Insurance design for epidemic outbreaks involving Cramér-Lundberg model

Naoyuki Ishimura (Chuo University, Japan)

joint work with A.M. Molina Barreto and K.Takaoka

Outline

§1. Introduction
§2. Preliminary
§3. Insurance Design
§4. Empirical Study
§5. Discussion

1 Itroduction

We are concerned with a simple model of the insurance coverage due to the loss of COVID-19.

Since the emergence of a new kind of pneumonia at the end of 2019, the worldwide confusion was followed and our way of life has been forced to change.
In the academia, intensive and enormous studies have been performed from various viewpoints, which we cannot recall all of outstanding results.

On the other hand, it is also important to provide suitable insurances in order to mitigate the disaster in some extent for the loss of epidemic outbreaks. In this respect, one of the authors considers certain insurance model for the epidemic bursts in 2016 (see Ishimura, Komadel, and Yoshizawa [4]), three years before COVID-19, where the modelling is based on (i) the onset of epidemic outbreaks, (ii) the estimate of ultimate number of removals with the use of SIR model, (iii) the market risk.

Here, taking account of the fact that such an epidemic outbreak has really occurred, we take a different approach to design better and much practical insurances for the loss of desease.

Our main idea is the use of the Cramér-Lundberg model in our principal risk process, which is known to provide a fundamental modelling tools in the risk theory. The application to the model of epidemic outbreaks, however, seems to be new.

2 Preliminary

2.1 Risk process

We begin with considering the well known Cramér-Lundberg type model in discrete setting:

$$U(n) = u + cn - C(n), \qquad n = 0, 1, 2, \cdots \quad (1)$$

where

$$C(n) = \sum_{k=1}^{N(n)} X_k. \quad (2)$$

Here $u = U(0)$ is the initial surplus at time 0, and c is the insurer's rate of premium income per unit time, which is assumed to be continuously received.
$\{N(n)\}$ denotes the number of claim process, which is originally assumed to be a Poisson process with parameter λ.

The sequence of insurer's aggregate claim amount $\{X_k\}_{k=1,2,\dots}$ takes nonnegative values and assumed to be independent and identically distributed random variables with common distribution function $F_X(x)$.

Moreover, we assume that the process $\{N(n)\}$ and the sequence $\{X_k\}_{k=1,2,\dots}$ are independent.

Later, we will fix the insurance covering period $[0, T]$ with the expiration date T $(> 0, \in \mathbb{N})$ and assume that.

$N(n)$: the cumulative number of infectives at date n.

$\{X_k\}$: the random variables which are modelled by the infectious distribution of real data.

Since the premium is usually stochasitc, we put

$$U(n) = u + P(n) - C(n), \qquad n = 0, 1, 2, \cdots \tag{3}$$

where $P(n)$ denotes the insurer's premium income process.
For example, if the premium is paid at the beginning and the middle of the covering period,

$$P(n) = c1_{\{n \geq 0\}} + c1_{\{n \geq T/2\}},$$

where 1_A is the indicator function of the set A.

2.2 SIR model

It is common to appeal to the famous Kermack-McKendrick theory, which is also known as SIR model, for the situation of epidemic outbreaks.
The original SIR model is the system of ordinary differential equations for three sub-populations: $S(t)$ is the number of susceptibles to the disease, $I(t)$ is the number of infectives, and

$R(t)$ is the number of removals.

$$\frac{dS(t)}{dt} = -\beta S(t)I(t)$$
$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t)$$
$$\frac{dR(t)}{dt} = \gamma I(t),$$

where the constants β is the infection parameter and γ is the removal parameter representing the rate at which infectives become immune.

The important parameter is

$$\varrho = \frac{\gamma}{\beta}, \tag{4}$$

which is related to the reproduction number. In this model the total population

$$N = S(t) + I(t) + R(t)$$

is preserved.

Following Singh, Lal and Kotti [7], which is based on the idea of Wacker and Schlüter [10], we employ the implicit discrete system as follows.

$$
\begin{aligned}
\frac{S_{k+1} - S_k}{t_{k+1} - t_k} &= -\frac{\beta_{k+1}}{N} S_{k+1} I_{k+1} \\
\frac{I_{k+1} - I_k}{t_{k+1} - t_k} &= \frac{\beta_{k+1}}{N} S_{k+1} I_{k+1} - \gamma_{k+1} I_{k+1} \qquad (5) \\
\frac{R_{k+1} - R_k}{t_{k+1} - t_k} &= \gamma_{k+1} I_{k+1}
\end{aligned}
$$

where $\{t_k\}_{k=1}^{M}$ denotes the considered time sequence. In our empirical study, we take $\{t_k\}$ as the successive date of observations.
We note that N is preserved also in this discretization: namely,

$$N = S_k + I_k + R_k = S_{k+1} + I_{k+1} + R_{k+1}.$$

3 Insurance design

As is well known, insurance has been employed to manage insurable risks, which include, for example, liability, property, fire, automobile, life, health, pension, and so on.

Here we intend to design suitable insurance for the loss of epidemic outbreaks, such as COVID-19.

Our strategy is to use the risk model (3) for the insurance coverage.

Let $[0, T]$ be the insurance coverage period where $T\ (> 0, \in \mathbb{N})$ means the expiration date. Empirically it seems natural to proclaim that

$N(n)$: the cumulative number of infectives at date n.

$\{X_k\}$: the random variables which are modelled by the infectious distribution of real data.

We note that the interpretation of X_k will be much more flexible.
We understand that it is possible to assume that X_k are deterministic variables.

We also remark that, in our setting, $N(n)$ may no longer be a stochastic process but a deterministic process governed by the SIR model.

Now, the insurance should be designed so that insurers are also able to cover the risk's loss, which is represented by the total claim process $C(n)$ over the insurance period. Usually, calculations are based on the so-called premium principle. Here, the premium refers to the payment that a policyholder makes for insurance cover against a risk. Following Dickson[2], let us denote by Π_Z the premium that an insurer charges to cover a risk Z.

Our proposed insurance for the loss of epidemic outbreaks can be formulated as follows.

Proposition 1.
Insurances against epidemic outbreaks will be designed if the next condition is satisfied.

$$c + E[P(T)] - \Pi_{C(T)} > 0.$$

We recall that there are known several types of premium principles. Here are the list of examples which we will treat later.
(i) The expected value principle:

$$\Pi_Z = (1+\theta)E[Z],$$

where $\theta > 0$ is the premium loading factor.
(ii) The variance principle:

$$\Pi_Z = E[Z] + aV[Z],$$

where $a > 0$ is the loading factor.

(iii) The standard deviation principle:

$$\Pi_Z = E[Z] + a\sqrt{V[Z]},$$

where $a > 0$ is the loading factor.

In any case, within our model, it is important to effectively estimate $N(n)$ and X_k from the real data, which is the subject of the next section.

4 Empirical study

4.1 Selection of COVID-19 data

Now, we have obtained data for COVID-19 from the website of the Minister of Health, Labour and Welfare of Japan
(https://covid19.mhlw.go.jp).
The data for Tokyo from January 5th 2022 to July 3rd 2022 for a total of 180 observations,

which includes:

$\hat{I}$: the cumulative number of infected;

$\hat{R}$: the number of cases that are discharged from hospital/released from treatment;

$\hat{D}$: the cumulative deaths.

For the recovered cases we define

$$R_k = \hat{R}_k + \hat{D}_k$$

and for new infected cases

$$I_k = \hat{I} - R_k.$$

For estimating the parameters of the SIR model, we divide the total data into successive two parts:

$1 : 120$ data from January 5th to May 4th, 2022;

$2 : 60$ data from May 5th to July 3rd, 2022.

The initial window of $\{t_k\}_{k=1}^{120}$ is employed for estimating the parameters, and the rest of the data of $\{t_k\}_{k=121}^{180}$ is used for the validation of the model and forecasting.

To eliminate noises from the data, we have applied a 7MA (moving average) filter as smoothing filter.
Figure 1 shows the cumulative cases of reported infected people and the cumulative cases of reported recovered people in Tokyo.
The upper graph shows the cumulative infectives and recovered cases.
The lower graph shows the daily new cases and the cumulative recovered cases.

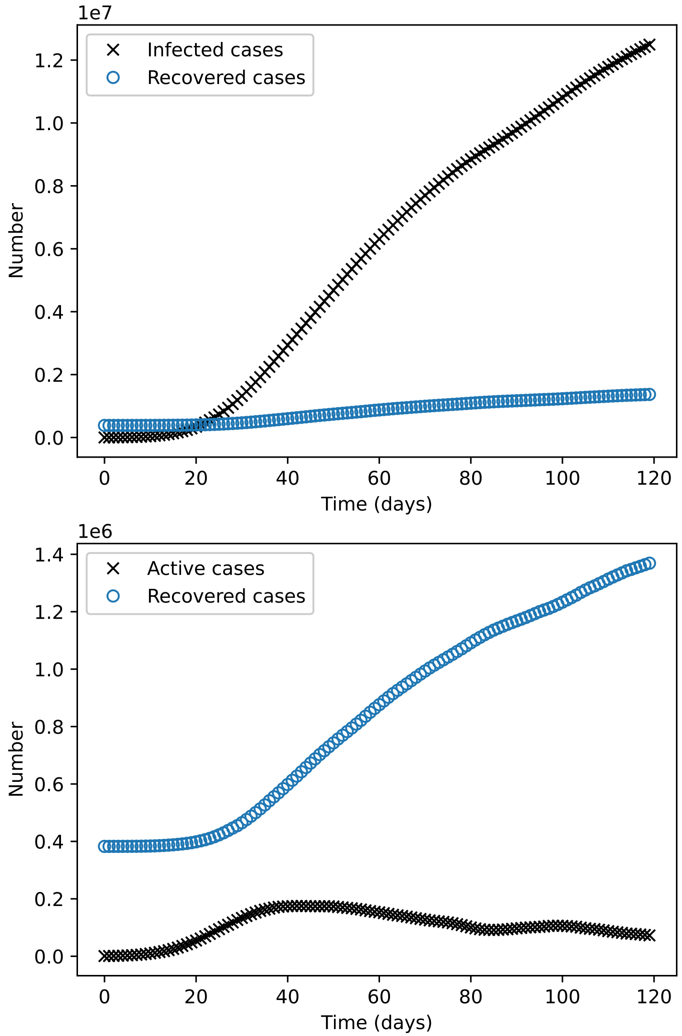

図1　COVID-19 data for Tokyo from $t = 1$ to $t = 120$.

4.2 Calculation of time-varying β and γ from observed data

We follow the methodology used in Singh, Lal and Kotti[7] for estimating the time-varying parameters for the SIR model.
Assuming that the discrete values of I and R for $k = 1, \ldots, M-1$ where $M = 120$, the time-varying parameter are computed as follows:

$$\beta_{k+1} = \frac{N\left(S_k - S_{k+1}\right)}{I_{k+1} \cdot S_{k+1} \cdot \Delta_{k+1}}$$

and

$$\gamma_{k+1} = \frac{R_{k+1} - R_k}{I_{k+1} \cdot \Delta_{k+1}},$$

where $\Delta_k = t_{k+1} - t_k = 1$ for $k = 1, \ldots, M-1$. With these parameters, we can calculate the time-discrete solution of the SIR model.

4.3 Model validation and short term forecast

We can assume that the time-varying transmission and recovery rate for $t \geq 1$ take the following form

$$\beta(t) := \beta_1 \cdot \exp(-\beta_2 t)$$

and

$$\gamma(t) := \gamma.$$

The constants β_1, β_2 and γ are real which can be determined by the sequences β_i^M and γ_i^M. It follows from a Maximum Likelihood estimation that the local minimum are given by

$$\hat{\gamma} = \frac{1}{M-1} \sum_{k=2}^{M} \gamma_k \,,$$

$$\widehat{\delta_2} = \frac{\sum_{k=2}^{M} t_k \cdot \ln\left(\beta_k\right) - \frac{1}{M-1}\left[\sum_{k=2}^{M} \ln\left(\beta_k\right)\right] \cdot \sum_{k=2}^{M} t_k}{\sum_{k=2}^{M} t_k^2 - \frac{1}{M-1} \cdot \left(\sum_{k=2}^{M} t_k\right)^2}$$

and

$$\widehat{\delta_1} = \frac{1}{M-1} \sum_{k=2}^{M} \left(\ln\left(\beta_k\right) - t_k \cdot \widehat{\delta_2} \right),$$

where $\delta_1 := \ln(\beta_1)$ and $\delta_2 := \beta_2$.
Using these solutions with the estimated model we can make short time forecast for the infected cases and recovered number for the COVID-19 infections in Tokyo.

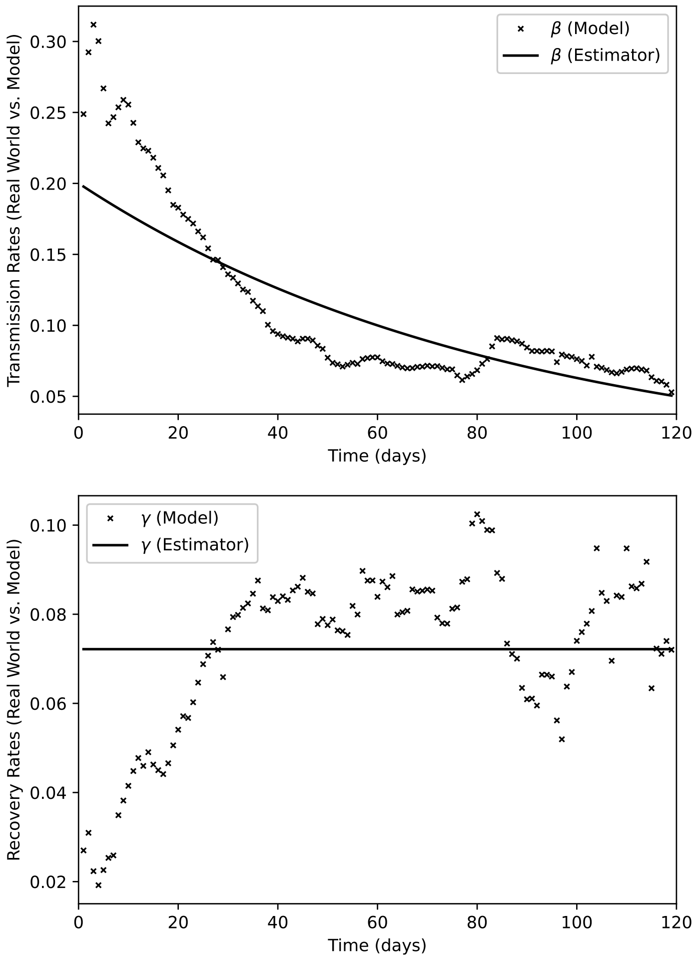

図2　Time-varying transmission and recovery rates from processed data for Tokyo and estimator functions.

In Figure 2 it is shown the transmission rate β for the model and the estimator as well values for the recovery rate γ. The estimated reproduction number ϱ is seen in Figure 3. The solution for the SIR model for $\{t_k\}_{k=1}^{120}$ is shown in Figure 4.
Finally, with the estimation of β and γ we can make a forecast for values of I_k and R_k where $\{t_k\}_{k=121}^{180}$ with the solution of (5) and MLE of transmission and recovery rates. In Figure 5 we can observe these solutions.

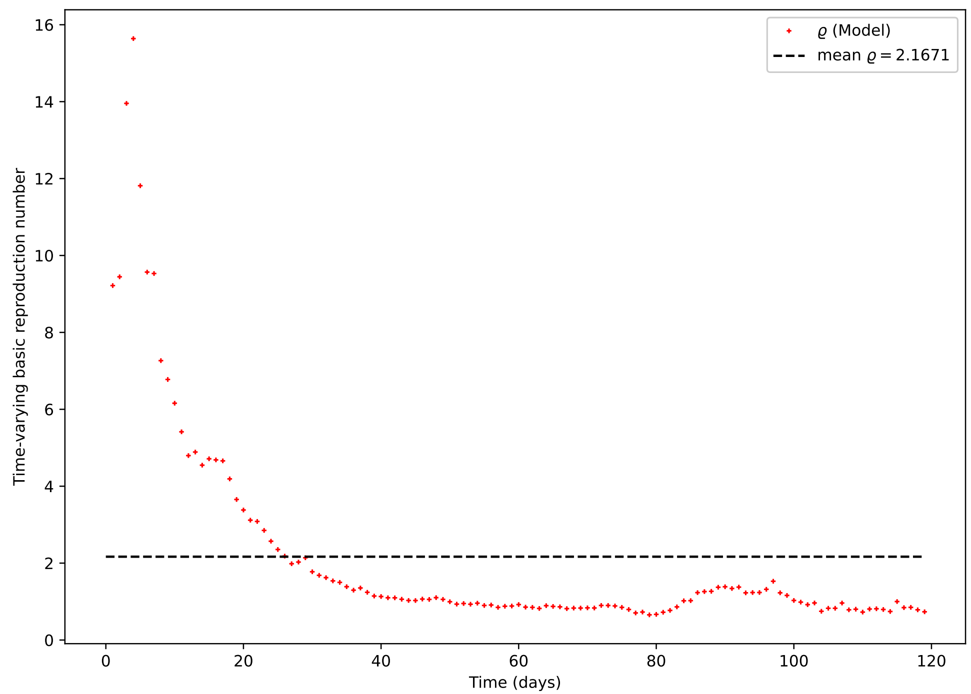

図3　Time-varying, and average effective reproduction number.

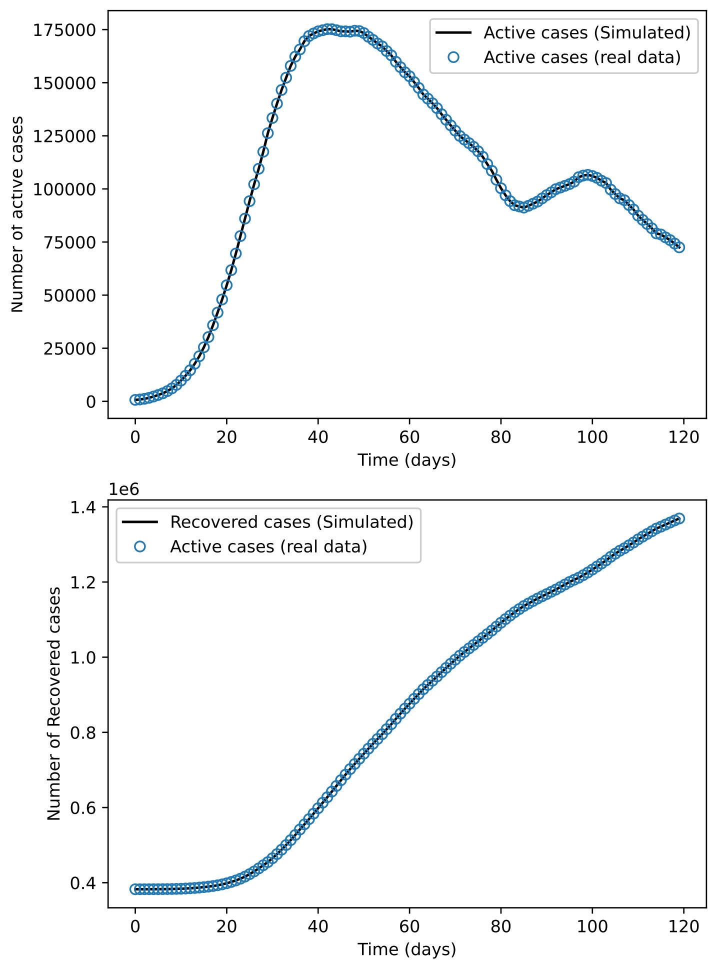

図4 Implicit time-discrete SIR model solution. In the left are shown data for (I_k) and to the right for (R_k).

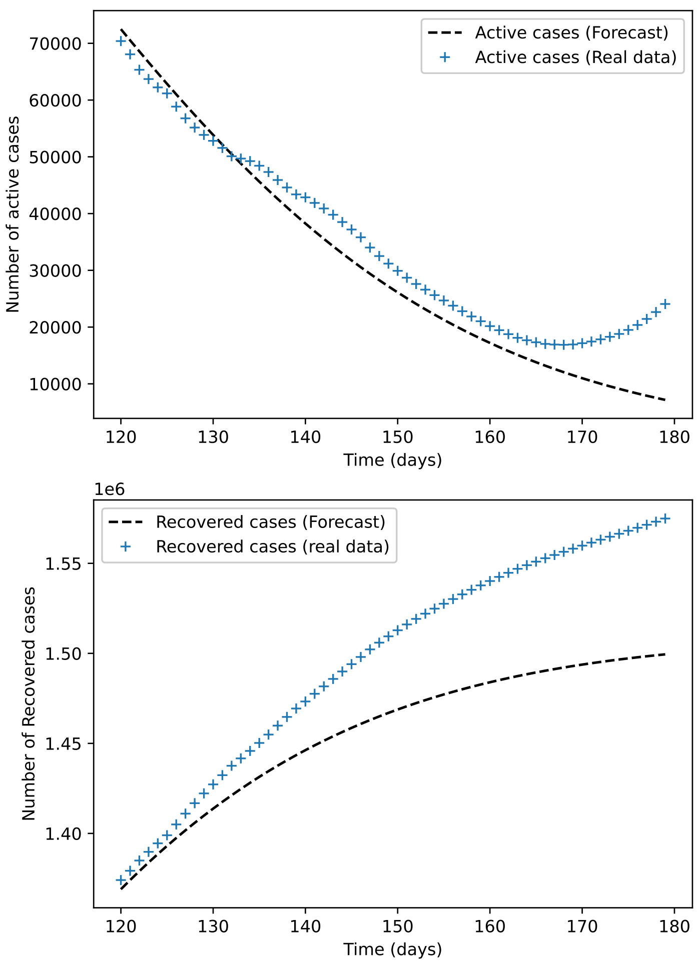

図5　SIR model solution forecast and real data. In the left are shown data for (I_i) and to the right for (R_i).

4.4 Examples of insurance products

We provide examples of insurance products.
The period we treat is 60 days $(T = 60)$ from May 5th to July 3rd, 2022.
We here regard $\{X_k\}$ as i.i.d. random variables which represent the loss due to COVID-19; the realization is assumed to result in $\{I_k\}$ with the abuse of notation, since I_k is used in the above for the computed deterministic value, but this time I_k is also regarded as a random variable.

We may thus infer that

$$C(T) = \sum_{k=1}^{N(T)} X_k = \sum_{k=1}^{T} I_k.$$

The right hand side can be interpreted and computed in two ways; one is the observed value from real data and the other is the calculated value as in §4.3.

The expectation and the variance are described as follows:

$$E[C(T)] = \frac{1}{T}\sum_{k=1}^{T} I_k,$$

$$V[C(T)] := \frac{1}{T}\sum_{k=1}^{T}(I_k - E[C(T)])^2.$$

For our window of period, we learn that the estimated values given by the SIR model to be:

$$E[C(T)] = 31337.19,$$
$$V[C(T)] = 374988335.05.$$

The values with real data, for comparison, are computed as

$$E[C(T)] = 35045.41,$$
$$V[C(T)] = 261796199.52.$$

The difference in variance is rather prominent for this period.
In both cases, we have estimated the risk $\Pi_{C(T)}$ with the principles of Proposition 1.

In this way, once the omen of epidemic outbreaks is observed, the insurers may be able to design suitable insurances for the coverage against diseases under appropriate premium principles.

4.5 Case study

We consider an insurance product of fixed payment. The period we treat is 59 days from February 1st 2022 to March 31th 2022. The insurance product against COVID-19 provided Sompo-Japan is three months, which includes the above period; however, we neglect the difference of the period in our analysis. The premium and the payment of this insurance product is 500 JPY and 50000 JPY, respectively.

Here we investigate the reason by our model why this insurance product has resulted in the suspension of insurance solicitation.

The model is just a simple version of (1)(2):

$$U(n) = u + cn - \sum_{k=1}^{N(n)} X_k.$$

We proclaim here that the surplus is for the individuals, which is different from above, and X_k is now interpreted as a deterministic fixed payment 50000 JPY.
As a result, we see that

$$c = \frac{500}{59}, \qquad E[X_k] = X_k = 50000,$$

since the current model is deterministic.

In the above period of 59 days, the number of infectives grows significantly everyday and we conclude that the maximum of $C(n)$ is attained at the last day of $n = 59$.
We estimate as in the previous subsections that $\hat{\beta} = 0.1175$ and $\hat{\gamma} = 0.09161$.
Moreover we assume that the population of Tokyo is $N = 14011487$ and we have $N(59) = 994028$.

The surplus for individuals then becomes

$$\begin{aligned} U(59) &= u + 59 \cdot \frac{500}{59} - 50000 \cdot \frac{994028}{14011487} \\ &\approx u - 3450. \end{aligned}$$

Therefore the initial surplus for the individuals should be over about 3500JPY.

5 Discussions

We have developed an insurance design for the loss of epidemic outbreaks.
The model is based on the well-known Cramér-Lundberg risk process, and with suitable premium principles, it is possible to introduce insurances against the risk of epidemic bursts.
Empirical studies are shown to support our model.
Further investigation is now in progress. .

参考文献

[1] A. Asmussen and H. Albrecher, *Ruin Probabilities, 2nd Edition* (World Scientific, Singapore, 2010).

[2] D. C. M. Dickson, *Insurance Risk and Ruin, 2nd Edition* (Cambridge University Press, Cambridge, 2017).

[3] A. Inoue, Y. Nakano, and K. Fukuda, *Finance to Hoken no Suuri (Mathematics for Finance and Insurance)* (Iwanami Shoten, Tokyo 2014, (in Japanese)).

[4] N. Ishimura, D. Komadel, and Y. Yoshizawa, Risk estimation model on epidemic outbreaks for an

insurer, presented in ASTIN Colloquim Lisbon 2016.
http://actuaries.org/lisbon2016/papers/19.NaoyukiIshim
[5] Y. Muroi, *Hoken to Kin'yu no Suuri (Mathematics for Insurance and Finance)* (Kyoritsu Shuppan, Tokyo 2017 (in Japanese)).
[6] H. H. Panjer and G. E. Wilmot, *Insurance Risk Models* (Society of Actuaries, Schaumburg, IL, 1992).
[7] R. A. Singh, R. Lal, and R. R. Kotti, Time-discrete SIR model for COVID-19 in Fiji, *Epidemiology and infection* **150** (2022) e75. 1–40,

https://doi.org/10.1017/S0950268822000590.

[8] H. Soutome, N. Ishimura, and H. Imai, Global in space numerical computation of the ruin probability, *Advances in Mathematical Sciences and Applications* **31** (2022) 397–406.

[9] M. Tomita, K. Takaoka, and M. Ishizaka, On the ruin probability of a generalized Cramér-Lundberg model driven by mixed Poisson processes, *J. of Appl. Prob.* **59** (2022) 849–859. https://doi.org/10.1017/jpr.2021.99

[10] B. Wacker and J. Schlüter, Time-continuous and time-discrete SIR models revisited: theory and

applications, *Advances in Difference Equations* **2020** (2020) 556.
https://doi.org/10.1186/s13662-020-02995-1

[11] Y. Yan, H. Pan, N. Shao, Y. Xuan, S. Wang, W. Li, X. Li, C. Y. Shen, X. Chen, X. Luo, Y. Chen, B. Xu, K. Liu, M. Zhong, X. Xu, Y. Jiang, S. Lu, G. Ding, J. Cheng, and W. Chen, COVID-19 in Singapore: Another story of success, *International J. Mathematics for Industry* **12** (2020) 205001.
https://doi.org/10.1142/S266133522050001X