Three fundamental problems in risk modelling on big data: an information theory view

Jiamin Yu

Received: date / Accepted: date

Abstract Since Claude Shannon founded Information Theory, information theory has widely fostered other scientific fields, such as statistics, artificial intelligence, biology, behavioural science, neuroscience, economics, and finance. Unfortunately, actuarial science has hardly benefited from information theory. So far, only one actuarial paper on information theory can be searched by academic search engines. Undoubtedly, information and risk, both as Uncertainty, are constrained by entropy’s law. Today’s insurance big data era means more data and more information. It is unacceptable for risk management and actuarial science to ignore information theory. Therefore, this paper aims to exploit information theory to discover performance limits of insurance big data systems and seek guidance for risk modelling and the development of actuarial pricing systems.

Keywords risk · information theory · information entropy · big data · actuarial science

1 Introduction

Since Claude Shannon published his most important paper “A mathematical theory of communication” in 1948, Information Theory has been established and widely fostered many scientific fields, such as statistics, artificial intelligence, biology, behavioral science, neuroscience, economics, and finance. Even in finance, which is most similar to actuarial science, for example, there is a lot of studies applying information theory, such as the financial value of information [1], credit ratings [2], and the impact of social media on financial markets [12]. In contrast, there is only one paper on information theory in actuarial science so far. Reference [10] discussed the residual entropy and past entropy in actuarial science and survival models for the deductible and limit of insurance claims. Unfortunately, from the search results of academic search engines, it can be said that actuarial science has hardly benefited from information theory.

First of all, theoretical research of risk theory and actuarial science requires information theory. By definition, both risk (in risk theory) and information (in information theory) are uncertainties theoretically. In the universe, everything is energy, and information is also energy. All spontaneous energy evolution processes, including risk processes, obey the second law of thermodynamics, also known as the Principle of Entropy Increase. Undoubtedly, information and risk, both as uncertainties, are constrained by information entropy’s law. Obviously, information theory also applies to risk theory and actuarial science.

Secondly, the practical application of risk management in the insurance industry (such as self-driving car insurance, cyber insurance etc.) also requires information theory. Today’s insurance industry has entered an era of insurance big data. This means that more huge amounts of data (i.e. information) can be used for actuarial modeling and risk management. Faced with massive information of big data, it is unacceptable for risk management and actuarial science to continue to ignore information theory.

Therefore, in order to meet the above theoretical and practical requirements, this paper aims to exploit information theory to discover the performance limits of
insurance big data systems, and seek guidance for risk modeling and the development of actuarial pricing systems.

The rest of this paper is organized as follows. Section 2 presents a novel communication system framework of risk model, which promotes the actuarial model to achieve both more risk classification and less pricing error. Section 3 ascribes risk classifications to a refinement problem of risk decoding, which can be solved by refinement of risk output event. Section 4 shows that the mutual information between risk input and output determines the performance of the actuarial model, and deterministic or causal model is the optimal actuarial model. Section 5 proposes that risk big data might cause risk information distortion, and lossless data collection design can significantly improve the performance of the actuarial model. Conclusions section summarizes main conclusions of this research and discusses future research outlook.

2 Communication system framework of risk model

2.1 Traditional risk model

The traditional actuarial risk model is the currently popular experience rating model, which predicts future losses (net risk premium) based on the historical loss (claim) experience of the insured subject. The experience rating model can be abstracted into the schematic diagram as below.

2.2 Communication system framework of risk model

2.2.1 Time-varying risk assumption

In the real world, risk changes over time generally. Assuming that the risk in Fig. 1 is time variant, the risk model becomes as follows.

![Fig. 2 The time-varying risk model schematic.](image)

Where, the original \( \lambda \) in Fig. 1 is replaced to \( \Lambda \), which represents the time-varying risk level, and \( \hat{\Lambda} \) is the corresponding estimated risk level. Mathematically, \( \Lambda \) can be regarded as a function \( \Lambda(t) \) of time \( t \), or even a stochastic process \( \Lambda_t \).

2.2.2 Risk informationization

According to the Merriam-Webster Dictionary, risk is possibility of loss or injury. Following the definition of information measurement in Shannon Information Theory [8], the information measurement of risk is the information (or uncertainty) measurement of loss or injury. Consequently, the actuarial problem in Fig. 2 becomes the information problem of the communication system in Fig. 3 as below.

As illustrated in Fig. 3, the original actuarial estimation process (from \( \Lambda \) to \( \hat{\Lambda} \)) in Fig. 2 is transformed into the risk information transmission process (from \( \Lambda \) to \( \hat{\Lambda} \)) in Fig. 3. Where, \( \Lambda \) originates from real-world risk source; With the risk encoder, the unknown risk \( \Lambda \) is collected as the risk input \( X \) by big data acquisition; \( X \) generates risk event \( Y \) through the risk mapper; Relying on the risk decoder, the event \( Y \) can be decoded into risk estimate \( \hat{\Lambda} \) finally. Further, this risk system is decomposed into three simple input-output subsystems, namely risk encoder, risk mapper, and risk decoder.

2.3 Equivalence of actuarial problems: information problems

In the context of insurance big data, the actuarial model adds two new requirements. One is more risk classifica-
tion, the other is higher pricing accuracy. If the set of various risk levels is represented as set $C$, the first requirement of risk classification can be described as the following mathematical problem
\[
\max \{ |C| \}
\]
Where, $|C|$ is the number of risk levels in the risk classification set $C$. Element $c$ represents a certain risk level in set $C$. The probability of pricing errors with risk level $c$, can be defined as follows
\[
P_e^c = P(\hat{A}_c \neq A_c), \quad c \in C
\]
(1)
Where, $A_c$ and $\hat{A}_c$ are the real and estimated risk of $c$ respectively. Similarly, the second requirement for precise pricing can be described as the following mathematical problem
\[
\min_{c \in C} (P_e^c)
\]
However, it seems that two mathematical problems $\max \{|C|\}$ and $\min_{c \in C} (P_e^c)$ are paradoxical: In case of more risk classification, it means $A_c$ contains more uncertainty; In case of more risk classification precise pricing, it implies $\hat{A}_c$ has less uncertainty. This paradox is consistent with the aggregation operation in experience rating: aggregation can reduce pricing errors but decay risk classification meanwhile.

Once the actuarial risk model is modeled as a risk communication system, the traditional actuarial problem can be transformed into one problem of information transmission from real-world risk $\Lambda$ to estimated risk $\hat{\Lambda}$. Thus, according to the mutual information theorem of information theory [5], the optimal risk modeling criterion is directly inferred to maximize the mutual information between real and estimated risk. According to the idea of information theory, the above two mathematical problems can be transformed into the following information expressions
\[
\max_{\hat{\Lambda}} (H(\hat{\Lambda})) \quad \text{and} \quad \max_{\Lambda} (I(\Lambda, \hat{\Lambda}))
\]
Where, $H(\hat{\Lambda})$ is the information entropy of $\hat{\Lambda}$, and $I(\Lambda, \hat{\Lambda})$ is the mutual information between $\Lambda$ and $\hat{\Lambda}$. $\max_{\hat{\Lambda}} (H(\hat{\Lambda}))$ is equivalent to $\max \{|C|\}$ (more uncertainty means greater entropy information entropy), and $\max_{\Lambda} (I(\Lambda, \hat{\Lambda}))$ is equivalent to $\min_{c \in C} (P_e^c)$ (less uncertainty means greater mutual information). Thus, information theory eliminates the paradox caused by the two requirements and creates the achievability of the optimal actuarial model.

Further, according to Fig. 3, the whole risk information transmission system is composed of risk encoder, risk mapper, and risk decoder three subsystems. Consequently, the information problems of the whole risk communication system can be decomposed into the information problems of three subsystems. The following sections will discuss these information issues in detail.

3.2 Solution: refinement of risk event

Let $\mathcal{Y}$ be the set of risk event $Y$, and divide $\mathcal{Y}$ into $k$ mutually disjoint sets, $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_k$, satisfying
\[
\bigcup_{i=1}^{k} \mathcal{V}_i = \mathcal{Y}
\]
(3)
Then $V$ is called a $k$-refinement of risk event set $Y$. Apparently, it holds

$$H(Y) \leq H(V) \quad \text{and} \quad I(Y, \hat{A}) \leq I(V, \hat{A}) \quad (4)$$

Equation (4) shows that, it is mathematically feasible to refine risk event $V$ for more risk classifications.

Fortunately, the Heinrich’s Safety Triangle (Heinrich’s Law) also ensures that the refinement of risk event has strong operability. Taking car insurance as an example, autonomous driving big data can refine risk event $Y$ as Fig. 5.

In Fig. 5, traditional car insurance can achieve a 2-refinement of risk event, $Y_1$ (Major-injury) and $Y_2$ (Minor-injury/property loss); With the aid of self-driving technology, insurance companies can achieve a 4-refinement of risk event, by appending $Y_3$ (Unreported violations and traffic hazards) and $Y_4$ (Unsafe behavior, conditions and hidden perils).

Significance The essence of risk event refinement is to ensure that risk events $V$ contain more risk information redundancy than estimated risks $\hat{A}$, so as to improve the capacity of risk decoding.

4 Capacity problem of risk mapping

4.1 Mutual information of risk mapping process

Next look at the risk mapper in Figure 3, an input and output system shown as Fig. 5.

Just like inequality (2), there is also a data processing inequality as follow

$$I(A, \hat{A}) \leq I(X, Y) \leq \min\{H(X), H(Y)\} \quad (5)$$

Where, $X$ and $Y$ are the risk input and output of risk mapper, and $I(X, Y)$ is the mutual information between $X$ and $Y$. $I(X, Y)$ describes the reliability (or shared information) that risk input variables $X$ generate (mapping) risk events $Y$. For a noisy (or erroneous) risk model, Risk Mapping Capacity is the maximum reliable mapping rate, which depends on the mutual information $I(X, Y)$ and information entropy $H(X)$ (inequality 5 gives these bounds). Therefore, the design of a good risk model should relate to optimize $I(X, Y)$ and $H(X)$.

4.2 Solution: maximization of risk mapping capacity

According to Shannon’s Channel Coding Theorem, the limit of the reliable risk mapping rate is the maximum mutual information $I(X, Y)$ (i.e. the “common uncertainty” of $X$ and $Y$). Thus, entropy and mutual information, can be used to determine the optimality of an actuarial system.

According to monotonicity and convexity of mutual information, $I(Y, X)$ is a convex function of $P(Y|X)$ for fixed $P(X)$ (which means that the actuary cannot change the pricing variables $X$ except for the actuarial model $P(Y|X)$). It can be derived from its convexity as

$$\lim_{P(Y|X)\rightarrow 1} I(X, Y) = \sup(I(X, Y)) \quad (6)$$

Where $\sup(I(X, Y))$ is the supremum of $I(X, Y)$. Equation (6) implies that when $X$ and $Y$ are close to one-to-one mapping, mutual information $I(X, Y)$ tends to achieve the maximum value.

According to another property of mutual information, $I(X, Y) \leq H(Y)$ with equality holding iff $Y$ is a function of $X$. That is

$$Y = G(X) \Rightarrow I(X, Y) = H(Y) \quad (7)$$

Combining equation (6) and (7), it shows that deterministic risk generating model $G(X)$ will cause

$$I(X, Y) = H(Y) = \sup(I(X, Y)) \quad (8)$$

In causal inference theory, $G(X)$ is called structural equation by [3].

Taking rear-end collision risk as an example, according to the vehicle dynamic model in reference [4], let

$$y = v_2 h_2 + \frac{v_1^2}{2a_1} - v_2 r_2 - \frac{v_2^2}{2a_2} \quad (9)$$

Where, $v_1$ and $a_1$ represent the initial speed and braking deceleration of the preceding vehicle 1, $v_2$ and $a_2$ are the initial speed and braking deceleration of the following vehicle 2, and $h_2$ and $r_2$ represent the time headway (gap) and reaction time of the following vehicle driver; $y$ is the risk event outcome variable. Here $v_1$, $v_2$, $a_1$, $a_2$, $h_2$, and $r_2$ belong to the risk input ($X$), and $y$ belongs to the risk output ($Y$).

Under the condition that $X$ is known, whether a collision occurs can be judged directly based on the calculation result of $Y$ ($y > 0$: no collision; $y \leq 0$: collision). Apparently the actuarial dynamics models similar to [3] can satisfy maximization of risk mapping capacity, because of their one-to-one mapping.
Three fundamental problems in risk modelling on big data: an information theory view

5 Distortion problem of risk variables

5.1 Information entropy of big data risk variables

Finally look at the risk encoder in Fig. 3, an input and output system shown as below. As shown in Fig. 6, the role of the risk encoder is to sample real-world risks (\(A\)) as big data risk input variables (\(X\)). Risk variables are usually continuous time variables, and \(X\) are discrete time variables. According to the Nyquist-Shannon sampling theorem, discrete sampling sequence may have information loss. Therefore, how to find the least-distortion and compact representation of real world risk information? This is the essential problem of risk encoder, which is directly related to the design of data collection on big data.

5.2 Constraints of risk big data design: multi-objective optimization

According to Shannon’s Source Coding Theorem, insurance big data variables (\(X\)) should be designed with redundant information to carry the information entropy of real risk as lossless as possible. This can be expressed mathematically as

\[
H(X) > H(A) \quad \text{or} \quad H(X) > H(\hat{A})
\]  

(10)

Where, \(H(X) > H(A)\) is a strong constraint, and \(H(X) > H(\hat{A})\) is a weak constraint, because \(A\) is usually unknown.

In practice, the cost of big data collection is unavoidable. Therefore, (10) must consider the cost constraints, namely

\[
\max_{H(X) > H(A), \text{Cost}(X) \leq \text{CoD}} H(X)
\]  

(11)

Where, \(\text{Cost}(X)\) is the cost function of big data collection (\(X\)), and \(\text{CoD}\) is the upper bound of cost.

In addition, it is worth emphasizing that the big data design of (\(X\)) will directly affect the risk mapping capacity of the actuarial model (i.e. risk mapper). According to monotonicity and convexity of mutual information, \(I(X,Y)\) is a concave function of \(P(X)\) for fixed \(P(Y|X)\) (which means the actuarial model \(P(Y|X)\) unchanged). It can be derived from its concavity as

\[
\max_{P(X)} H(X) = \sup_{P(X)} I(X,Y)
\]  

(12)

Equation (12) implies the fact that, not all insurance big data can be modeled as an excellent actuarial pricing model, only the big data collection design optimized
by mutual information can realize the best performance of the actuarial model.

(11) and (12) show that, the insurance big data design of \( X \) must not only extract the real-world risk characteristics without distortion, but also help the actuary find a good enough actuarial model.

Reference [11] presents an interesting and extreme example, assuming that driving risk (safety) model does not consider human factors such as driver's distraction, and only depends on the frequencies of wearing shoes observation variables in accident samples, it is easy to conclude: “Wearing shoes is the inevitable cause of the crash, because almost all drivers were wearing shoes when crashes occurred.” If the probability of driver wearing shoes is 99.999%, substituting the probability \( P(X) \) into (12), there is

\[
\lim_{P(1)=0.99999, P(0)=0.00001} I(X, Y) \rightarrow 0 \ll \sup I(X, Y)
\]  

According to \( I(X, Y) \)'s concavity, it can be directly inferred that the risk variable of wearing shoes is a very poor risk characteristic variable, so it is impossible to model it as a good actuarial model.

6 Conclusions

The traditional actuarial model is prone to cause ecological paradox and Simpson's paradox due to their correlation statistical assumptions, which were reported in Reference [6], a research on traffic big data. In this paper, the big data risk model is tackled as a risk information input-output system. This novel modeling framework can apply well-developed information theory to simultaneously achieve more risk classifications and less pricing errors.

The most important contribution of this work is to provide a new perspective of big data actuarial modeling, which enables actuaries to use more sophisticated mathematical tools such as information theory and system dynamics for risk modeling. In particular, specific approaches for optimizing the performance of big data actuarial models are proposed, including big data collection design, risk event refinement, and even causal (deterministic) modelling.

The work in this paper could have the potential to be widely used in insurance big data risk modeling, especially self-driving car insurance, cyber insurance.

References

1. Keith Ambachtsheer, Beyond portfolio theory: The next frontier, Financial Analysts Journal 61 (2005), no. 1, 29–33.
2. Aurelio Fernandez Bariviera, Luciano Zunino, M. Belen Guercio, Lisana B. Martinez, and Osvaldo A. Rosso, Efficiency and credit ratings: a permutation-information-theory analysis, Research Papers in Economics (2015).
3. A.R. Barron and T.M. Cover, A bound on the financial value of information, IEEE Transactions on Information Theory 34 (1988), no. 5, 1097–1100.
4. Edward A. Brill, A car-following model relating reaction times and temporal headways to accident frequency, Transportation Science 6 (1972), no. 4, 343–353.
5. Thomas M. Cover and Joy A. Thomas, Elements of information theory (wiley series in telecommunications and signal processing), 2006.
6. Gary A. Davis, Possible aggregation biases in road safety research and a mechanism approach to accident modeling, Accident Analysis & Prevention 36 (2004), no. 6, 1119–1127.
7. H. W. Heinrich, Industrial accident prevention: a scientific approach, 1959.
8. Stefan M. Moser and Po-Ning Chen, A student’s guide to coding and information theory, 2012.
9. Judea Pearl, Theoretical impediments to machine learning with seven sparks from the causal revolution, Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018, pp. 3–3.
10. Athanasios Sachlas and Takis Papaiannou, Residual and past entropy in actuarial science and survival models, Methodology and Computing in Applied Probability 16 (2014), no. 1, 79–99.
11. David Shinar, Crash causes, countermeasures, and safety policy implications, Accident Analysis & Prevention 125 (2019), 224–231.
12. Ilya Zheludev, Robert Smith, and Tomaso Aste, When can social media lead financial markets, Scientific Reports 4 (2015), no. 1, 4213–4213.