the higher end of which all the risks are identical and the actuary's job is easy. A price that is calculated to be fair for one risk will be fair for every other risk within the same group/class. As an illustration, imagine an actuary sitting at a desk, staring at a spreadsheet of 10,000 car insurance claims categorized by car brand (e.g., BMW, Toyota, or Jaguar). The actuary finds that all BMW drivers make claims of 432 pounds, while all Toyota drivers make claims of 156 pounds. These classes are considered homogeneous. What is the likelihood that this will occur by chance and is it realistic? Does the car brand really tell us that much information? What if we considered additional subcategories, such as the driver's age? Would all 30-year- old Toyota drivers make claims for the same amount? If an actuary continues to subdivide data until each class has one or two data points, the data thereby becomes small and statistically redundant. Thus, seeking homogeneity does not make sense.

To go back to the similarity spectrum, even if homogeneity is not achievable, it does not mean it is impossible that claim amounts within a class would be 'close' to one another. After all, repairs on cheaper cars are not as costly as those for more expensive ones. However, the best way to address this is not by setting premiums for a safe driver based solely on the high-risk group to which he or she belongs. This approach might discourage safe driving habits among policyholders. Thus, heterogeneity within risk classes is a good reason to consider individual experiences in pricing policies.

Brief History: How to Incorporate Individual Experience in Pricing?

In light of Goulet's work [5] on credibility theory, there are two recognized approaches. The earlier of the two, introduced by Mowbray, dates back to 1914 and is known as the limited fluctuation credibility method. It answers the question: How many observations are sufficient for past individual experience data to be fully credible? It is an All-or-nothing approach. For example, without mathematical details, once a benchmark has been obtained as an outcome of 400 observations, if two individuals are taken from the same class and one has 390 past experience observations and the other has 405, then the first policyholder's past experience has 0% credibility despite being relatively close to the benchmark, while the latter has 100% credibility.

The void between zero and full credibility was later filled with the category of partial credibility by Albert W. Whitney in 1918. Whitney decided that both information from the overall class and from the individual's experience are important and gave a method for determining a price that is balanced between the two sources. This work was built on, and a newer application was formulated in two papers by Bailey, one in 1945 and the other in 1950. The more modern approach is known as greatest accuracy credibility, and gives credibility ranges from 0% to 100% depending on how significantly the individual's experience differs from the rest of the class and on its quality. Partial credibility is more realistic and flexible approach especially when data is limited. More on the history of credibility theory can be found in Goulet [5].

Paper Overview

Risks are inherently random. Loss functions are used to map random loss to particular quantifiable terms. Furthermore, it is possible to formulate premium calculation principles to convert losses into monetary terms. In risk theory, various loss functions were introduced. One

can name, for example, the famous quadratic loss function, which is used to derive the equivalence premium principle. The focus in this paper will be on the exponential loss function and on the derivation of the exponential premium principle from it, and how it can be equipped with past experience data. The presence of the equivalence principle premium in this work serves the purposes of illustration and clarification, as it is assumed that the reader is familiar with this premium from classical risk theory. In Chapter 2, two objectives are pursued: The first one is to recall Heilmann's approach in [7] of deriving different premium calculation principles from different loss functions. The second one is to examine the properties of the exponential principle from risk management perspectives (We refer in this regard to Dhaene et al. [8] and to Gerber and Jones [6]). Chapter 3 begins with a review of the concept of Bayesian learning through Bayes' theorem. The discussion then shifts to individual experience-based rate-making, and relevant estimations using Bayesian inference on two levels: non-linear Bayesian inference, and linear Bayesian inference. Credibility theory in the actuarial world is an application of the latter. The chapter goes on to elaborate on the derivation of credibility premiums under equivalence and exponential principles. A concise explanation of credibility theory, particularly under the equivalence principle, is provided in Bühlmann & Gisler [3]. Regarding the credibility premium under the exponential principle, a thorough investigation is laid out in Wen et al. [12], and Wen et al. [13].

Statement of the Problem

Exponential premium is frequently mentioned in actuarial literature. However, little they explain how it links to credibility theory. This work aims to clarify that link. Future researchers may benefit from mathematical reasoning and derivations to improve the research on the application of exponential premiums.

Loss Functions and Premium Principles

It is the decision-maker's task to choose an appropriate loss function towards risks arising from a particular insured event (e.g. fire, car crash) and its consequences (e.g. medical treatment, paying rent for a temporary car). In the simplest words, this concerns the modeling of such events taking place, while the details of the choice process are outside the scope of this investigation. However, a number of considerations are briefly mentioned. The objectives of this chapter are twofold: to pursue Heilmann's approach in [7] for generating theoretic premium calculation principles after the proper loss function is selected, and to emphasize the advantages of the exponential premium principle in the risk management and ruin theory contexts, as it is the principle of interest in this article.

All calculations throughout this chapter are for an arbitrary loss random variable, meaning a loss that is not classified. Later, a foundation will be established, from which we will move on to calculating premiums for parameterized or classified losses in the next chapter.

Definition 2.1. A loss function is a function L: $L^0 \times G \rightarrow R$ that assigns a random variable to a loss random variable in L^0 and an action in G, where L^0 is the space of all real random variables and G is the set of possible premiums. The associated expected loss function is defined by l(X, P). It is assumed hereafter that the loss random variable is a continuous bounded and non-negative random variable that its first and second moments exist in addition to its moment-generating function.

Definition 2.2. A premium calculation principle is a function H: L0 → R that assigns a real number to a loss random variable.

Losses have different distributions and some of these are symmetric. The quadratic loss function is appropriate in that case and it will be used here to illustrate Heilmann's approach

for the derivation of a number of premium calculation principles. Let

$$l_{se}(X,P) = E(X-P)^2 \qquad (2.1)$$

Where E stands for the expectation under the probability measure. An insurer does not have control over the risk X occurring, however, he or she has control over the premium parameter$P$.

According to Heilmann [7], the insurer has to calculate a price that, intuitively, reduces the expected losses to the lowest level possible. In other terms, the appropriate premium $\mathcal{H}(X)$ in$\mathbf{G}$, for a risk X is the one that minimizes the expected loss function$l_{se}(X,P)$. This is expressed mathematically as the following minimization problem:

$$\min_{P\in\mathbf{G}_{se}} [l_{se}(X,P)] = l_{se}(X,\mathcal{H}(X)) \qquad (2.2)$$

The solution $\mathcal{H}(X)$ to the minimization problem (2.2) can be determined by differentiation. Start by simplifying the expression using the linearity property of the expectation operator (cf. A.11) as follows:

$$\begin{aligned} l_{se}(X,P) &= \mathbb{E}[(X-P)^2] \\ &= \mathbb{E}[X^2 - 2XP + P^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]P + P^2, \end{aligned}$$

We suppose no further constraint on the premium and therefore$\mathbf{G} = \mathbb{R}$, the real line. Take partial differentiation with respect to P and equate to zero

$$\begin{aligned} \frac{\partial}{\partial P}\mathbb{E}[X^2] - 2\mathbb{E}[X]\frac{\partial}{\partial P}P + \frac{\partial}{\partial P}P^2 &= 0 \\ -2\mathbb{E}[X] + 2P &= 0 \end{aligned}$$

Finally, solve for P and get the premium principle

$$\mathcal{H}_e(X) = \mathbb{E}[X] \qquad (2.3)$$

This is called the equivalence principle and it uses only the first moment of the loss distributionF_X. This principle is used in case of risk indifference or neutrality. It suggests that charging a higher premium is as bad as charging a lower premium. This supports the fairness of this principle, but because selling products at the break-even price is not profitable, Wen et al. suggest in [12] that ruin is inevitable if the insurer relies exclusively on this principle. A possible way to avoid that is to calculate a risk loading separately and include it in the calculation of the premium. This can be expressed as the following:

$$\mathcal{H}_{ex}(X) = (1+\beta)\mathbb{E}(X), \beta > 0 \qquad (2.4)$$

This form is called the expectation principle and it can view as the equivalence principle with a risk loading. Bühlmann and Gisler (p.9) [3] described a risk loading as having an economic advantage: it protects against market volatility, and thus it can be interpreted as the cost of risking the firm's capital. Therefore, not having a built-in risk loading presents a drawback for a premium principle. The expectation principle has the advantage of risk loading. However, a downside of the expectation principle is that when two groups have the same distribution mean and different degrees of fluctuation, an identical premium will be charged to both under this principle; this applies to the equivalence principle as well. The variation of the loss distribution F_X is not considered.

As can be seen in Figures 2.1 and 2.2, both loss distributions have the same expected value, and thus by the equivalence principle they are to be charged the same premium despite the significant difference in their other characteristics other than the first moment. Here are two other approaches for pricing that consider the variation in loss distribution in addition to its

mean: Variance principle where a risk loading that is proportional to the variance is added:

$$\mathcal{H}_{\text{var}}(X) = \mathbb{E}(X) + \beta\sigma^2(X), \beta > 0$$

Also, the standard variation principle, where, this time, a risk loading proportional to the standard deviation is added:

$$\mathcal{H}_{\text{dev}}(X) = \mathbb{E}(X) + \beta\sigma(X), \beta > 0$$

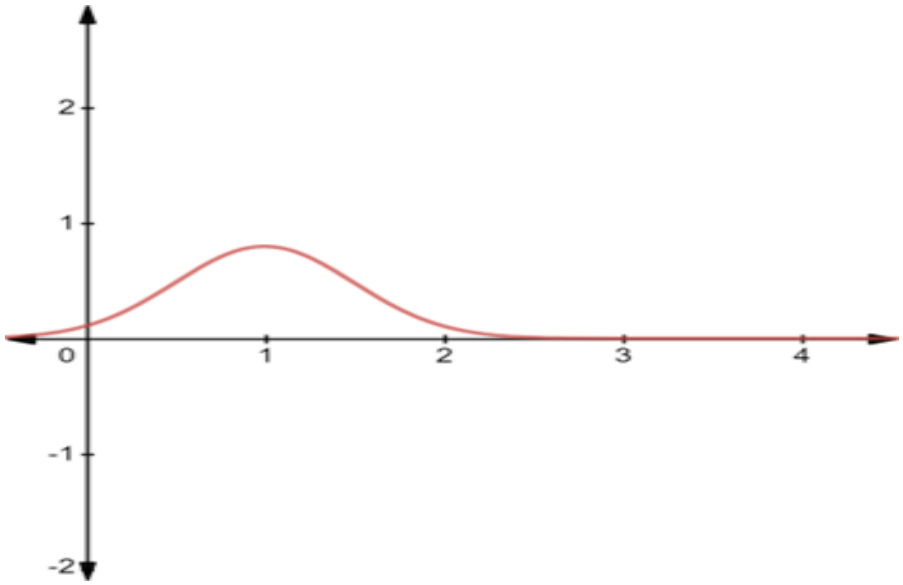


Figure 2.1: Normal Distribution Losses with Mean of 1 and Variance of 0.5

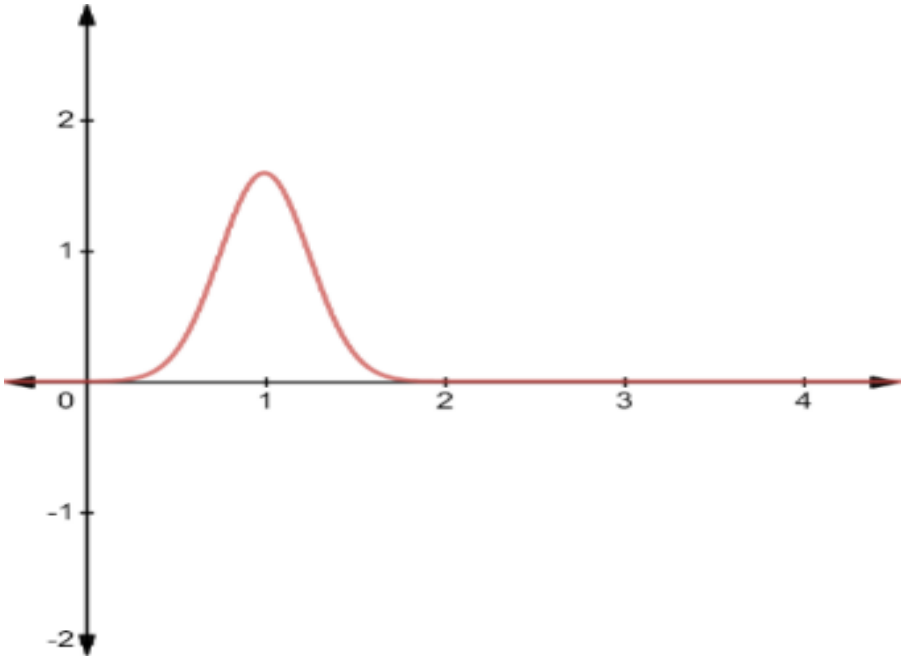


Figure 2.2: Normal Distributed Loses with Mean of 1 and Variance of 0.25

The variance and standard deviation principles produce higher rates for risks with higher volatility by including an extra term that serves as a guard against risk variation in comparison with the equivalence principle. Shortage of statistical data can make it difficult to recognize the loss distribution F_X fully, which may leave the decision-maker with no choice but to use principles that require minimum knowledge ofF_X, such as principles that only require the first and second moments of loss distribution.

Remark 2.3. If the calculated value of P is infinite then the insurer views the underlying risk as too high to cover.

When a decision-maker fails to choose an appropriate loss function that respects the characteristics of the loss distribution, thus producing an inappropriate premium principle, it leads either to unfair premium pricing or to wrongful disregard of the risk.

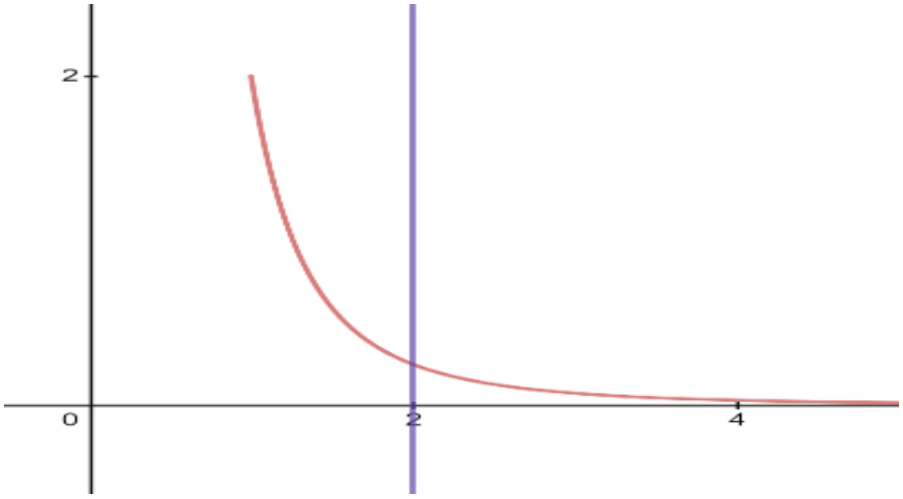


Figure 2.3: Pareto I Distributed Losses with A Mean of 2. The Purple Line Represents the Equivalence Principle Premium

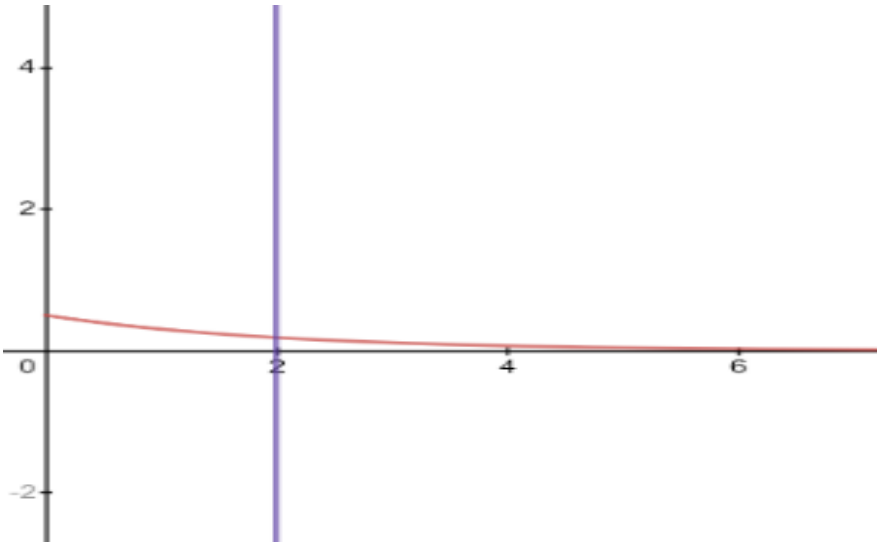


Figure 2.4: Exponentially Distributed Losses with Mean of 2. The Purple Line Represents the Equivalence Principle Premium

Bühlmann and Gisler (p.8 [3]) referred to the above-mentioned premium principles as 'classical' pricing rules. This is their summary list:

1. Expectation premium principle
2. Standard deviation principle
3. Variance principle
4. Exponential principle.

In the following section, the discussion is dedicated to the *exponential premium principle*, its derivation, and some of its properties.

The Exponential Premium Calculation Principle

As stated previously, some loss distributions are asymmetric. When a loss distribution is skewed and light-tailed (see Figures 2.3 and 2.4), such characteristics of the loss distribution cause the decision-maker to consider an exponential loss function with the following form:

$$l_{ex}(X, P) = E[(e^{\alpha X} - e^{\alpha P})^2], \quad \alpha > 0 \qquad (2.7)$$

Heilmann [7] stated this exponential loss function and proposed that it is possible to use the previous optimization procedure applied to the derivation of the equivalence principle (2.3) from the quadratic loss function (2.1). Partial differentiation with respect to P is taken and equated to zero to attain the desired minimizer $H_{ex}(X)$:

$$\frac{\partial}{\partial P} l_{ex}(X,P) = 0$$

$$\begin{aligned}\frac{\partial}{\partial P}\mathbb{E}[(e^{\alpha X} - e^{\alpha P})^2] &= \frac{\partial}{\partial P}\mathbb{E}[e^{2\alpha X} - 2e^{\alpha X}e^{\alpha P} + e^{2\alpha P}] \\ &= \frac{\partial}{\partial P}\mathbb{E}[e^{2\alpha X}] - 2\mathbb{E}[e^{\alpha X}]\frac{\partial}{\partial P}e^{\alpha P} + \frac{\partial}{\partial P}e^{2\alpha P} \\ &= -2\alpha\mathbb{E}[e^{\alpha X}]e^{\alpha P} + 2\alpha e^{2\alpha P} = 0\end{aligned}$$

So, $e^{2\alpha P} = \mathbb{E}[e^{\alpha X}]e^{\alpha P}$

$$P = \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha X}]$$

We conclude that:

$$\mathcal{H}_{ex}(X) = \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha X}] \qquad (2.8)$$

The Exponential Premium Principle as a Risk Measure

Just like a thermometer translates heat to Fahrenheit or degrees Celsius or as a weight scale translates weight to kilograms, pounds or whatever unit suits the intended objective, economists use risk measures to quantify a random risk, into a real number. Understandably, any economic entity wants to monitor the risks surrounding its area of business. This is especially true for the insurance sector, where risk is the commodity traded for profit.

Definition 2.4**.** A risk measure is any function $\rho: L^0 \to \mathbb{R}$ that maps a loss random variable to a real number.

Premium principles, as defined in Definition 2.2, clearly match the definition of a risk measure. Premium principles are risk measures in the insurance context. However, they must satisfy at least some properties to justify their use. In the following text it will be described how the various properties of the exponential premium principle as a risk measure make it a desirable approach to pricing in the actuarial realm. Wen et al. [12] and Kaas et al., p. 120-121 [9] showed that the exponential principle fulfills the following properties for any loss random

variables X, Y and a constant$\in \mathbb{R}$:

1. Monotonicity: $\mathcal{H}_{ex}(X) \leq \mathcal{H}_{ex}(Y)$, if $X \leq Y$ almost surely

Proof. Since the logarithmic and exponential functions are increasing, then

$$\begin{aligned} X \leq Y &\Rightarrow \alpha X \leq \alpha Y\,, \alpha > 0 \\ &\Rightarrow e^{\alpha X} \leq e^{\alpha Y} \\ &\Rightarrow \mathbb{E}(e^{\alpha X}) \leq \mathbb{E}(e^{\alpha Y}) \\ &\Rightarrow \ln \mathbb{E}(e^{\alpha X}) \leq \ln \mathbb{E}(e^{\alpha Y}) \\ &\Rightarrow \frac{1}{\alpha} \ln \mathbb{E}(e^{\alpha X}) \leq \frac{1}{\alpha} \ln \mathbb{E}(e^{\alpha Y}) \\ &\Rightarrow \mathcal{H}_{ex}(X) \leq \mathcal{H}_{ex}(Y) \end{aligned}$$

This means that this premium principle is somewhat logical, in that it will set a higher premium for higher risk.

2. No unjustified loading: $\mathcal{H}(\mathrm{c}) = \mathrm{c}$

Proof.

$$\mathcal{H}_{\mathrm{ex}}(\mathrm{c}) = \frac{1}{\alpha} \ln \mathbb{E}[\mathrm{e}^{\alpha \mathrm{c}}] = \frac{1}{\alpha} \ln \mathrm{e}^{\alpha \mathrm{c}} = \frac{\alpha \mathrm{c}}{\alpha} = \mathrm{c} \qquad (2.9)$$

3. No rip-off: $\mathcal{H}_{\mathrm{ex}}(\mathrm{X}) \leq \max[\mathrm{X}]$

Proof. Since the exponential premium principle is monotone and $X \leq \max(X)$:

$$\mathcal{H}_{ex}(X) \leq \mathcal{H}_{ex}(\max(X)) = \max(X) \qquad (2.10)$$

This is a useful property because whatever the price premium principle produces, it will never exceed the maximum value the loss X can possibly take. For example, if a museum owner wishes to buy a policy against theft, and the pieces on exhibit are worth one million pounds in total, it would be unreasonable to charge a premium over that amount to cover the policy.

4. Safety loading: $\mathcal{H}_{\mathrm{ex}}(\mathrm{X}) \geq \mathbb{E}[\mathrm{X}]$

Proof. Let the function $g: \mathbb{R} \to \mathbb{R}$ be defined by$g(x) = e^{\alpha x}$, then g is convex because$g'' > 0$. Hereby Jensen's inequality (cf. Definition A.17) states the following:

$$\begin{aligned} \mathbb{E}[e^{\alpha X}] &\geq e^{\alpha \mathbb{E}[X]} \\ \ln \mathbb{E}[e^{\alpha X}] &\geq \ln e^{\alpha \mathbb{E}[X]} \\ \frac{1}{\alpha} \ln \mathbb{E}[e^{\alpha X}] &\geq \frac{\alpha \mathbb{E}[X]}{\alpha} \\ \frac{1}{\alpha} \ln \mathbb{E}[e^{\alpha X}] &\geq \mathbb{E}[X] \end{aligned} \qquad (2.11)$$

The exponential premium principle guarantees a price greater than the loss mean. It anticipates positive gain on average$\mathcal{H}_{ex}(X) - \mathbb{E}[X] \geq 0$.

Combining the previous two properties ensures that the exponential premium principle gives a price somewhere between the mean and the maximum value of X, a range where both insurer and insured are, at a minimum, safe from extreme pricing situations, an observation noted down by Gerber and Jones in [6].

Mean loss ≤ exponential premium ≤ maximum loss

5. Invariant to displacements: $\mathcal{H}_{\text{ex}}(\text{X}+\text{c}) = \mathcal{H}_{\text{ex}}(\text{X}) + \text{c}$

Proof.

$$\begin{aligned}\mathcal{H}_{ex}(X+c) &= \frac{1}{\alpha}\ln \mathbb{E}\left[e^{\alpha(X+c)}\right] \\ &= \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha X}e^{\alpha c}] \\ &= \frac{1}{\alpha}\ln(e^{\alpha c}\mathbb{E}[e^{\alpha X}]) \\ &= \frac{1}{\alpha}(\ln e^{\alpha c} + \ln \mathbb{E}[e^{\alpha X}]) \\ &= \frac{1}{\alpha}(\alpha c + \ln \mathbb{E}[e^{\alpha X}]) = \mathcal{H}(X) + c\end{aligned}$$

The interpretation of property (5) is that if a loss is shifted by a certain amount, then the premium is shifted by that same amount without extra charges.

6. Additivity for independent risks: $\text{H}_{\text{ex}}(\text{X}+\text{Y}) = \text{H}_{\text{ex}}(\text{X})+\text{H}_{\text{ex}}(\text{Y})$ for independent X, Y

Proof.

$$\begin{aligned}\mathcal{H}_{ex}(X+Y) &= \frac{1}{\alpha}\ln \mathbb{E}\left[e^{\alpha(X+Y)}\right] \\ &= \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha X}e^{\alpha Y}] \\ &= \frac{1}{\alpha}\ln(\mathbb{E}[e^{\alpha X}]\mathbb{E}[e^{\alpha Y}]) \text{ by independence of } X \text{ and } Y \text{ (cf. A. 11)} \\ &= \frac{1}{\alpha}(\ln \mathbb{E}[e^{\alpha X}] + \ln \mathbb{E}[e^{\alpha Y}]) \\ &= \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha X}] + \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha Y}] \\ &= \mathcal{H}_{ex}(X) + \mathcal{H}_{ex}(Y)\end{aligned}$$

This proof implies that if an insurer wishes to calculate the overall premium for two independent risks, then this simply achieved by adding together each risk's premium.

7. Iteratively: $\mathcal{H}_{ex}(X) = \mathcal{H}_{ex}(\mathcal{H}_{ex}(X \mid Y))$; for all X, Y and the conditional premium principle given Y is defined by $\mathcal{H}_{ex}(X \mid Y) = \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha X} \mid Y]$

Proof.

$$\begin{aligned}\mathcal{H}_{ex}(\mathcal{H}_{ex}(X \mid Y)) &= \frac{1}{\alpha}\ln \mathbb{E}\left[e^{\alpha\mathcal{H}}\text{Hex}^{(X|Y)}\right] \\ &= \frac{1}{\alpha}\ln \mathbb{E}\left[\exp\left(\alpha\frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha X} \mid Y]\right)\right] \\ &= \frac{1}{\alpha}\ln \mathbb{E}\left[\mathbb{E}[e^{\alpha X} \mid Y]\right] \\ &= \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha X}] = \mathcal{H}_{ex}(X)\end{aligned}$$

Notice here that premium principle is applied twice, and this seems to cancel out the effect of conditioning on Y as a result of the Tower property of expectations (cf. Proposition

A.12). It is worth mentioning that the famous Tower property of expectations is also called the law of iterated expectations

The Exponential Premium Principle is not Coherent

Risk measures are the main topics discussed in quantitative risk management. When risk measures were introduced, their coherency as it is termed by Artzner et al. [2], was also discussed. However, the exponential principle is not coherent, and this impairs the claim that it is a good pricing tool.

Definition 2.5. (Coherent premium principle/risk measure) a risk measure ρ is said to be coherent if, for any losses X and Y and a constant c ∈ R, it fulfills the following properties:

1. Positive Homogeneity: $\rho(cX) = c\rho(X)$, $c > 0$
2. Invariant to displacements: $\rho(X + c) = \rho(X) + c$
3. Monotonicity: $\rho(X) \leq \rho(Y)$, if $X \leq Y$ almost surely
4. Subadditivity: $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

The exponential premium principle violates both the positive homogeneity and subadditivity conditions, showing it is not coherent.

Proof. The exponential premium principle is not sub-additive for dependent X, Y. Here we provide a counterexample. We consider a uniformly distributed random variable U on [0, 1] and define X = Y = ln (U)/α, then

$$\mathcal{H}_{ex}(X+Y) = -\ln(3)/\alpha \text{ and } \mathcal{H}_{ex}(X) + \mathcal{H}_{ex}(Y) = -2\ln(2)/\alpha$$
$$\text{then, } \mathcal{H}_{ex}(X+Y) > \mathcal{H}_{ex}(X) + \mathcal{H}_{ex}(Y)$$

Subadditivity relates to risk diversification: it means that aggregating risks together in a single portfolio is less risky than taking each risk individually. As a premium principle. For example take the scenario: (i) A portfolio of fire insurance for an entire city's housing stock. On the other hand, another portfolio (ii) consists of houses, each located in a different city. The portfolio (i) is charged a lower premium, although it is a better candidate for greater simultaneous loss; the houses are next to each other, so if one catches fire it is more likely that the adjacent house, which is also in that portfolio, catches fire. This property does not differentiate between positively dependent and independent losses.

Regarding the property of positive homogeneity, it does not hold. The counterexample given above states that$\mathcal{H}_{ex}(2X) \neq 2\mathcal{H}_{ex}(X)$.

Criticisms of subadditivity, positive homogeneity, and other properties of risk measures in insurance practice can be found in Dhaene et al. [8].

Premium Principles in a Bayesian Framework

In Chapter 1, it was mentioned that insurers classify data and group them based on some degree of similarity in their quantifiable characteristics. However, al- though it is possible to sort, for example, insured car drivers by the number of cylinders in the engines of the cars that they drive, it is not possible to quantify those drivers' sense of responsibility. The latter characteristics are called risk pro- files, and the groups are called collectives. Pricing rules introduced in Chapter 2 are functions that take the unclassified loss random variable as an argument. In this chapter, Bayes theory and risk theory are simultaneously used to enable premium calculation principles, particularly equivalence and exponential principles, to take the classified risk as an argument and give the desired corresponding price for the insured's individual experience. Ways of estimating individual premiums under equivalence and exponential principles are introduced. This helps in the identification of the link between credibility theory and the exponential premium principle

that takes place later in this chapter.

Throughout this chapter the following definitions and assumptions hold:

- We consider the sequence $(X_1, \dots, X_n, X_{n+1})$ of identically distributed random variables as losses in a collective in years $1,2, \dots, n, n+1$ respectively. X_1 Is a bounded non-negative random variable, with finite moments and its moment generating function exists, and denote f_{X_1} the probability distribution function (pdf), and by F_{X_1} the cumulative distribution function forX_1. Further let the vector $\vec{X} = (X_1, X_2, \dots, X_n)$ to be the previous claim observations vector from year 1 through$n$, and $\vec{x}$ is a realization of$\vec{X}$.
- We define Range $(V) := \{v \in \mathbb{R} \mid f_V(v) > 0\}$ where f_V is the pdf of the random variable$V$.

Assumption 3.1. The risk profile that produced a risk X_j is unknown/random. The risk profile is denoted θ as a sample from the random variableΘ, with f_Θ the probability distribution function forΘ, and F_Θ the cumulative distribution function. The sample space of Θ is the set Range (Θ).

Assumption 3.2**.** Given the risk profile $\Theta = \theta, X_j$'s are independent and identically distributed. It is to be noted that X_j 's and Θ are dependent.

- We define prior cumulative distribution function:

$$F_\Theta(\theta) \to [0,1], \theta \in \text{Range}(\Theta)$$

- Define posterior cumulative distribution function:

$$F_{\Theta|X}(\theta) \to [0,1], \theta \in \text{Range}(\Theta)$$

- We define loss function l as in Definition 2.1 from Chapter 2.

Bayesian Learning

The a priori distribution represents the actuary's initial view about the likely distribution of Θ. That view is formed at the beginning of the year, before the arrival of any claims. At the end of the year, in the light of new data, a better estimate of the initial distribution is formed. This updated view is called posterior distribution and it is mathematically computed via Bayes theorem (cf. Theorem A.5).

The following example is intended to illustrate the corresponding distributions at each stage of Bayesian learning.

Example 3.3. The number of clients' meetings in a given day j(j = 1, 2...) at a law firm $N_j|\Lambda = \lambda$ are independent Poisson random variables with mean λ that varies according to the time a client spends waiting before entering an attorney's office. The Poisson's pdf is given by

$$f_{N|\Lambda}(n \mid \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}, \lambda > 0, n = 0,1,2, \dots$$

The firm hires a statistician to give an insight into the distribution of waiting times. As a starting point, the statistician uses the prior distribution ofΛ, which is a gamma distribution expressed by the following pdf

$$f_\Lambda(\lambda) = \frac{2^5}{24}\lambda^{5-1}e^{-2\lambda}, \lambda > 0$$

The firm allows the statistician to spend time in the firm to collect data. In the following two

days, a randomly selected attorney meets five clients on day one and three clients on day two. Now, the statistician applies Bayes theorem (cf. Theorem A.5) to update the prior distribution as follows:

$$f_{\Lambda|N_1=5,N_2=3}(\lambda) = \frac{f_{N_1,N_2|\Lambda}(5,3 \mid \lambda) \cdot f_{\Lambda}(\lambda)}{f_{N_1,N_2}(5,3)}$$
$$= \frac{1}{f_{N_1,N_2}(5,3)} \cdot \left(\frac{\lambda^5 e^{-\lambda}}{5!}\right)\left(\frac{\lambda^3 e^{-\lambda}}{3!}\right)\left(\frac{2^5}{24}\lambda^{5-1}e^{-2\lambda}\right)$$

The numerator $f_{N_1=5,N_2=3}(n)$ is a constant. So, for simplicity, all the constants are gathered under one notation$c$, then the posterior distribution is given as:

$$f_{\Lambda|N_1=5,N_2=3}(\lambda) = c \cdot \lambda^{12}e^{-4\lambda}$$

The posterior found is a pdf function and it must have an integral of one over all values of λ (cf. Definition A.2),

$$\int_0^{\infty} c \cdot \lambda^{12}e^{-4\lambda}d\lambda = 1$$

This looks like a form of Gamma pdf. It is possible to pull the constant c out of the integral and multiply and divide by another proper constant to create an integral of the Gamma~ (13, 4) pdf to facilitate the integration

$$c \cdot \frac{\Gamma(13)}{4^{13}} \underbrace{\int_0^{\infty} \frac{4^{13}}{\Gamma(13)}\lambda^{13-1}e^{-4\lambda}d\lambda}_{=1} = 1$$

Consequently: $c = \frac{4^{13}}{\Gamma(13)}$

And hence: Λ| (N1, N2) = (2, 3) ~ Gamma (13, 4). This is the distribution of Λ after learning. It is the posterior distribution.

Collective Risk Premium: Price for Everyone

In insurance, a premium is always set for the amount of a claim that will occur in the future; otherwise, it is not insurance. Thus, proper forecasting is a necessity. An actuary currently in year n has to set a price for a claim that may arrive in year n+1. It is believed that the risks in a collective are similar but different (see Chapter 1), and the premium that expresses this similarity is called the collective risk premium. It is, loosely speaking, an average claim value for all profiles. Therefore, Xn+1 is not conditioned on another random variable representing the profile. Under the equivalence principle, the collective premium is

$$\mu_0 := \mathbb{E}[X_{n+1}]$$

Under the exponential premium principle, the collective premium is:

$$\tilde{\mu}_0 := \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha X_{n+1}}]$$

Why not stop here?

As in any line of business, insurance companies want to earn more than they lose while managing competition with their peer companies. Of course, if a premium is calculated as in (3.1) and (3.2), the actual risks will sometimes be higher and sometimes lower. It is easy to imagine that this would be very convenient for customers with higher risks, while the same cannot be said for those with lower risks. This is called anti-selection, where the company

becomes attractive to the worst risks, while the good risks seek fairer premiums elsewhere. This phenomenon is explained in Bühlmann and Gisler [3] (p. 11). In Chapter 1, it is emphasized that no collective of risks, called classes, is homogeneous. Grouping risks reduces the variation between them, but it does not thereby vanish. All collectives are naturally heterogeneous because individuals/policyholders are naturally different. Therefore, collective risk premiums are especially unfair to those who stand out the most, among other risks, in their collective. This motivates insurers to consider individual risk premiums, which represent, again in loose terms, the average, but specifically the average

Over risks within a risk profile.

Individual Risk Premium: More Tailored Rating

Here, the concept of experience rating becomes relevant, with the idea of a premium that incorporates past information for rate-making. In other words, the data produced by the individual's profile is used. The following definition is found in Bühlmann and Gisler , [3] (p. 9).

Definition 3.4**.** (Individual equivalence premium) For Xj (j = 1, ..., n) an observed loss in year j and risk profile Θ, the individual premium with respect to the equivalence principle for next year's loss Xn+1 is

$$\mu_\Theta := \mathbb{E}[X_{n+1} \mid \Theta]$$

Heilmann [7] also showed how a 'profiled' individual exponential premium eμΘ can be obtained by minimizing the conditional exponential loss function with respect to eμΘ. The following definition is proposed by Heilmann [7]

Proposition 3.5**.** (Individual exponential premium) For Xj (j = 1, 2..., n) an observed loss in year j and risk profile Θ, the individual premium with respect to the exponential loss function (2.7) for next year's loss Xn+1 is:

$$\tilde{\mu}_\Theta := 1/\alpha \ln E[e^{\alpha X_{n+1}} \mid \Theta]$$

Proof.

$$\begin{aligned}
\frac{\partial}{\partial \tilde{\mu}_\Theta} l(X_{n+1}, \tilde{\mu}_\Theta)\Big| \Theta &= \frac{\partial}{\partial \tilde{\mu}_\Theta} \mathbb{E}[(e^{\alpha X_{n+1}} - e^{\alpha \tilde{\mu}_\Theta})^2 \mid \Theta] \\
&= \frac{\partial}{\partial \tilde{\mu}_\Theta} \mathbb{E}\left[e^{2\alpha X_{n+1}} - 2e^{\alpha X_{n+1}} e^{\alpha \tilde{\mu}_\Theta} + e^{2\alpha \tilde{\mu}_\Theta} \mid \Theta\right] \\
&= \frac{\partial}{\partial \tilde{\mu}_\Theta} \mathbb{E}[e^{2\alpha X_{n+1}} \mid \Theta] - 2\mathbb{E}[e^{\alpha X_{n+1}} \mid \Theta] \frac{\partial}{\partial \tilde{\mu}_\Theta} e^{\alpha \tilde{\mu}_\Theta} + \frac{\partial}{\partial \tilde{\mu}_\Theta} e^{2\alpha \tilde{\mu}_\Theta} \\
&= -2\alpha \mathbb{E}[e^{\alpha X_{n+1}} \mid \Theta] e^{\alpha \tilde{\mu}_\Theta} + 2\alpha e^{2\alpha \tilde{\mu}_\Theta}
\end{aligned}$$

Equating the differentiated Bayes risk to zero we get

$$\begin{aligned}
&\to e^{2\alpha \tilde{\mu}_\Theta} = \mathbb{E}[e^{\alpha X_{n+1}} \mid \Theta] e^{\alpha \tilde{\mu}_\Theta} \\
&\to e^{\alpha \tilde{\mu}_\Theta} = \mathbb{E}[e^{\alpha X_{n+1}} \mid \Theta] \\
&\to \ln e^{\alpha \tilde{\mu}_\Theta} = \ln \mathbb{E}[e^{\alpha X_{n+1}} \mid \Theta] \\
&\to \alpha \tilde{\mu}_\Theta = \ln \mathbb{E}[e^{\alpha X_{n+1}} \mid \Theta] \\
\to \tilde{\mu}_\Theta &= \frac{1}{\alpha} \ln \mathbb{E}[e^{\alpha X_{n+1}} \mid \Theta]
\end{aligned}$$

The same idea of Bayesian learning applied in Example 3.3 can also be applied to premium principles. In contrast to collective premiums, the premiums to be introduced in the next section are ones that can actually learn and improve as new data becomes available.

Bayes Estimators and Premium Principles

The process throughout this section is essentially that of decision theory, but random variables here have a priori and posterior distributions. Thus, it is decision theory with a Bayesian structure. As in Chapter 2, an insurer wants to find a premium that minimizes the expected loss. In this section, the goal is particularly to find prices that are estimators of the individual premiums (3.4) and (3.5) by performing minimization procedures to the expected loss functions, to achieve the best estimator among all possible estimators, that is, the Bayes rule. However, there is an alternative equivalent method that gives the same best estimator or Bayes rule which is minimizing the posterior risk function.

Definition 3.6. (Posterior risk) For the past observations vector$\vec{X}$, let β be some unknown parameter. $\hat{\beta}(\vec{X}) = \hat{\beta}$ Is some estimator of β and for a loss function l defined as in (Definition 2.1), $\mathbb{E}[l(\beta, \hat{\beta}) \mid \vec{X}]$ is the posterior risk of$\hat{\beta}$.

Bayes Premium under the Equivalence Principle

As the individual equivalence premium μ_Θ given in Theorem 3.7 is a conditional expectation, it is a random variable. Its exact value can only be estimated. The estimation is obtained by solving a minimization problem of the posterior risk of $\hat{\mu}_\Theta$ over the set of all possible individual equivalence premium estimators $\mathbf{A} = \{\hat{\mu}_\Theta(\vec{X}): \hat{\mu}_\Theta < \infty\}$ using the quadratic loss function l (2.1),

$$\min_{\widehat{\mu_\Theta} \in A} \mathbb{E}[l(\mu_\Theta, \widehat{\mu_\Theta})] \tag{3.3}$$

In order to attain the best individual equivalence premium estimator$\hat{\mu}_\Theta^*$, Heilmann [7] gave the form of the Bayes equivalence premium $\hat{\mu}_\ominus^*$ via the following theorem:

Theorem 3.7. (Bayes equivalence premium) For the previous observations vector $\vec{X}$ and risk profileΘ, the Bayes premium with respect to the quadratic loss function (2.1), assuming $\hat{\mu}_\Theta^*$ is a realistic premium to be charged,$\hat{\mu}_\Theta^* \in \mathbf{A}$, is

$$\hat{\mu}_\ominus^* := \mathbb{E}[\mu_\Theta \mid \vec{X}] \tag{3.4}$$

Proof. Conditional expectations are linear operators, as are non-conditional expectations. Additionally, the Tower property of expectation is used. Now, for any possible individual premium estimator $\hat{\mu}_\Theta \in \mathbf{A}$

$$\begin{aligned}\mathbb{E}[L(\mu_\Theta, \hat{\mu}_\Theta) \mid \vec{X}] &= \mathbb{E}[(\mu_\Theta - \hat{\mu}_\Theta)^2 \mid \vec{X}] \\ &= \mathbb{E}[(\mu_\Theta - \hat{\mu}_\Theta)^2 \mid \vec{X}] \\ &= \mathbb{E}[\mu_\Theta^2 - 2\mu_\Theta\hat{\mu}_\Theta + \hat{\mu}_\Theta^2 \mid \vec{X}] \\ &= \mathbb{E}[\mu_\Theta^2 \mid \vec{X}] - 2\mathbb{E}[\mu_\Theta \mid \vec{X}]\hat{\mu}_\Theta + \hat{\mu}_\Theta^2\end{aligned}$$

Minimization is obtained via partial differentiation with respect to $\hat{\mu}_\Theta$

$$\frac{\partial}{\partial \hat{\mu}_\Theta}\left(\mathbb{E}[\mu_\Theta^2 \mid \vec{X}] - 2\mathbb{E}[\mu_\Theta \mid \vec{X}]\hat{\mu}_\Theta + \hat{\mu}_\Theta^2\right) = -2\mathbb{E}[\mu_\Theta \mid \vec{X}] + 2\hat{\mu}_\Theta \tag{3.5}$$

Equating (3.5) to zero

Equating (3.5) to zero

$$-2\mathbb{E}[\mu_\Theta \mid \vec{X}] + 2\hat{\mu}_\Theta = 0$$

$$\rightarrow \hat{\mu}_\Theta^* = \mathbb{E}[\mu_\Theta \mid \vec{X}]$$

The Bayes premium under the quadratic loss function has a nice property: if the insurer does

not have past claim experience, it produces the collective risk premium.

Bayes Premium under the Exponential Principle

In Definition 3.4, the individual exponential premium $\tilde{\mu}_\Theta$ is random due to the fact that Θ is unknown. Hence, it has to be estimated. Heilmann [7] obtained the optimal individual exponential premium estimator $\hat{\tilde{\mu}}^*_\Theta$ by minimizing Posterior risk of $\hat{\tilde{\mu}}_\Theta$ this time using the exponential loss function over the set of all individual exponential premium estimators$\tilde{A} = \{\hat{\tilde{\mu}}_\Theta(\vec{X}): \hat{\tilde{\mu}}_\Theta < \infty\}$.

$$\min_{\tilde{\mu}_\Theta \in \tilde{\mathbb{A}}} \mathbb{E}\left[L\left(\tilde{\mu}_\Theta, \hat{\tilde{\mu}}_\Theta\right) \mid \vec{X}\right] = \mathbb{E}\left[L\left(\tilde{\mu}_\Theta, \tilde{\tilde{\mu}}^*_\Theta\right) \mid \vec{X}\right]$$

The following theorem can be found in Heilmann [7]:

Theorem 3.8. (Bayes exponential premium) For the previous observations vector $\vec{X}$ and a risk profileΘ, the Bayes premium with respect to the exponential loss function (2.7), assuming $\hat{\tilde{\mu}}^*_\Theta$ is a realistic premium to be set, that is,$\hat{\tilde{\mu}}^*_\Theta \in \tilde{A}$, is

$$\hat{\tilde{\mu}}^*_\ominus := \frac{1}{\alpha} \ln \mathbb{E}\left[e^{\alpha\tilde{\mu}\tilde{\theta}} \mid \vec{X}\right]$$

Proof. As stated above, conditional expectations are linear operators, as are nonconditional expectations. Now, for any possible individual exponential premium estimator $\hat{\tilde{\mu}}_\Theta \in \tilde{A}$

$$\begin{aligned}
\mathbb{E}\left[L\left(\tilde{\mu}_\Theta, \hat{\tilde{\mu}}_\Theta\right) \mid \vec{X}\right] &= \mathbb{E}\left[\left(e^{\alpha_\mu} - e^{\alpha^\theta_{\hat{\mu}}}\right)^2 \mid \vec{X}\right] \\
&= \mathbb{E}\left[e^{2\alpha\tilde{\mu_\theta}} - 2e^{\alpha\tilde{\mu}\tilde{\mu}_\theta} e^{\alpha\hat{\mu}} + e^{2\alpha\tilde{\mu_\theta}} \mid \vec{X}\right] \\
&= \mathbb{E}\left[e^{2\alpha\tilde{\mu}_\theta} \mid \vec{X}\right] - 2\mathbb{E}\left[e^{\alpha\tilde{\mu}_\theta} \mid \vec{X}\right] e^{\alpha\hat{\tilde{\mu}_\theta}} + e^{2\alpha\tilde{\mu}_\Theta}
\end{aligned}$$

Minimization is obtained via partial differentiation with respect to $\hat{\tilde{\mu}}_\Theta$

$$\frac{\partial}{\partial \hat{\tilde{\mu}}_\Theta}\left(\mathbb{E}\left[e^{2\alpha\tilde{\mu}\tilde{\mu}_\theta} \mid \vec{X}\right] - 2\mathbb{E}\left[e^{\alpha\tilde{\mu}\Theta} \mid \vec{X}\right] e^{\alpha\hat{\tilde{\mu}}_\theta} + e^{2\alpha\hat{\tilde{\mu}}_\Theta}\right) = -2\mathbb{E}\left[e^{\alpha\tilde{\mu}\Theta} \mid \vec{X}\right] + 2e^{\alpha\tilde{\mu}_\Theta}$$

Equating (3.8) to zero

$$\begin{aligned}
-2\mathbb{E}\left[e^{\alpha\tilde{\mu}_\Theta} \mid \vec{X}\right] + 2e^{\alpha\tilde{\mu}_\Theta} &= 0 \\
e^{\alpha\tilde{\mu}_\Theta} &= \mathbb{E}\left[e^{\alpha\tilde{\mu}_\Theta} \mid \vec{X}\right] \\
\rightarrow \hat{\tilde{\mu}}^*_\Theta &= \frac{1}{\alpha} \ln \mathbb{E}\left[e^{\alpha\tilde{\mu}_\Theta} \mid \vec{X}\right]
\end{aligned}$$

Why not stop here?

Bühlmann and Gisler, [3] (p. 49) imposed simplicity as a requirement of a premium formula. They had a real world insight regarding the simplicity of the Bayes premium. In order to find the Bayes premium, it is necessary to know the prior and the posterior distributions. Neither of these is easy to specify in practice. The Bayes premium is typically found through complicated methods and it cannot be found by any simple means.

The Bayes premium is the best individual premium estimator among the class of all possible individual premium estimators. To solve the complexity problem as per Bühlmann and Gisler, [3] (p. 55), it is possible to confine the 'search' to some narrower class. That is, the class of the individual premium estimators that are linear in observations. So, this time the Bayes rule is the best linear Bayes premium among all possible linear Bayes premiums. An application of linear Bayesian methods is exhibited in credibility theory, as per Wen et al. [12]. In the context

of credibility theory the best linear Bayes premium is called the credibility premium; this is developed further in the next section.

Premium Principles and Credibility Theory

Take the case of car insurance, for example. Ideally, those insured should be charged premiums that correspond to their own behavior. If an insured person drives safely but belongs to a high-risk collective, then one might expect an insurer to charge him or her a higher premium than the insured 'deserves', and unfortunately, this might be the case. It is stated in Bühlmann and Gisler, [3] (p.2) that, in practice, the observed individual experience data is too limited to be eligible for the application of the law of large numbers. Bayesian statistics allow the use of both sources of information: from the collective experience, and from the individual experience via credibility theory; this combination of information

Contributes to the calculation of the credibility premium.

The Credibility Premium under the Equivalence Principle

Estimating individual premiums has been the goal since Subsection 3.1.2. For the individual equivalence premium (see Theorem 3.7), the intuition here is similar to the one in Subsection 3.2.1. The objective is to find an individual equivalence premium estimator that minimizes the expectation of the quadratic loss function (2.1). However, an important distinction is that there is a restriction to the pool of μ_Θ estimators. Now, the minimization is over the set of all individual equivalence premiums estimators are linear with respect to observations, denoted by the setD $= \{P^c: P^c = a_0 + \sum_{j=1}^{n} a_j X_j;\ a_0, a_{j(j=1,\dots,n)} \in \mathbb{R}\}$. The goal is to find the best linear estimator$\hat{P}^c$.

Let L be the quadratic loss function (2.1). The optimization problem can be performed as:

$$\min_{P \in D} \mathbb{E}[L(\mu_\Theta, P^c)] \qquad (3.9)$$

In order to attain the best linear individual premium estimator, that is, the credibility premium$\hat{P}^c$.

Assumption 3.9. (Simple credibility model building blocks)

- Assumptions 3.1 and 3.2.
- $\mu_\Theta = \mathbb{E}[X_j \mid \Theta]$
- $\mu_0 = \mathbb{E}[X_j] = \mathbb{E}\left[\mathbb{E}[X_j \mid \Theta]\right]$ using Tower property (cf. Proposition A.12)
- $\sigma_\Theta^2 = \mathrm{Var}(X_j \mid \Theta)$
- $\tau^2 = \mathrm{Cov}(X, \mu_\Theta) = \mathrm{Var}(\mu_\Theta)$

A Recap from Probability theory

- $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$
- $\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n} X_j\right] = \frac{n\mathbb{E}[X_1]}{n} = \mathbb{E}[X_1] = \mu_0$
- $\mathrm{Var}(X_1) = \mathbb{E}[\mathrm{Var}(X_1 \mid \Theta)] + \mathrm{Var}(\mathbb{E}[(X_1 \mid \Theta)]$ by law of total variation (cf. A.15)

The following theorem can be found in Bühlmann and Gisler [3], (p. 56). It supplies the general

form of the credibility premium:

Theorem 3.10. Under Assumption 3.9, the general credibility estimator is as follows:

$$\hat{P}^c = Z\bar{X} + (1-Z)\mu_0 \tag{3.10}$$

Where $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$ and Z is the credibility factor

$$Z = \frac{n}{n+\sigma^2/\tau^2} \tag{3.11}$$

Proof. The goal is to minimize the quadratic loss function (2.1) with respect to $\hat{P}^c$

$$\min_{\widehat{P_c^c}\in D} \mathbb{E}\left[\left(\mu_\Theta - \widehat{Pc}\right)^2\right] = \min_{a_0, a_{j(j=1,\dots,n)}\in\mathbb{R}} \mathbb{E}\left[\left(\mu_\Theta - a_0 - \sum_{j=1}^{n} a_j X_j\right)^2\right]$$

Let $\widehat{a_0}$ and $\hat{a}_{j;j=1,\dots,n}$ be the solution to the minimization problem, then:

$$\mathbb{E}\left[\left(\mu_\Theta - \hat{a}_0 - \sum_{j=1}^{n} \hat{a}_j X_j\right)^2\right] = \min_{a_0, a_{j(j=1,\dots,n)}\in\mathbb{R}} \mathbb{E}\left[\left(\mu_\Theta - a_0 - \sum_{j=1}^{n} a_j X_j\right)^2\right]$$

Hence the target credibility premium has the form:

$$\hat{P}^c = \hat{a}_0 + \sum_{j=1}^{n} \hat{a}_j X_j \tag{3.12}$$

Losses in the collective have the same distribution. Thus, temporal order does not interfere with the calculations here and since there is only one best linear individual premium estimator, then:

$$\hat{a}_1 = \hat{a}_2 = \cdots = \hat{a}_n$$

Then it is possible to unify the notation for simplicity$\hat{a}_1 = \hat{a}_2 = \cdots = \hat{a}_n = \hat{b}_n$. Afterward, it will be easier to get rid of the summation operator

$$\hat{P}^c = \hat{a} + \hat{b}\bar{X} \tag{3.13}$$

Where what we are trying to obtain, $\hat{a}$ and$\hat{b}$, are the solutions to the following

$$\mathbb{E}\left[\left(\mu_\Theta - \hat{a} - \hat{b}\bar{X}\right)^2\right] = \min_{a,b\in\mathbb{R}} \mathbb{E}[(\mu_\Theta - a - b\bar{X})^2] \tag{3.14}$$

There are two variables to minimize, thus, partial differentiation is to be carried out to minimize the expected loss. Starting with a

$$\frac{\partial}{\partial a}\mathbb{E}[(\mu_\Theta - a - bX)^2] = -2\mathbb{E}[(\mu_\Theta - a - bX)] \tag{3.15}$$

Equating (3.15) to zero

$$\begin{aligned} -2\mathbb{E}[\mu_\Theta - a - b\bar{X})] &= 0 \\ \mathbb{E}[\mu_\Theta] - \mathbb{E}[a] - b\mathbb{E}[\bar{X}] &= 0 \\ \mu_0 - a - b\mu_0 &= 0 \\ \rightarrow \hat{a} &= (1-b)\mu_0 \end{aligned} \tag{3.16}$$

Substituting (3.16) in (3.14)

$$\begin{aligned}\mathbb{E}[(\mu_\Theta-(1-b)\mu_0-b\bar{X})^2] &= \mathbb{E}[(\mu_\Theta-\mu_0+b\mu_0-b\bar{X})^2]\\ &= \mathbb{E}\left[\left((\mu_\Theta-\mu_0)-b(\bar{X}-\mu_0)\right)^2\right]\\ &= \mathbb{E}[(\mu_\Theta-\mu_0)^2-2b(\mu_\Theta-\mu_0)(\mu_0-\bar{X})+b^2(\mu_0-\bar{X})^2]\\ &= \mathbb{E}[(\mu_\Theta-\gamma_0)^2]-2b\mathbb{E}[(\mu_\Theta-\mu_0)(\mu_0-\bar{X})]+b^2\mathbb{E}[(\mu_0-\bar{X})^2],\end{aligned}$$

Taking partial differentiation with respect to b and equating to zero

$$\begin{aligned}-2\mathbb{E}\big[(\bar{X}-\mathbb{E}[\bar{X}])\big((\mu_\Theta-\mathbb{E}[\mu_\Theta])\big]+2b\mathbb{E}[(\bar{X}-\mathbb{E}[\bar{X}])^2]=0\\ \mathbb{E}\big[(\bar{X}-\mathbb{E}[\bar{X}])\big((\mu_\Theta-\mathbb{E}[\mu_\Theta])\big]=b\mathbb{E}[(\bar{X}-\mathbb{E}[\bar{X}])^2],\end{aligned}$$

Spotting covariance and variance forms appearing (cf. A.3) yields the following:

$$\mathrm{Cov}(\bar{X},\mu_\Theta)=b\mathrm{Var}(\bar{X})$$

Solving for b to get the minimizer

$$\hat{b}=\frac{\mathrm{Cov}(\bar{X},\mu_\Theta)}{\mathrm{Var}(X)} \qquad (3.17)$$

Given that the model Assumption 3.9 holds, and via the dependency of X and Θ (cf. A.3)

$$\mathrm{Cov}(\bar{X},\mu_\Theta)=\mathrm{Var}(\mu_\Theta)=:\tau^2 \qquad (3.16)$$

$$\mathrm{Var}(\bar{X})=\frac{\mathbb{E}[\sigma_\Theta^2]}{n}+\mathrm{Var}(\mu_\Theta)=:\frac{\sigma^2}{n}+\tau^2 \qquad (3.17)$$

The b minimizer (3.14) is formed as follows

$$\hat{b}=\frac{\tau^2}{\tau^2+\frac{\sigma^2}{n}}=\frac{n}{n+\frac{\sigma^2}{\tau^2}} \qquad (3.18)$$

Finally, let $Z=\hat{b}$ then plugging Z into the equation (3.16), it holds $\hat{a}=(1-\mathrm{Z})\,\mu_0$.

Discussion: Credibility Premium Components

The estimator X that appears in the formula (3.10), represents the individual observed risks average. It can be said that $\bar{X}$ is the component that speaks for past individual experience.

This makes the word credibility very intuitive. How credible are the past experience X_j s ?

The element n in the credibility factor is the number of years/data points. By the law of large numbers, the precision of individual premium estimation can be associated with the amount of data an insurer has at his or her disposal. As the amount of data increases, the credibility factor also increases. Ultimately, if the amount tends to approach infinity, then the credibility factor is 1.

$$\lim_{n\to\infty}\frac{\tau^2}{\tau^2+\frac{\sigma^2}{n}}=1$$

This, of course, is an impossible scenario. But the underlying idea is that, the greater the knowledge of an individual's profile, the more an insurer tends to depend on individual experience, and the less reliance is placed on the collective experience. From another perspective, the insurer starts with the collective premium and adjusts toward individual experience, according to the quality of new individual experience data.

It must be borne in mind that μ_Θ is a random variable. It contributes to the general credibility factor with its variationτ^2. In Statistics, the term variance refers to how a random variable fluctuates about its mean. The mean of μ_Θ is μ_0 (see Assumption 3.9). Hence, the variance of μ_Θ represents how scattered the individual premium is from the collective premiumμ_0.

Attending to the credibility factor, we take two extreme cases of τ^2 defined as in equation (3.16):

$$\lim_{\tau^2 \to \infty} \frac{n}{n+\frac{\sigma^2}{\tau^2}} = 1 \qquad (3.19)$$

$$\lim_{\tau^2 \to 0} \frac{n}{n+\frac{\sigma^2}{\tau^2}} = 0 \qquad (3.20)$$

An insight into the heterogeneity of the collective can be gained by looking atτ^2. If an individual experience is absolutely different from the collective, as in equation (3.19), then, there are no grounds for including the collective premium in that particular credibility premium estimation procedure and vice versa.

$X_j \mid \Theta$ Are the data within a risk profile Θ.$\mathrm{Var}(X_j \mid \Theta)$ measures the volatility of data within a risk profile and, if its expected value is taken, the result is$\mathbb{E}[\mathrm{Var}(X_j \mid \Theta)] = \sigma^2$. It is a component of the credibility factor that measures the mean variation of risks within a risk profile. In other words, lower σ^2 means that the within risks are 'close' to each other and this increases the credibility of these data. Hence, the smaller the σ^2 the higher the credibility factor$Z$, since quality within risks data means quality individual experience data.

The previous discussion of credibility premium components presents an understanding of what Bühlmann and Gisler [3] (p. 58) offered as their remarks on Theorem 3.10.

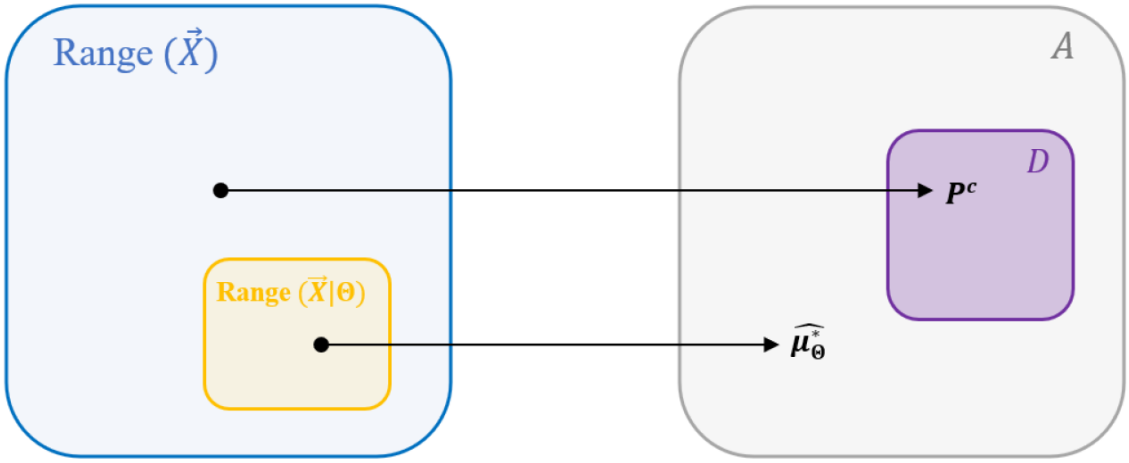


Figure 3.1: A recap of Bayes Premium $\hat{\mu}^*_\Theta$ Defined as in Theorem 3.7 and Credibility Premium P^c

From Theorem 3.10. Range $(\vec{X})$ is the set of all past observed losses and Range $(\vec{X}) \supset$ Range$(\vec{X} \mid \Theta)$ is the subset of losses produced by some risk profileΘ. The above Ven diagram shows how they map on to the elements $\hat{\mu}^*_\Theta \in$ A *and*$P^c \in$ D.

Remark 3.11. Restricting the 'search' of individual premium estimators to those that have linear observations ensures that the credibility premium is in between $\bar{X}$ andμ_0. That is, it always falls into one of two cases:

$$\bar{X} \leq \text{ credibility premium } \leq \mu_0, \text{ if } \bar{X} \leq \mu_0$$
$$\bar{X} \geq \text{ credibility premium } \geq \mu_0, \text{ if } \bar{X} \geq \mu_0$$

The equivalence principle is assumed in the general/simple credibility model. However, the same concepts can be extrapolated to other premium principles. In the following sections, the same analogy is applied to derive the credibility premium under the exponential principle. Of

course, different assumptions, appropriate to the task, are to be imposed in the next subsection.

The Credibility Premium under the Exponential Principle

All the theorems and definitions used throughout this paper now feed into the formation of the credibility premium for the exponential principle. Again, the goal is to find the credibility exponential premium $\tilde{P}^c$ which is the estimator for $\tilde{\mu}_\Theta$ (see Definition 3.5) that minimizes the expected exponential loss function (2.7). The pool of estimators $\hat{\tilde{P}}^c$ are chosen from those that have linear observations$\tilde{\mathbf{D}} = \left\{\hat{\tilde{P}}^c : \hat{\tilde{P}}^c = a_0 + \sum_{j=1}^{n} a_j X_j, a_0, a_{j(j=1,\dots,n)} \in \mathbb{R}\right\}$.

$$\min_{\tilde{P}^c} \mathbb{E}\left[\left(e^{\alpha\tilde{\mu}_\Theta} - e^{\alpha\vec{\tilde{P}}_c^c}\right)^2\right] = \mathbb{E}\left[\left(e^{\alpha\tilde{\mu}\omega} - e^{\alpha\tilde{p}_c^c}\right)^2\right] \qquad (3.21)$$

However, this attempt has been found to be unsuccessful (see the full presentation of the attempt in Appendix A.4.1) due to the complexity encountered during derivation. The following represents another attempt. It should first be noted that Bühlmann and Gisler, [3] (p.65) allowed the past individual experience data used in estimation to be transformations of observed losses, instead of explicitly using$X_j (j = 1,2,\dots,n)$. Wen et al. [12] and Wen et al. [13] proposed the following lemma and assumed the appropriate related assumptions: they suggested that the credibility estimator for the conditional moment generating function can be used to derive the credibility exponential premium.

Assumption 3.12.

- Given $\Theta = \theta$, the transformations of risks $e^{\alpha X_j}$'s are iid with distribution function $F_{X_j|\Theta}$
- Assumption 3.1
- $\gamma_\Theta = \mathbb{E}[e^{\alpha X_1} \mid \Theta]$
- $\gamma_0 = \mathbb{E}[e^{\alpha X_1}]$
- $\sigma^2_{\gamma_\Theta} = \mathrm{Var}(e^{\alpha X_1} \mid \Theta)$
- $\tau^2_\gamma = \mathrm{Cov}(\bar{Y}, \gamma_\theta) = \mathrm{Var}(\gamma_\theta)$

Lemma 3.13. Under Assumption 3.12, the credibility estimator for γ_θ using quadratic loss function (2.1) is

$$\tilde{P}_\gamma^c = Z_\gamma \bar{Y} + (1 - Z_\gamma)\gamma_0$$

Where $\bar{Y} = \frac{\sum_{j=1}^{n} e^{\alpha X_j}}{n}$ as $Y_j = e^{\alpha X_j}$ and $Z_\gamma = \frac{n}{n+\sigma^2_\gamma/\tau^2_\gamma}$ is the credibility factor.

Proof. Performing the following minimization using the quadratic loss function (2.1) to find P_γ^c the best estimator in the set of estimators $\hat{P}_\gamma^c$ that are linear in$Y_j = e^{\alpha X_j}$, that is, the set$\tilde{\mathbf{C}} = \left\{\hat{P}_\gamma^c : \hat{P}_\gamma^c = a_0 + \sum_{j=1}^{n} a_j Y_j\right\}$.
P_γ^c Is the solution for

$$\min_{\hat{P}_\gamma \in \mathbf{C}} \mathbb{E}\left[\left(\gamma_\ominus - \hat{P}_\gamma^c\right)^2\right] = \mathbb{E}\left[\left(\gamma_\Theta - P_\gamma^c\right)^2\right] \qquad (3.22)$$

And it has to be of the form

$$P_\gamma^c = \hat{a}_0 + \sum_{j=1}^{n} \hat{a}_j Y_j$$

Where $\hat{a}_0, \hat{a}_j$ solves the minimization problem

$$\min_{a_0,a_{J(j=1,\dots,n)}\in\mathbb{R}} \mathbb{E}\left[\left(\gamma_\Theta - a_0 - \sum_{j=1}^{n} a_j Y_j\right)^2\right] = \mathbb{E}\left[\left(\gamma_\Theta - \hat{a}_0 - \sum_{j=1}^{n} \hat{a}_j Y_j\right)^2\right]$$

The same argument is presented in the proof of Theorem 3.10. Since $Y_j's$ are identically distributed and through the uniqueness of the credibility estimator:

$$\hat{a}_1 = \hat{a}_2 = \cdots = \hat{a}_n$$

It should now be rewritten:

$$P_Y^c = \hat{a} + \hat{b}\bar{Y} \tag{3.23}$$

Where

$$\bar{Y} = \frac{\sum_{j=1}^{n} e^{\alpha X_j}}{n}$$

For $\hat{a}$ and $\hat{b}$ are the solutions to

$$\mathbb{E}\left[\left(\gamma_\Theta - \hat{a} - \hat{b}\bar{Y}\right)^2\right] = \min_{a,b\in\mathbb{R}} \mathbb{E}[(\gamma_\Theta - a - b\bar{Y})^2] \tag{3.24}$$

There are two variables to minimize, thus, partial differentiation is to be carried out to minimize the expected loss. Starting with a

$$\begin{aligned}
\frac{\partial}{\partial a}\mathbb{E}[(\gamma_\Theta - a - b\bar{Y})^2] &= \frac{\partial}{\partial a}\mathbb{E}[(\gamma_\Theta^2 - 2\gamma_\Theta(a + b\bar{Y}) + (a + b\bar{Y})^2)] \\
&= \frac{\partial}{\partial a}\left(\mathbb{E}\left[\gamma_\ominus^2\right] - 2\mathbb{E}[\gamma_\Theta](a + b\bar{Y}) + \mathbb{E}[a^2 + 2ab\bar{Y} + b^2\bar{Y}^2]\right) \\
&= \frac{\partial}{\partial a}(\mathbb{E}[\gamma_\Theta^2] - 2\mathbb{E}[\gamma_\Theta](a + b\bar{Y}) + a^2 + 2ab\mathbb{E}[\bar{Y}] + \mathbb{E}[b^2\bar{Y}^2]) \\
&= -2\mathbb{E}[\gamma_\Theta] + 2a + 2b\mathbb{E}[Y] \qquad (3.25)
\end{aligned}$$

After equating (3.25) to zero, we get:

$$\begin{aligned}
a &= \mathbb{E}[\gamma_\theta] - b\mathbb{E}[Y] \\
a &= \gamma_0 - b\gamma_0 \qquad (3.26) \\
\rightarrow \hat{a} &= (1-b)\gamma_0
\end{aligned}$$

Now, substituting equation (3.26) in (3.24)

$$\begin{aligned}
\mathbb{E}\left[(\gamma_\Theta - \gamma_0) - b(\bar{Y} - \gamma_0)\right)^2\right] &= \mathbb{E}[(\gamma_\Theta - \gamma_0)^2 - 2b(\gamma_\theta - \gamma_0)(\gamma_0 - \bar{Y}) + b^2(\gamma_0 - \bar{Y})^2] \\
&= \mathbb{E}[(\gamma_\Theta - \gamma_0)^2] - 2b\mathbb{E}[(\gamma_\theta - \gamma_0)(\gamma_0 - \bar{Y})] + b^2\mathbb{E}[(\gamma_0 - \bar{Y}
\end{aligned}$$

Then taking partial differentiation with respect to b and equating to zero, we get:

$$\begin{aligned}
-2\mathbb{E}[(\gamma_\Theta - \gamma_0)(\gamma_0 - \bar{Y})] + 2b\mathbb{E}[(\gamma_0 - \bar{Y})^2] &= 0 \\
\mathbb{E}[(\gamma_\Theta - \gamma_0)(\gamma_0 - \bar{Y})] &= b\mathbb{E}[(\gamma_0 - \bar{Y})^2]
\end{aligned}$$

Spotting covariance and variance forms appearing (cf. A.3), we get:

$$\mathrm{Cov}(\bar{Y}, \gamma_\Theta) = b\mathrm{Var}(\bar{Y})$$

Solving for b to get the minimiser

$$\hat{b} = \frac{\mathrm{Cov}(\bar{Y}, \gamma_\Theta)}{\mathrm{Var}(\bar{Y})}$$

Given the model Assumption 3.12 holds, and by the dependency of Y_1 and Θ

$$\mathrm{Cov}(\bar{Y}, \gamma_\Theta) = \mathrm{Var}(\gamma_\Theta) =: \tau_\gamma^2 \qquad (3.26)$$

$$\mathrm{Var}(\bar{Y}) = \frac{\mathbb{E}\left[\sigma_{\gamma_\Theta}^2\right]}{n} + \mathrm{Var}(\gamma_\Theta) =: \frac{\sigma_\gamma^2}{n} + \tau_\gamma^2 \qquad (3.27)$$

The b minimizer (3.24) is reformed as follows:

$$\hat{b} = \frac{\tau_\gamma^2}{\tau_\gamma^2 + \frac{\sigma_\gamma^2}{n}} = \frac{n}{n + \frac{\sigma_\gamma^2}{\tau_\gamma^2}} \qquad (3.27)$$

Finally, let $Z_\gamma = \hat{b}$ then plugging Z_γ in equation (3.26), it holds$\hat{a} = \left(1 - Z_\gamma\right)\gamma_0$. Now looking to equation (3.23), the credibility estimator is written:

$$P_\gamma^c = Z_\gamma \bar{Y} + \left(1 - Z_\gamma\right)\gamma_0 \qquad (3.27)$$

The form of the minimization problem in Wen et al. [12] and Wen et al. [13] used for finding the credibility exponential premium $\tilde{P}^c$ can be traced back

As being a minimization across the individual exponential premium estimators $\hat{\tilde{P}}^c$ that are ln-linear in$e^{\alpha X_j}$. The set is defined: $\widetilde{W} = \left\{\hat{\tilde{P}}^c : \hat{\tilde{P}}^c = \frac{1}{\alpha}\ln(a_0 + \sum_{j=1}^n a_j e^{\alpha X_j}, a_0, a_{j(j=1,\dots,n)} \in \mathbb{R})\right\}$. Now, it is written, using the exponential loss function (2.7):

$$\min_{\tilde{p}^c \in \overline{\mathbf{W}}} \mathbb{E}\left[\left(e^{\alpha\tilde{\mu}_\Theta} - e^{\alpha\cdot\hat{\tilde{P}}^c}\right)^2\right] = \mathbb{E}\left[\left(e^{\alpha\tilde{\mu}_\Theta} - e^{\alpha\cdot\tilde{p}c}\right)^2\right] \qquad (3.28)$$

It can be seen that the minimization problem on the left-hand side of Lemma 3.13 is equivalent to the following:

$$\min_{a_0, a_j(j=1,\dots,n)\in\mathbb{R}} \mathbb{E}\left[\left(\gamma_\theta - a_0 - \sum_{j=1}^n a_j e^{\alpha X_j}\right)^2\right]$$
$$= \min_{a_0, a_j(j=1,\dots,n)\in\mathbb{R}} \mathbb{E}\left[\left(e^{\alpha_\mu\tilde{\mu}_\Theta} - e^{\alpha\cdot\frac{1}{\alpha}\ln\left(a_0 + \sum_{j=1}^n a_j e^{\alpha X_j}\right)}\right)^2\right]$$

Then (3.28) and (3.22) are equivalent

$$\min_{\hat{P}_\sigma \in \mathbb{C}} \mathbb{E}\left[\left(\gamma_\Theta - \hat{P}_\gamma^c\right)^2\right] = \min_{\hat{P}^c \in \overline{\mathbf{W}}} \mathbb{E}\left[\left(e^{\alpha\tilde{\mu}_\Theta} - e^{\alpha\cdot\hat{P}^c}\right)^2\right]$$

The main objective is to find the credibility estimator for the individual exponential premium$\tilde{\mu}_\Theta$, as defined in Theorem 3.8, not forγ_Θ. Thus, it is written:

$$\mathbb{E}\left[\left(\gamma_\Theta - P_\gamma^c\right)^2\right] = \mathbb{E}\left[\left(e^{\alpha\tilde{\mu}_\Theta} - e^{\alpha-\tilde{p}c}\right)^2\right]$$

Notice

$$\gamma_\Theta = e^{\alpha\tilde{\mu}_\Theta} \Leftrightarrow \tilde{\mu}_\Theta = \frac{1}{\alpha}\ln\gamma_\Theta$$

Then

$$P_\gamma^c = e^{\alpha\tilde{P}c} \Leftrightarrow \tilde{P}^c = \frac{1}{\alpha}\ln P_\gamma^c$$

This concludes Wen et al.'s [12] proof of the following theorem:

Theorem 3.14. Under Assumption 3.12, and defining P_γ^c and Z_γ as in Lemma 3.13, the credibility premium under the exponential principle (cf. Section 2.1) is:

$$\tilde{P}^c = \frac{1}{\alpha}\ln P_\gamma^c = \frac{1}{\alpha}\ln\left[Z_\gamma Y + (1 - Z_\gamma)\gamma_0\right]$$

Where $\bar{Y} = \frac{\sum_{j=1}^{n} e^{\alpha X_j}}{n}$ as $Y_j = e^{\alpha X_j}$ and $Z_\gamma = \frac{n}{n+\sigma_\gamma^2/\tau_\gamma^2}$ is the credibility factor.

Example 3.15. Company XYZ sells house-fire insurance policies in Manchester. The occurrence of house fires is associated with whether or not a house has a working fire-alarm system. The actuary at XYZ believes that the random variable Θ representing the proportion of houses in Manchester that have no working fire alarm system in a given year follows a Beta distribution (see Definition A.7) with the following density function

$$f_\Theta(\theta) = \frac{1}{\mathrm{B}(2,2)}\theta^{2-1}(1-\theta)^{2-1}, 0 \le \theta \le 1$$

The number of house fires occurring during year j in Manchester $N_j \mid \Theta = \theta (j = 1,2, \dots)$ follows a Binomial distribution (see Definition A.6) with parameters $m = 1$ and$p = \theta$. A total number of eight covered houses burned in the last 12 years. XYZ's actuary wants to know the expected number of house fires in the coming year, that is, the 13th year.

Model characteristics and notations

- Let N_j be the identically-distributed observed number of house fires in year$j$.
- Let $\hat{\mu}_\Theta = \mathbb{E}[\mu_\Theta \mid N_1, \dots, N_{12}]$ be the Bayes equivalence premium as in Theorem 3.7, where μ_Θ is the individual equivalence premium as in Definition 3.4. Set $\mu_0 = \mathbb{E}\left[N_j\right]$ as in equation (3.1). Also, let$\tau^2 = \mathrm{Var}(\mu_\Theta)$.
- Let $\hat{\tilde{\mu}}_\Theta = \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha\tilde{\mu}_\Theta} \mid N_1, \dots, N_{12}]$ be the Bayes exponential premium as in Theorem 3.8 where $\tilde{\mu}_\Theta$ is the individual exponential premium as in Proposition 3.5. Set $\tilde{\mu}_0 = \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha N_j}]$ as in equation (3.2). Also, let $\gamma_\Theta = \mathbb{E}[e^{\alpha N_j} \mid \Theta]$ and$\tau_\gamma^2 = \mathrm{Var}(\gamma_\Theta)$, then define $\gamma_0 = \mathbb{E}[e^{\alpha N_j}]$ as in Assumption 3.12.
- Beta and Binomial are conjugate priors (see Proposition A.8) as listed in Heilmann [7] or in Buhlmann and Gisler, [3] (p.33). This means that for n_j the number of observed claims in year j and m is the parameter of $N_j \mid \Theta \sim \mathrm{Binomial}(m, \Theta)$, then the following random variables are distributed as: a priori $\Theta \sim \mathrm{Beta}(a, b)$ a posteriori

$$\Theta \mid N_j \sim \mathrm{Beta}\left(a' = a + n_j, b' = b + m - n_j\right)$$

$$\mathbb{E}[\Theta] = \frac{a}{a+b} = \frac{1}{2}, \text{ for, } a = b = 2$$

$$\mathrm{Var}(\Theta) = \frac{ab}{(1+a+b)(a+b)^2} = \frac{4}{(5)(16)} = 0.05$$

Collective perspective: Let us find overall 'averages' for the number of house fires in Manchester for the 13th year, regardless of the condition of fire alarms in the city's houses, under each of the following principles:

- Equivalence principle (3.1)

$$\mu_0 = \mathbb{E}[N_{13}] = \mathbb{E}\left[\mathbb{E}[N_{13} \mid \Theta]\right] = \mathbb{E}[1 \cdot \Theta] = 1 \cdot 0.5 = 0.5$$

- Exponential principle (3.2)

$$\tilde{\mu}_0 = \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha N_{13}}] = \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha \mu_\Theta}] \text{ by iterative 7.Property , } \alpha > 0$$

After viewing (Definitions A. 3 and A.6), take

$$\begin{aligned}\tilde{\mu}_\Theta &= \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha N_{13}} \mid \Theta] \\ &= \frac{1}{\alpha}\ln m_{N_{13}|\Theta}(\alpha), m_{N_{13}|\Theta} \text{ is the mgf of } N_{13} \mid \Theta \\ &= \frac{1}{\alpha}\ln(1-\Theta+\Theta e^{\alpha}), \alpha > 0\end{aligned}$$

Now

$$\begin{aligned}\tilde{\mu}_0 &= \frac{1}{\alpha}\ln \mathbb{E}\left[e^{\alpha\cdot\frac{1}{\alpha}\ln(1-\Theta+\Theta e^{\alpha})}\right] \\ &= \frac{1}{\alpha}\ln \mathbb{E}[(1-\Theta+\Theta e^{\alpha})] \\ &= \frac{1}{\alpha}\ln\big((1-\mathbb{E}[\Theta]+e^{\alpha}\mathbb{E}[\Theta])\big) \\ &= \frac{1}{\alpha}\ln\big(0.5(1+e^{\alpha})\big), \alpha > 0\end{aligned}$$

Bayesian perspective: To recap from Probability Theory, the sum of Binomial random variables, with the same probability p of success, is a Binomial random variable with the 'number of experiments' parameter being the sum of each Binomial's number of experiments. Since there are 12 years and the claims for each year $N_j \mid \Theta \sim \text{Binomial}(1,\Theta)$, then claims over the sum of 12 years are: $(N_1+\cdots+N_{12})|\Theta = S|\Theta \sim \text{Binomial}(12,\Theta)$. A priori $\Theta \sim \text{Beta}(2,2)$ a posteriori $\Theta \mid S \sim \text{Beta}(a'=2+8, b'=2+12-8)$ then

$$\Theta|N_1,\dots,N_{12} \sim \Theta|S \sim \text{Beta}(a'=10, b'=6)$$

$$\mathbb{E}[\Theta \mid S] = \frac{a'}{a'+b'} = \frac{10}{10+6} = \frac{5}{8}$$

- Bayes estimator for the expected number of house fires under the Equivalence principle (see Theorem 3.7)

$$\hat{\mu}^*_{\ominus} = \mathbb{E}[\mu_\Theta \mid N_1,\dots,N_{12}] = \mathbb{E}[1\cdot\Theta \mid N_1,\dots,N_{12}] = \mathbb{E}[\Theta \mid S] = \frac{5}{8}$$

- Bayes estimator for the expected number of house fires under the exponential principle (see Theorem 3.8)

$$\begin{aligned}\hat{\tilde{\mu}}^*_\Theta &= \frac{1}{\alpha}\ln \mathbb{E}[e^{\alpha\tilde{\mu}_\Theta} \mid N_1,\dots,N_{12}], \alpha > 0 \\ &= \frac{1}{\alpha}\ln \mathbb{E}\left[e^{\alpha\cdot\frac{1}{\alpha}\ln\left(1-\Theta+\Theta_c^{\alpha}\right)} \,\middle|\, N_1,\dots,N_{12}\right] \\ &= \frac{1}{\alpha}\ln \mathbb{E}[(1-\Theta+\Theta e^{\alpha}) \mid N_1,\dots,N_{12}]\end{aligned}$$

Then

$$\hat{\hat{\mu}}_{\ominus} = \frac{1}{\alpha}\ln \mathbb{E}[(1-\Theta+\Theta e^{\alpha}) \mid S], \alpha > 0$$
$$= \frac{1}{\alpha}\ln\left(\frac{3}{8}+\frac{5}{8}e^{\alpha}\right)$$

Credibility perspective: let the average observed claims be

$$\bar{N} = \frac{\sum_{j=1}^{12} N_j}{12} = \frac{8}{12} = \frac{2}{3}$$

Further find

$$\sigma^2 = \mathbb{E}[\mathrm{Var}(N_{13} \mid \Theta)]$$
$$= \mathbb{E}[1 \cdot \Theta(1-\Theta)]$$
$$= \mathbb{E}[\Theta] - \mathbb{E}[\Theta^2]$$
$$= 0.5 - (0.05 + 0.5^2) = 0.2$$

- Credibility estimator for expected number of house fires under the equivalence principle P^c (see Theorem 3.10)

Let Z be the credibility factor for the equivalence principle model

$$Z = \frac{8}{8+\frac{\sigma^2}{\tau^2}}$$

Where

$$\tau^2 = \mathrm{Var}(\mathbb{E}[N_{13} \mid \Theta]) = \mathrm{Var}(1 \cdot \Theta) = 0.05$$

Then

$$Z = \frac{8}{8+\frac{1}{6}} = 0.978$$
$$P^c = ZN + (1-Z)\mu_0$$
$$= 0.978 \cdot \frac{2}{3} + (1-0.978) \cdot 0.5$$
$$= 0.663$$

- Credibility estimator for the expected number of house fires under the exponential principle $\tilde{P}^{c_S}$ (see Theorem 3.14 and Lemma 3.13)

Let Z_γ be the credibility factor under the exponential principle model Find

$$\gamma_\Theta = \mathbb{E}[e^{\alpha N_j} \mid \Theta] = (1-\Theta+\Theta e^{\alpha})$$

And

$$\tau_\gamma^2 = \mathrm{Var}(\gamma_\Theta) = (e^{\alpha}-1)^2\mathrm{Var}(\Theta)\big) = (e^{\alpha}-1)^2 \cdot 0.05$$

$$\begin{aligned}
\sigma_\gamma^2 &= \mathbb{E}[\mathrm{Var}(\gamma_\Theta)] \\
&= \mathbb{E}[\mathbb{E}[e^{2\alpha N_j} \mid \Theta] - (\mathbb{E}[e^{\alpha N_j} \mid \Theta])^2] \\
&= \mathbb{E}[(1-\Theta+\Theta e^{2\alpha}) - (1-\Theta+\Theta e^{\alpha})^2] \\
&= \mathbb{E}[(1-\Theta+\Theta e^{2\alpha}) - (1-2\Theta+\Theta^2+2\Theta e^{\alpha}-2\Theta^2 e^{\alpha}+\Theta^2 e^{2\alpha})] \\
&= \mathbb{E}[1-\Theta+\Theta e^{2\alpha}-1+2\Theta-\Theta^2-2\Theta e^{\alpha}+2\Theta^2 e^{\alpha}-\Theta^2 e^{2\alpha}] \\
&= \mathbb{E}[\Theta+\Theta e^{2\alpha}-\Theta^2 e^{2\alpha}-2\Theta e^{\alpha}+2\Theta^2 e^{\alpha}-\Theta^2] \\
&= \mathbb{E}[\Theta]+\mathbb{E}[\Theta e^{2\alpha}]-\mathbb{E}[\Theta^2 e^{2\alpha}]-2\mathbb{E}[\Theta e^{\alpha}]+2\mathbb{E}[\Theta^2 e^{\alpha}]-\mathbb{E}[\Theta^2] \\
&= 0.5+0.5e^{2\alpha}-0.3e^{2\alpha}-2(0.5e^{\alpha})+2(0.3e^{\alpha})-0.3 \\
&= 0.5+(0.5-0.3)e^{2\alpha}-e^{\alpha}+0.6e^{\alpha}-0.3 \\
&= 0.5-0.3+0.2e^{2\alpha}-0.4e^{\alpha} \\
&= 0.2+0.2e^{2\alpha}-0.4e^{\alpha} \qquad (3.29)
\end{aligned}$$

Then

$$\gamma_0 = \mathbb{E}[e^{\alpha N_j}] = \mathbb{E}\Big[\mathbb{E}[e^{\alpha N_j} \mid \Theta]\Big] = \mathbb{E}[(1-\Theta+\Theta e^{\alpha})] = 0.5+0.5e^{\alpha}$$

Consequently

$$Z_\gamma = \frac{\tau_\gamma^2}{\tau_\gamma^2+\frac{\sigma_\gamma^2}{n}} = \frac{(e^{\alpha}-1)^2\cdot 0.05}{(e^{\alpha}-1)^2\cdot 0.05+0.025+0.025e^{2\alpha}-0.05e^{\alpha}}$$

$$\begin{aligned}
P_\gamma^c &= Z_\gamma N + (1-Z_\gamma)\gamma_0 \\
&= Z_\gamma \frac{2}{3} + (1-Z_\gamma)(0.5+0.5e^{\alpha})
\end{aligned}$$

Finally:

$$\tilde{P}^c = \frac{1}{\alpha}\ln\left(Z_\gamma \frac{2}{3} + 0.5(1-Z_\gamma)(1+e^{\alpha})\right)$$

For$\alpha > 0$, the exponential premium principle always seems to produce an outcome that is larger than that produced by the equivalence principle. This aligns with the Safety loading Property 4. Discussed earlier.

CONCLUSION AND RECOMMENDATIONS

Conclusion

This paper has gradually developed tools for identifying connections between the exponential premium and credibility theory, borrowing methods from other theories, such as risk theory and decision theory. More detailed proofs from a range of relevant resources have also been identified and presented. The thinking behind the structure of the paper was to start Chapter 1 by introducing a clearer picture of data sorting and demonstrating that a lack of harmony among data is the reason for considering experience rating. The historical background provided, which has been summarized from Goulet [5], covered the rise of experience rating procedures in insurance practice, with credibility theory giving one such procedure. All examples and graphs in this paper represent the author's efforts to try to convey ideas drawn from a review of the related literature, as understood by the author.

Due to the random nature of individual premiums, discussed in Chapter 3, Bayesian statistics were applied in order to estimate the Bayes premium under the equivalence principle, adding more intermediate steps in the proof of Theorem 3.7, as well as in the proof of the exponential

principle in Theorem 3.8. A broader point of view on rate-making, represented by collective premiums (3.1) and (3.2) was also introduced. When introducing both collective and individual premiums, the drawbacks of each were intentionally noted, to render the idea of combining both in pricing more intuitively sound. More intermediate steps and mathematical clarifications have also been added to the proof of the credibility premium under the equivalence principle in Theorem 3.10, compared with the same proof as it appears in Bühlmann and Gisler [3], (p.56).

Recommendations

An attempt was made to derive an exponential credibility estimator that is linear in observations. However, this led to a differentiation task requiring the interchange of differentiation and expectation operators (see Appendix A.4.1), a gap that was not addressed before in the related papers. It is possible, even with its complexity, for an interested researcher to fill this gap and continue the derivation in (Appendix A.4.1), using partial differentiation.

Acknowledgments

I would like to thank Deemah for all her support and love throughout the progress of this work. And my sister Danah for all the coffee cups she made me.

REFERENCES

[1] Asmussen, S., & Constantinescu, C. (2021). On the risk of credibility premium rules. Scandinavian Actuarial Journal. Taylor & Francis.

[2] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999). Coherent measures of risk. Mathematical Finance 9, 203-228.

[3] Bühlmann, H. and Gisler, A. (2005). A course in credibility theory and its applications. Springer Science & Business Media.

[4] Gómez-Déniz, E., & Vázquez-Polo, F. J. (2022). Exact credibility reference Bayesian premiums. Insurance: Mathematics and Economics. Elsevier.

[5] Goulet, Vincent (1998). Principles and application of credibility theory. Journal of Actuarial Practice 6, 5-62.

[6] Gerber, H. and Jones, D. (1976). Some practical considerations in connection with the calculation of stop-loss premiums. Transactions of the Society of Actuaries 28, 215-231.

[7] Heilmann, W. R. (1989). Decision theoretic foundations of credibility theory. Insurance: Mathematics and Economics 8, 77-95.

[8] Dhaene, J. and Goovaerts, M. & Kaas, R. (2003). Economic capital allocation derived from risk measures. North American Actuarial Journal 7: 2, 44-56.

[9] Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2008). Modern actuarial risk theory using R. Springer Science and Business Media.

[10] Schmidtli, H., & Schmidtli, H. (2017). Credibility theory. Risk Theory. Springer.

[11] Tsanakas, A. and Desli, E. (2003). Risk measures and theories of choice. British Actuarial Journal 9, 959-991.

[12] Wen, L., Yu, J., Mei, G., & Zhang, Y. (2017). The credibility premiums based on estimated moment-generating function. Communications in Statistics-Theory. Taylor & Francis.

[13] Wen, L., Wang, W., & Wang, J. (2011). The credibility premiums for exponential principle. Acta Mathematica Sinica, English Series 27, 2217–2228.

[14] Yong, Y., Zeng, P., & Zhang, Y. (2024). Credibility theory for variance premium principle. North American Actuarial Journal. Taylor & Francis.

Appendix A

In what follows, we consider a Probability space(Ω, F, P).

A. 1 Probabilities and distributions

Definition A.1. In what follows, we consider a Probability space

$$F_X(x) = \mathbb{P}(X \leq x) \text{ for } x \in \mathbb{R} \qquad (1.1)$$

Definition A.2. For any continuous random variable X that has a cdfF_X, and for the integrable function$f_X: \mathbb{R} \to [0, \infty)$, the probability density function (pdf) of X can be expressed as:

$$f_X(x) = F_X'(x), \forall x \in \mathbb{R} \qquad (1.2)$$

And the following relationship also holds:

$$F_X(x) = \int_{-\infty}^{x} F_X'(t)dt = \int_{-\infty}^{x} f_X(t)dt, \forall x \in \mathbb{R}$$

Since $\lim_{x \to +\infty} F_X(x) = 1$ it holds that:

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \qquad (1.3)$$

Definition A.3. (Moment generating function) For any random variable X its

Moment generating function $m_X: \mathbb{R} \to (0, \infty]$ is defined as:

$$m_X(t) = \mathbb{E}[e^t X] \text{ for t} \in \mathbb{R}$$

Definition A.4. For any random variables X, Y with the joint distribution function$f_{X,Y}$, and $f_X, f_Y > 0$ are the marginal distribution functions of X and Y respectively. The conditional distribution function of X given $Y = y$ is

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Theorem A.5. (Bayes Theorem) For any random variables X, Y with the joint distribution function$f_{X,Y}$, and $f_X > 0, f_Y > 0$ are the marginal distribution functions of X and Y respectively. The conditional distribution function of Y given $X = x$ is$f_{Y|X=x}$, then

$$f_{Y|X}(y \mid x) = \frac{f_{X|Y}(x \mid y) f_Y(y)}{f_X(x)}$$

Definition A.6. (Binomial distribution) If a discrete random variable N has a Binomial distribution with parameters $m \in \mathbb{N}$ and$p \in [0,1]$, where m represents the number of independent Bernoulli experiments and p is the probability of a successful experiment occurring, then its probability mass function f_N is

$$f_N(k) = \binom{m}{k} p^k (1-p)^{m-k}, k = 0,1 \ldots m$$

And its moment-generating function m_N is expressed as

$$
\begin{aligned}
m_N(t) &= \sum_{k=0}^{m} e^{tk} \cdot f_N(k) \\
&= \sum_{k=0}^{m} e^{tk} \cdot \binom{m}{k} p^k (1-p)^{m-k} \\
&= \sum_{k=0}^{m} \binom{m}{k} (e^t p)^k (1-p)^{m-k}, \text{ that is a binomial series} \\
&= [1 - p + pe^t]^m
\end{aligned}
$$

Furthermore, its mean and variance are $\mathbb{E}[N] = mp, \mathrm{Var}(N) = mp(1-p)$

Definition A.7. (Beta distribution) If a continuous random variable X has a Beta distribution with parameters $a > 0$ and$b > 0$. Then its probability density function f_X is

$$f_X(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, 0 \le x \le 1$$

For

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Where the Gamma function$\Gamma(.) is defined as$

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt, x > 0$$

And its mean and variance are

$$\mathbb{E}[X] = \frac{a}{a+b}, \mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Proposition A.8. (Beta-Binomial model) If the prior distribution $f_\Theta(\theta)$ follows a Beta distribution with parameters $a > 0$ and $b > 0$ and the likelihood function's $f_{X|\Theta}(k)$ distribution is a Binomial distribution with parameters m andθ, then the posterior distribution $f_{\Theta|X}(\theta \mid k)$ is Beta with parameters $a' = a + k$ and $b' = b + m - k$.

Proof. To begin with, the prior of Θ is

$$f_\Theta(\theta) = \frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1}, 0 \le \theta \le 1$$

Then

$$f_{X|\Theta}(k \mid \theta) = \binom{m}{k} \theta^k (1-\theta)^{m-k}, k = 0,1 \dots m$$

Now, using Bayes Theorem A. 5

$$
\begin{aligned}
f_{\Theta|X}(\theta \mid k) &= \frac{f_{X|\Theta}(k \mid \theta) \cdot f_\Theta(\theta)}{f_X(k)} \\
&= \binom{m}{k} \theta^k (1-\theta)^{m-k} \cdot \frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1} \frac{1}{f_X(k)}
\end{aligned}
$$

Gather constants in one notation c

$$f_{\Theta|X}(\theta \mid k) = c \cdot \theta^{a+k-1}(1-\theta)^{b+m-k-1}, 0 \le \theta \le 1$$

It is possible to recognize the form of $f_{\Theta|X}(\theta \mid k)$ as a form of a Beta distribution with parameters $a+k$ and$b+m-k$.

A. 2 Expectation and its properties

Definition A.9. (Expectation)

(i) For any continuous random variable X with probability density functionf_X, the expectation of X is

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)dx$$

(ii) For any discrete random variable X with probability mass function f_X the expectation of X is

$$\mathbb{E}[X] = \sum_{x} x \cdot f_X(x)$$

Definition A.10. (Conditional Expectation)

(i) For any continuous random variables X, Y with conditional probability density function$f_{X|Y}$, the conditional expectation of X given $Y = y$ is

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y=y}(x \mid y)dx$$

(ii) For any discrete random variables X, Y with conditional probability mass function $f_{X|Y}$ the conditional expectation of X given $Y = y$ is

$$\mathbb{E}[X \mid Y = y] = \sum_{x} x \cdot f_{X|Y=y}(x \mid y)$$

Proposition A.11.

(i) For any X and Y two random variables the following linearity property holds:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

For any two functions $g: \mathbb{R} \to \mathbb{R}$ and $h: \mathbb{R} \to \mathbb{R}$ it also holds

$$\mathbb{E}[g(X) + h(Y)] = \mathbb{E}[g(X)] + \mathbb{E}[h(Y)]$$

(ii) For independent random variables X and$Y$, we have

$-\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ Simalarly, for any two functions $g: \mathbb{R} \to \mathbb{R}$ and $h: \mathbb{R} \to \mathbb{R}$ it also holds$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$
$-\mathbb{E}[X \mid Y] = \mathbb{E}[X]$. The same notion applies for any function $g: \mathbb{R} \to \mathbb{R}$ it holds $\mathbb{E}[g(X) \mid Y] = \mathbb{E}[g(X)]$

The property of linearity is used in this paper as a conditional expectation operator and it also holds.
(iii) For any X, Y and Z random variables the following linearity property holds

$$\mathbb{E}[X + Y \mid Z] = \mathbb{E}[X \mid Z] + \mathbb{E}[Y \mid Z]$$

Proposition A.12. (Tower property) For X, Y two random variables such that $\mathbb{E}[X] < \infty$ and $\mathbb{E}[X \mid Y]: \Omega \to \mathbb{R}$ is a random variable and its randomness is inherited from$Y$, that is, $\mathbb{E}[X \mid Y]$ is a function of$Y$, the following holds:

(i) For any random variable X and a continuous random variable Y with probability density function f_Y the expectation of $X \mid Y$ is

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y = y] \cdot f_Y(y)dy = \mathbb{E}[X]$$

(ii) For any random variable$X$, and a discrete random variable Y with probability mass functions f_Y the expectation of $X \mid Y$ is

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \sum_{y} \mathbb{E}[X \mid Y = y] \cdot f_Y(y) = \mathbb{E}[X]$$

A. 3 Variance/Covariance and their properties

Definition A.13. (i) For any continuous random variables X, Y with probability density functions f_X, f_Y respectively

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 \cdot f_X(x)dx = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}^2[X]$$

$$\mathrm{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) \cdot f_{X,Y}(x, y)dxdy = \mathbb{E}[(Y - \mathbb{E}[Y])(X - \mathbb{E}[X])]$$

(ii) For any discrete random variables X, Y with probability mass functionsf_X, f_Y respectively

$$\mathrm{Var}(X) = \sum_{x} (x - \mathbb{E}[X])^2 \cdot f_X(x) = \mathbb{E}[(x - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}^2[X]$$

$$\mathrm{Cov}(X, Y) = \sum_{x,y} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) \cdot f_{X,Y}(x, y) = \mathbb{E}[(Y - \mathbb{E}[Y])(X - \mathbb{E}[X])]$$

Proposition A.14.

(i) For any X and Y two random variables the following variance property holds

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X, Y)$$

(ii) For independent random variables X and$Y$, we have $\mathrm{Cov}(X, Y) = 0$ then,

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

(iii) For any random variable X and a constant$c \in \mathbb{R}$, we have

$$-\mathrm{Var}[X + c] = \mathrm{Var}[X]$$
$$-\mathrm{Var}[cX] = c^2\mathrm{Var}[X]$$

Proposition A.15. (Law of Total Variation) For any random variables X, Y and $X \mid Y: \Omega \to \mathbb{R}$ the following holds

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y])$$

A. 4 Miscellaneous topics

Definition A.16. A function $f: \mathbb{R} \to \mathbb{R}$ is said to be linear if for all constants $c \in \mathbb{R}$

$$f(cx) = cf(x) \forall x \in X$$

Lemma A.17. (Jensen's inequality) Let X be a random variable and $g: \mathbb{R} \to \mathbb{R}$ is a convex function, then

$$g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)]$$

A.4.1 Proof attempt: Credibility exponential premium

The goal is to find a credibility exponential premium $\tilde{P}^c$ which is the estimator of $\tilde{\mu}_\Theta$ (see Definition 3.5) that minimizes the expected exponential loss function (2.7). But this time as

American Journal of Statistics and Actuarial Science
ISSN 2958-5244(Online)
Vol.6, Issue 1, pp 1–34, 2025
www.ajpojournals.org

opposed to what is done in Theorem 3.10, the estimator is linear in $e^{\alpha X_j} = Y_j (j = 1, \dots, n)$ identically distributed transformations of original observed claims, and it is selected from the set $\widetilde{\mathrm{D}} = \left\{\hat{\tilde{P}}^c : \hat{\tilde{P}}^c = a_0 + \sum_{j=1}^{n} a_j X_j, a_0, a_{j(j=1,\dots,n)} \in \mathbb{R}\right\}$.

The minimization problem is expressed as follows:

$$\min_{\hat{\tilde{p}}^c \in \mathbb{D}} \mathbb{E}\left[\left(e^{\alpha \tilde{p}_\theta} - e^{\alpha \hat{P}^c}\right)^2\right] = \mathbb{E}\left[\left(e^{\alpha \tilde{\mu} \theta} - e^{\alpha \tilde{P}^c}\right)^2\right]$$

Or equivalently

$$\min_{a_0, a_j (j=1,\dots,n) \in \mathbb{R}} \mathbb{E}\left[\left(e^{\alpha_{\tilde{\mu}}^{\theta}} - e^{\alpha\left(a_0 + \sum_{j=1}^{n} a_j X_j\right)}\right)^2\right] = \mathbb{E}\left[\left(e^{\alpha_{\tilde{\mu}}} - e^{\alpha\left(\tilde{a}_0 + \sum_{j=1}^{n} \hat{a}_j X_j\right)}\right)^2\right]$$

The same argument is presented in the proof of Theorem 3.10, since $X_j's$ are identically distributed and by the uniqueness of the credibility exponential premium. Then

$$\hat{a}_1 = \hat{a}_2 = \cdots = \hat{a}_n$$

Following this, it possible to write the credibility exponential premium as

$$\tilde{P}^c = \hat{a} + \hat{b}\bar{X},$$
$$\text{where } X = \frac{\sum_{j=1}^{n} X_j}{n}$$

$\hat{a}$ And $\hat{b}$ are the solutions to

$$\min_{a,b \in \mathbb{R}} \mathbb{E}\left[\left(e^{\alpha \tilde{\mu}_\Theta} - e^{\alpha(a + b\bar{X})}\right)^2\right] = \mathbb{E}\left[\left(e^{\alpha \tilde{\mu}_\Theta} - e^{\alpha(\hat{a} + \hat{a}X)}\right)^2\right]$$

Now to recap the form of $\tilde{\mu}_\Theta$ from Theorem 3.5

$$e^{\alpha \tilde{\mu} \Theta} = e^{\alpha - \frac{1}{\alpha} \ln \mathbb{E}\left[e^{\alpha X_j} | \Theta\right]} = \mathbb{E}[e^{\alpha X_j} \mid \Theta],$$

And rewritten this is

$$\min_{a,b \in \mathbb{R}} \mathbb{E}\left[\left(\mathbb{E}[e^{\alpha X_j} \mid \Theta] - e^{\alpha(a + b\bar{X})}\right)^2\right] = \mathbb{E}\left[\left(\mathbb{E}[e^{\alpha X_j} \mid \Theta] - e^{\alpha(\hat{a} + \hat{a}\bar{X})}\right)^2\right]$$

In simplified form

$$\begin{aligned}
\mathbb{E}\left[\left(\mathbb{E}[e^{\alpha X_j} \mid \Theta] - e^{\alpha(a + b\bar{X})}\right)^2\right] &= \mathbb{E}\left[\mathbb{E}[e^{\alpha X_j} \mid \Theta]^2 - 2\mathbb{E}[e^{\alpha X_j} \mid \Theta] e^{\alpha(a + b\bar{X})} + e^{2\alpha(a + b\bar{X})}\right] \\
&= \mathbb{E}[\mathbb{E}[e^{\alpha X_j} \mid \Theta]^2] - 2\mathbb{E}\left[\mathbb{E}[e^{\alpha X_j} \mid \Theta]\right] \mathbb{E}\left[e^{\alpha(a + b\bar{X})}\right] + \mathbb{E}\left[e^{2\alpha(a + b\bar{X})}\right] \\
&= \mathbb{E}[\mathbb{E}[e^{\alpha X_j} \mid \Theta]^2] - 2e^{\alpha a} \mathbb{E}[e^{\alpha X_j}] \mathbb{E}\left[e^{\alpha b\bar{X}}\right] + e^{2\alpha a} \mathbb{E}\left[e^{2\alpha b\bar{X}}\right]
\end{aligned}$$

There are two variables of interest, so, partial differentiation and equating to zero is to be carried out to minimize the expected loss. Starting with a we get

$$\begin{aligned}
-2\alpha e^{\alpha a} \mathbb{E}[e^{\alpha X_j}] \mathbb{E}\left[e^{\alpha b\bar{X}}\right] + 2\alpha e^{2\alpha a} \mathbb{E}\left[e^{2\alpha b\bar{X}}\right] &= 0 \\
\rightarrow \mathbb{E}[e^{\alpha X_j}] &= e^{\alpha a} \mathbb{E}\left[e^{\alpha b\bar{X}}\right] \\
\hat{a} &= \frac{1}{\alpha} \ln \frac{\mathbb{E}[e^{\alpha X_j}]}{\mathbb{E}\left[e^{\alpha b\bar{X}}\right]}
\end{aligned}$$

Then, the following is substituted (1.8) in (1.5)

It is difficult to take the partial derivative of the resulting form (1.9) with respect to b without interchanging expectation and differentiation operators.

License