情報処理学会論文誌 Vol.60 No.9 1528–1537 (Sep. 2019)

個人情報漏洩の損害額の新しい数理モデルの提案

山田 道洋[1,a)] 菊池 浩明[2,b)] 松山 直樹[2] 乾 孝治[2]

受付日 2018年12月10日, 採録日 2019年6月11日

概要：個人情報漏洩の被害額の算出にはセキュリティの保険の観点から大きな需要がある．JNSA は本人の特定容易度や企業の社会的責任度から損害賠償額を見積もる JO モデルを提案している．しかしながら，その係数や算出式は専門家が経験に基づいて主観的に決めたものであり，その信頼性が疑問視されている．それに対して，我々は企業規模などを説明変数とし，その企業の特別損失額を目的変数として重回帰を適用した新しい数理モデルを提案する．さらに，15,000 件から抽出した 144 件のインシデントにモデルを適用し，算出される損害額について先行研究の 2 つのモデルとの比較を行う．

キーワード：セキュリティ保険，情報漏洩，サイバーインシデント

Proposal on a Mathematical Model to Estimate Loss of Leakage of Personal Data

Michihiro Yamada[1,a)] Hiroaki Kikuchi[2,b)] Matsuyama Naoki[2] Koji Inui[2]

Received: December 10, 2018, Accepted: June 11, 2019

Abstract: There is a great demand from the viewpoint of security insurance to calculate the value of damage due to leakage of personal information. The Japan Network Security Association (JNSA) proposed a model to calculate the damage compensation amount. However, the coefficient was determined by experts' subjective evaluations for which there is no basis. We propose a new mathematical model by applying multiple regression using cyber incident records and information such as enterprise size as explanatory variables and the value of extraordinary losses to a company as a target variable. We apply the damage model to 144 cyber incidents chosen from 15,000 candidates, compare the two models' loss amounts, and consider the relationship between them.

Keywords: security insurance, data breach, cyber incident

1. はじめに

近年，不正アクセスや内部犯行などによる個人情報の流出事件が増加している．2014 年にはベネッセコーポレーション社の業務委託先の元社員が，与えられていた権限を利用し約 3,504 万件の個人情報を名簿業者 3 社へ売却していた [2]．また，幻冬舎は運営するウェブサイトへの不正アクセスにより，最大で 93,014 名のメールアドレスやユーザ ID が流出した可能性を報告している [3]．

このような個人情報漏洩による被害額の算出にはセキュリティ保険の観点から大きな需要がある．Gartner 社のような IT 戦略コンサルタントは，サイバー保険を効果的に使用するためのガイドラインを提供している [18]．保険業界の予測では，2015 年の約 20 億ドルから 2025 年には約 200 億ドルまでになると予測されている [17]．英国などの各国政府は，サイバーセキュリティリスク管理の改善のためにサイバー保険市場の成長を支援している [11]．Franke は，インタビューの結果からスウェーデンのサイバー保険

1 明治大学大学院先端数理科学研究科
Graduate School of Advanced Mathematical Sciences, Meiji University, Nakano, Tokyo 164–0001, Japan
2 明治大学総合数理学部
School of Interdisciplinary Mathematical Sciences, Meiji University, Nakano, Tokyo 164–0001, Japan
a) my.yama34@gmail.com
b) kikn@meiji.ac.jp

本稿は，2013 年 3 月の CSEC 研究会での論文「個人情報漏洩の損害額の新しい数理モデルの提案」[1] を基にしている．

市場の特徴を報告している [12].

日本ネットワークセキュリティ協会 (JNSA) は，2002 年に，本人の特定容易度や企業の社会的責任度から各組織が所有する個人情報の潜在的リスクを把握するための 1 つの推定手法として，想定損害賠償額算定式（JNSA Damage Operation Model for Individual Information Leak：JO モデル）を提案している [4]．JO モデルは 1 人あたりの基本情報価値を 500 円とし，氏名と住所が同時に漏れたときはその 3 倍とするなどのシンプルなルールから構成されたもので，広く知られている．しかしながら，我々は，JO モデルには次の問題点があることを指摘する．

(1) 500 円，3 倍などの定数は専門家の主観で定められたものであり，その根拠が不足しており，信頼性が低い．
(2) 16 年前に設計された古いモデルであり，最近の法改正などの事情が考慮されていない．
(3) 予測された損害額の信頼性が不明である．

そこで，本研究では，2005 年からの 12 年分の 1 万 5 千件の情報漏洩事件のビッグデータを解析し，より精度の高い最新の損失額の数理モデルを定式化することを試みる．同様の先行研究に，米国の 1 万件のインシデント情報から定式化した Romanosky のモデル [7] がある．しかし，このモデルは，Advisen 社の米国のインシデントデータセットに基づいたモデルであり，日本国内のインシデントについては同様のデータセットが存在しないため，国内のケースには適用が困難と考える．

そこで，我々は，企業が公開している会計情報 [8] に注目した．大規模な漏洩インシデントが生じたとき，その対応にかかるコストを特別損失に計上しているためである．こうして，漏洩した個人情報や企業規模などの情報を説明変数とし，その企業の特別損失額を目的変数として重回帰を適用し，新しい数理モデルを提案する．1 万 5 千件から抽出した 144 件のインシデントに損害モデルを適用し，損失額と先行する JO モデルとの比較を示し，両者の関係を考察する．

本稿の構成は次のとおりである．2 章は，関連研究の紹介を行い，3 章では提案方式について，4 章では提案方式の評価，5 章で考察を与え，6 章で本研究の結論を述べる．

2. 関連研究

2.1 JO モデル

JNSA セキュリティ被害調査ワーキンググループは 2002 年より新聞やインターネットニュースなどで報道されたインシデントの記事，組織からリリースされたインシデントに関連した文書の情報を集計し，漏えいした組織の業種，漏えい人数，漏えい経路などの分類・評価を行っており，「日付，情報管理・保有責任者（企業名），業種名，社会的貢献度，被害人数，漏洩情報区分，漏洩原因，漏洩経路，事後対応姿勢，漏洩情報（氏名，住所，電話番号，生年月日など）」といった事件の特性を記録している [6]．2005 年から 2016 年のインシデントの統計量を**表 1** に示す．

また，これらの情報から各企業の顧客 1 人あたりへの想定損害賠償額を算出する JNSA Damage Operation Model for Individual Information Leak（JO モデル）を提案している [4]．算出式を次に示す．

$$
\begin{aligned}
&\text{損害賠償額}\\
&\quad = \text{漏洩情報価値} \times \text{社会的責任度} \times \text{事後対応評価}\\
&\quad = (\text{基礎情報価値} \times \text{機微情報度} \times \text{本人特定容易度})\\
&\qquad \times \text{社会的責任度} \times \text{事後対応評価}\\
&\quad = \text{基礎情報価値 } [500]\\
&\qquad \times \text{機微情報度 } [\max(10^{\max(x)-1} + 5^{\max(y)-1})]\\
&\qquad \times \text{本人特定容易度 } [6, 3, 1]\\
&\qquad \times \text{社会的責任度 } [2, 1]\\
&\qquad \times \text{事後対応度 } [2, 1]
\end{aligned}
\tag{1}
$$

ここで，[ ] 内の値は，それぞれの項目の値域である[*1]．機微情報度，本人特定容易度は漏洩した情報によって定められている，機微情報度は，3 値を取る精神的レベル x と経済的レベル y の 2 変数で与えられる．たとえば，氏名の場合，x=y=1 であるが，病名は x=1，y=2 である．本人特定容易度は

$$
\text{本人特定容易度} = \begin{cases} 6 & \text{氏名 and 住所} \\ 3 & \text{氏名 or (住所 and 電話番号)} \\ 1 & \text{その他} \end{cases}
$$

と定められている．JO モデルを用いて，2014 年に発生したベネッセコーポレーション社のインシデントによる損害額を算出すると，パラメータはそれぞれ 精神的レベル $x = 2$, 経済的レベル $y = 1$, 本人特定容易度 $= 6$, 社会的責任度 $= 1$, 事後対応度 $= 1$ となり，1 人あたりの損害賠償額は 33,000 円となる．同インシデントの被害人数は 4,858 万人とされており，損害賠償額は全体で 1 兆 6 千 31 億 4 千万円となる．

2.2 Romanosky モデル

Romanosky は Advisen 社[*2]より入手した 2005 年から 2014 年のアメリカの企業の 11,705 件のインシデント情報をもとに，各年に企業が被った総コストを算出するモデルを次のように提案している [7]．なお，単位は百万ドルである．

[*1] これらの定数は，セキュリティ被害調査 WG が独自に策定した想定損害賠償額算定式に基づくものである．文献 [23] によると，過去に発生した会員情報インシデントにおける訴訟参加率 0.02 = 10 人/5 万人から算出している．また，文献 [5] によると，平成 13 年の宇治市住民基本台帳データ大量漏えい事件を参考にしたとされている

[*2] https://www.advisenltd.com/

表 1 JNSA のデータセットの統計量
Table 1 Statistics of JNSA dataset.

期間	レコード数	企業数	属性数	平均被害人数	平均インシデント数/年	平均想定損害賠償額（円/人）	平均想定損失額（百万円）
12 年間	15,569	8,853	25	11,764.32	1,297.42	42,361.73	460.27

表 2 Romanosky の提案モデルの各係数（一部）[7]
Table 2 Coefficients of Romanosky's model [7].

係数			Estimate	
定数	β_0		−3.858	*
$\log(revenue_{i,t})$	β_1		0.133	**
$\log(record_{i,t})$	β_2		0.294	***
$repeat$	β_3		−0.352	
$malicious$	β_4		−0.029	
$lawsuit$	β_5		0.444	
		Government	−1.339	
$FirmType_{i,t}$	α	Private	−1.032	
		Public	−0.065	

$$\begin{aligned}\log(cost_{i,t}) = \beta_0 &+ \beta_1 \cdot \log(revenue_{i,t}) \\ &+ \beta_2 \cdot \log(records_{i,t}) \\ &+ \beta_3 \cdot repeat_{i,t} + \beta_4 \cdot malicious_{i,t} \\ &+ \beta_5 \cdot lawsuit_{i,t} + \alpha \cdot FirmType_{i,t} \\ &+ \lambda_t + \rho_{ind} + \mu_{i,t}\end{aligned} \quad (2)$$

ここで，各係数の値を**表 2** に示す．i，t は t 年の企業 i のデータを参照すること示し，$revenue$ は収益[*3]，$records$ は漏洩情報の件数を示している．$rpeat$，$lawsuit$ はブール値，$FirmType$ はダミー変数として，過去に事件を起こしているか，事件について提訴されたかどうか，政府機関か一般企業かなど，それぞれあてはまる場合に 1，それ以外は 0 を取る．λ_t は，t 年のみを 1 とする年次のベクトル，ρ_{ind} はその企業の業種 ind を表すベクトル，$\mu_{i,t}$ はエラー項である．これらの係数は 12,000 件の観測から得られたインシデントとコストの関係を表している．これらの中で，コストの主要な要因は，収益であると主張している．

しかし，このモデルの係数はアメリカの企業の情報をもとにしたものであり，日本の企業についても同じ係数が適用できるかは疑問である．

2.3 その他関連研究

アメリカでは，個人情報の盗難による企業および個人の損失が 2005 年には 560 億ドルとなり，35%は企業において発生した個人情報漏洩事件によるものであった．Romanosky らは，2002 年から 2009 年の間の個人情報漏洩事件に関する法律（data breach disclosure laws）の影響を推定した [16]．彼らは，同法の採用により，個人情報の漏えいが平均で 6.1%減少することを報告している．文献 [15] によると，個人が金銭的損害を被ると企業が訴訟を起こされる確率は，オッズ比で 3.5 倍高くなる．さらに，原告が金銭的損失を被った場合や，認定された集団訴訟に直面したときには訴訟が和解する割合は 30%増になることが示されている．

Gordon らは与えられた情報を保護するために投資する最適な金額を決定するモデルを提案している [13], [14]．ここでは，企業における情報セキュリティに対する投資が，情報漏洩やサイバー攻撃から生じると予想される損失の 37%を超えてはならないことを推奨している．

Edward らは一般的な公開データセットを研究し，ベイズの一般化線形モデルを開発してデータ侵害の傾向を調査した [19]．

櫻井は，ネットワーク利用者に対する 2 回の社会調査から得られたデータに基づき，コンジョイント分析を利用してネットワーク利用者が個人情報についてどのような評価をするかを求め，主要な 6 種類の個人情報に対して，個人情報漏洩事故が発生した場合に慰謝料として受け入れる受入補償額を推定した．その結果，推定した金銭的評価は，現実に個人情報漏洩被害者に支払われた慰謝料をはるかに上回ることが示された [21]．

石川らは，個人情報漏洩インシデント発生時に企業が実際に支払った補償額を調査し，JO モデルで算出される想定賠償額と少なくとも 2 倍以上のギャップが存在することを報告している [22]．また，JO モデルについて，「検索容易性」，「変更容易性」，「回収容易性」の 3 点をモデルに組み込むことを提言している．

3. 提案方式

3.1 概要

本研究では，QICK Astra Manager [8] より購入した本決算（連結優先）データの個人情報流出インシデントが発生した年の会計情報，および，2005 年から 2016 年までの JNSA データセット [4], [6] に記載されている情報漏洩インシデントのデータを使用する．日本には Advisen 社に該当する企業がない．そこで，公開されている会計情報に注目した．

JO モデルは，個人情報を取り扱う組織の潜在的なリスクを把握することを目的として，個人情報の漏えいが生じたときに被害者全員が賠償請求する仮定の下で試算された想定損害賠償額を定式化している [5]．一方，本研究では，個人情報の漏えいインシデントが発生したときに，組織が被る被害者へのお詫び料と情報セキュリティ対策費用の評価モデルを提案する．これらの情報は，企業の財務内容を報告する年次事業報告書であるアニュアルレポートから複

[*3] $revenue$ には「純利益」，「歳入」などの意味があるが，本稿ではこれを「売上額」と解釈して算出する．

表 3 関連研究との比較
Table 3 Comaprison of models.

		JO モデル [5]	Romanosky[16]	提案モデル
アプローチ		発見的	回帰	回帰
評価対象	呼称	想定損害賠償額	*cost*	被害額
	損害賠償額	含む	含む	含まない
	セキュリティ対策	含まない	含む	含む
	お詫び料	含まない	含む	含む
試算のソース		JNSA	Advisen	JNSA+決算短信

数の企業について観測することができる．客観的に評価を行うために，インシデントが生じた年の特別損失額を目的変数として重回帰を適用してモデルを構築する．したがって，先行研究 JO モデルの損害賠償額とは厳密には異なる概念であるが，個人情報にかかわる潜在的なリスクを定量化するという観点で目的を共有しており，評価モデルの 1 つとして本稿にて提案する．関連研究との違いを**表 3** に整理する．式 (1) と (2) は形式が積と和で大きく異なるように見えるが，等価な形に変形できることを 4.1 節で示す．

3.2 特別損失額

企業の通常の経営活動では発生しない，特別な要因によって一時的に発生した損失を特別損失という．ベネッセホールディングスは 2014 年に発生した個人情報流出事件のお詫びにかかる費用などとして約 260 億円を特別損失として計上しており [9]，他の企業においてもインシデントによる損害額は特別損失として計上されると考えられる．そこで本研究では，各企業の特別損失額がそのインシデントにかかったコストを代表していると考える．

3.3 データのクレンジング

特別損失額には，インシデントによる損害額が計上されると考えられるが，特別損失額には「システム開発中止にともなう損失」や「事業構造改善費用」なども含まれており，特別損失額の全額がインシデントに関与しているわけではない．そこで，特別損失額，JNSA データセットに対して，重回帰を適用するための次のようにデータの加工，精査を行う．

3.3.1 年度ごとのデータの集約

サイバーエージェント社では 2016 年 5 月 11 日に不正ログインが発生した後，同年 11 月 29 日に再び不正ログインが発生している．このように，同一企業で同年度中に複数回インシデントが発生していた 21 社の，「被害人数」「漏洩原因」「漏洩情報」「事後対応度」「経済的ランク」「精神的ランク」「本人特定容易度」の項目についてデータの集約を行う．各項目の集約方法を以下に示す．

- 被害人数：1 年間の被害人数の合計
- 漏洩原因：「内部犯行」，「不正な情報持ち出し」などの故意による漏洩が 1 件でもあったかどうか
- 漏洩情報：1 年間で漏洩したすべての項目の和集合

表 4 アニュアルレポートの調査対象
Table 4 Objects of the annual report survey.

対象年数	対象件数	対象企業数
2005–2016	105	90

表 5 セキュリティ対策費と特別損失額の単回帰結果
Table 5 Incident-related and extraordinary losses estimated via single regression.

係数	*Estimate*	*p.value*	
定数	−11.956	0.443	
特別損失額	0.850	$3.48E-9$	***
定数	0	N/A	
特別損失額	0.849	$6.23E-12$	***

- 事後対応度：1 年間の最大値
- 経済的ランク：1 年間の最大値
- 精神的ランク：1 年間の最大値
- 本人特定容易度：1 年間の最大値

3.3.2 アニュアルレポートの調査

インシデント 105 件の企業の決算短信，アニュアルレポートなどを被害人数の多い順の 90 社を調査した．調査対象の統計を**表 4** に示す．

調査の結果，表 6 の 5 件のレポートに特別損失の内訳として「情報セキュリティー対策」と記載されていた[*4]．情報セキュリティー対策と記載されていた企業名とその金額を**表 6** に整理する．このセキュリティ対策費を真の損失額とし，特別損失額による単回帰を行った結果を**表 5** に示す．回帰の結果から，セキュリティ対策費と特別損失額には強い相関がある．また，定数項の −11.956 は全体の特別損失額からは誤差の範囲であり，p 値も高い．そこで，本研究では定数項を 0 として回帰の近似を行った結果から，

$$\text{漏洩損害額 } y = 0.849 \cdot \text{特別損失額}$$

と定義する．表 6 に示すとおり，セキュリティ対策費（真値）と漏洩損害額の誤差は平均で 10.87 百万円であり，95%信頼区間は [−18.37 百万円, +40.11 百万円] である．

3.3.3 特異なデータの除外

特別損失額は災害や社会情勢などの影響によっても増加する．2008 年に起こったリーマンショックにより多くの企業が影響を受け，2008 年前後の特別損失額が大きく増加している企業が存在していた．このようなリーマンショックによる影響を排除するため，本研究では 2010 年以降のデータを使用する．

銀行では，貸付債権と有価証券の損失を数年にわたって

[*4] セキ株式会社のレポートには，「昨年 9 月 15 日付で『当社お客様情報の流出に関するお詫びとお知らせ』を公表しました．その後の二次的な被害に関しましては，現在のところ報告されておりません．外部からの不正アクセスにより個人情報が外部に流出した懸念があり，それらにかかわる対応費用を情報セキュリティ対策費として計上しております．」[10] というように情報流出インシデントへの補償によるものであると明記されている．

表 6 情報セキュリティー対策費 [百万円]
Table 6 Information security countermeasure [M JPY].

企業名	年度	セキュリティ対策費	特別損失額	0.849· 特別損失額	誤差
ベネッセホールディングス	2015	26,039	30,642	26,045.7	+6.7
セキ	2016	210.67	234	198.9	−11.77
ストリーム	2014	5.56	66	56.1	+50.54
ミサワ	2012	27.24	42	35.7	+8.46
アークン	2016	8.92	11	9.35	+0.43
平均		5,256.69	6,199.4	5,269.15	+10.87
信頼区間（95%）					10.87 ± 29.24

表 7 重回帰対象のデータセット
Table 7 Dataset for multiple regression.

期間	レコード数	企業数	平均被害人数	平均売上高（百万円）	平均特別損失額（百万円）
2010～2016	144	115	356,630.2	407,317.46	5,812.27

表 8 提案モデルにおける係数
Table 8 Coefficient of proposed model.

係数				*Estimate*	*p.value*		定義域	平均値
		β_0		−3.9632	0.0093	***		
	log(被害人数)	$\log(x_1)$	β_1	0.0379	0.4612			6.15
	log(売上高)	$\log(x_2)$	β_2	0.9904	$2.18E-23$	***		11.40
	故意	x_3	β_3	0.6261	0.6808		0, 1	0.15
	事後対応度	x_4	β_4	N/A	N/A		0, 1	0
	経済的ランク	x_5	β_5	0.1590	0.5025		1, 2, 3	1.31
	精神的ランク	x_6	β_6	0.0128	0.9772		1, 2, 3	1.11
	本人特定容易度	x_7	β_7	0.2079	0.6930		1, 3, 6	4.26
業種	不動産業	x_8	β_8	−0.0773	0.9071		0, 1	0.08
	建設業			−1.4450	0.0258	*		0.10
	情報通信業			−0.1350	0.8014			0.19
	林業			−0.4030	0.7309			0.01
	電気・ガス			−0.9330	0.3081	*		0.03
	生活関連サービス業，娯楽業			−1.004	0.3771			0.01
	卸売業，小売業			−0.4550	0.4203			0.17
	医療，福祉			−0.6319	0.5045			0.03
	宿泊業，飲食サービス業			−0.4607	0.53801			0.04
	製造業			−0.7577	0.1728			0.17
	教育，学習支援業			−0.0654	0.9531			0.02
	学術研究，専門・技術サービス業			−0.1173	0.9219			0.01
	金融業，保険業			−1.7570	0.0572	*		0.03
	運輸業，郵便業			−0.8893	0.7994			0.03
	氏名	x_9	β_9	−0.6231	0.6007		0, 1	0.82
	住所	x_{10}	β_{10}	−0.5169	0.7406		0, 1	0.55
	電話番号	x_{11}	β_{11}	−0.5337	0.7562		0, 1	0.51
	生年月日	x_{12}	β_{12}	−0.2348	0.5105		0, 1	0.26
	性別	x_{13}	β_{13}	0.2624	0.5296		0, 1	0.17
	職業	x_{14}	β_{14}	0.1453	0.7767		0, 1	0.07
	メールアドレス	x_{15}	β_{15}	−0.3845	0.2318		0, 1	0.46
	ID/PASS	x_{16}	β_{16}	−0.2810	0.5025		0, 1	0.12

特別損失に乗せて計上し，ある閾値を超えた時点で急激に増額する．つまり一般の企業における特別損失額と計上年とは意味が大きく異なる．そのため，本研究では銀行のインシデントデータは除外する．

3.3.4 決算データを取得できないデータの除外

売上高や特別損失額は証券コードをもとに取得しているため，日本年金機構や日本郵政といった証券コードのない企業や団体のインシデントデータを除外する．

3.4 線形重回帰モデル

以上のデータの加工，精査を行った結果データは 144 件となった．対象としたデータセットの統計を**表 7** に示す．144 件のデータに対して，漏洩損害額 y を目的変数として重回帰を適用して得た次の線形モデルを提案する．重回帰には R の lm 関数を用いた．

$$\begin{aligned}\log(y) &= f(x_1, x_2, \cdots, x_{16}) \\ &= \beta_0 + \beta_1 \cdot \log(x_1) + \beta_2 \cdot \log(x_2) \\ &\quad + \beta_3 \cdot x_3 + \cdots + \beta_{16} \cdot x_{16}\end{aligned} \tag{3}$$

式 (3) の各係数と，説明変数の閾値を**表 8** に示す．*を $p < 0.1$（有意水準 10%）とし，**を $p < 0.05$（有意水準 5%），***を $p < 0.01$（有意水準 1%）とする．提案モデルでは売上高に高い有意差（***）がみられ，漏洩損害額は売上高に強く依存している．また，建設業などの業種の一部にも有意差が見えるが，これらは対象インシデントの数が少ないことが原因である．

4. 評価

売上高についての漏洩損害額の散布図と提案損失コストモデルの回帰直線を**図 1** に示す．なお売上高 x_2 と漏洩損害額 y 以外の項目 $x_1, x_3, \cdots x_{16}$ には各値の平均値を入力している．

被害人数 x_1 についての漏洩損害額 y の散布図と提案損失モデルの回帰直線を**図 2** に示す．なお被害人数と漏洩損害額以外の項目 $x_2, x_3, \cdots$ には各値の平均値を入力している．図 1 と比較して，被害人数 x_1 は損害額への相関が高くない．

被害人数 x_1 についての漏洩損害額の 3 つのモデルを**図 3** に示す．JO モデルは被害人数によって損害額は比例し，影響が大きいのに比べ，提案モデルと Romanosky のモデルでは被害人数による損害額は影響が小さいことが分かる．

4.1 JO との比較

式 (1) の JO モデルは，1 人あたりの賠償額が漏洩した情報によって定数倍されるモデルである．一方，提案の式 (3) のモデルは線形式で両者は異なるように見える．しかし，提案モデルを下記のように変形することによって JO モデルと等価であることを示す．

$$\begin{aligned}y &= e^{f(x)} \\ &= e^{\beta_0+\beta_1\cdot\log(x_1)+\beta_2\cdot\log(x_2)+\beta_3\cdot x_1+\beta_4\cdot x_2+\cdots} \\ &= e^{\beta_0} \cdot e^{\beta_1\cdot\log(x_1)} \cdot e^{\beta_2\cdot\log(x_2)} \cdot e^{\beta_3\cdot x_1} \cdot \cdots \\ &= e^{\beta_0} \cdot x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot e^{\beta_3\cdot x_3} \cdot \cdots\end{aligned} \tag{4}$$

提案モデルでの算出額と JO モデルでの想定損害賠償額の比較を**表 9** に示す．No.1〜5 は情報セキュリティ対策費を取得できた企業であり，No.6 以降はそれらを除いた被害人数の多い順に 15 社を示している．

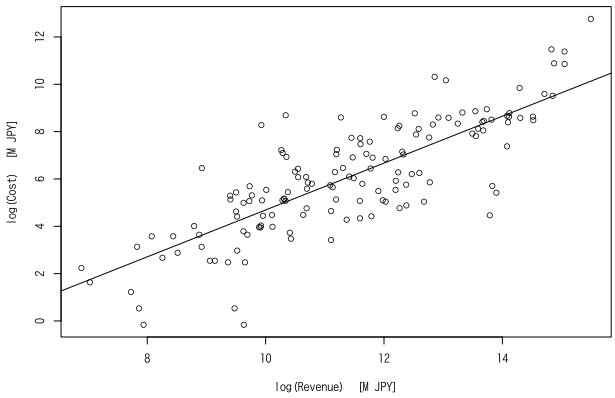

図 1 売上高と漏洩損害額の散布図
Fig. 1 Scatter plot between revenue and cost.

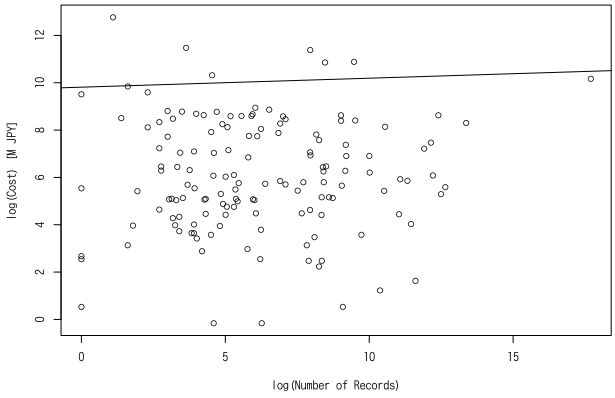

図 2 被害人数と漏洩損害額の散布図
Fig. 2 Scatter plot between # record and cost.

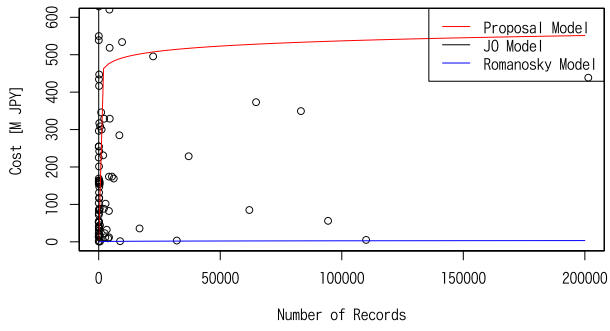

図 3 被害人数と漏洩損害額の散布図：各モデルの比較
Fig. 3 Scatter plot between # record and cost (comparison of three models).

提案モデルと JO モデルでは算出額に大きな違いが見られ，JO モデルでは平均で約 116 億円の損害賠償額が生じていた．しかし，表 3 に示したように，本研究での想定損害額および，漏洩損害額として定義した損失には損害賠償額を含まないが，先行研究では損害賠償額を含んで損害額を算出する．そのため，誤差率の単純な比較にはならないことに注意が必要である．

JO モデルでは経済的ランク，精神的ランク，本人特定容易度について，その段階によって 10 倍，5 倍のように想

表 9 各モデルでの漏洩損失額（被害人数の上位 20 社）
Table 9 Cost in each model (top 20 companies of many victims).

No	企業名	日付	被害人数	JO モデル	Romanosky	提案モデル	漏洩損害額	情報セキュリティ対策費
1	ベネッセホールディングス	2014/7/9	48,580,000	160,3140	2,367.6	13,287.4	26,045.7	26,039
2	セキ	2015/9/15	267,000	41,652	325.2	87.4	198.9	210.7
3	ストリーム	2014/1/30	94,359	566.2	256.6	152.9	56.1	5.6
4	ミサワ	2011/5/26	16,798	1,310.2	126.9	17.1	35.7	27.2
5	アークン	2016/1/13	3,859	23.2	66.9	4.4	9.4	8.9
6	サイバーエージェント	2016/11/29	640,368	742.2	466.4	3,532.6	4,021.4	
7	コシダカホールディングス	2014/9/17	310,000	930.0	403.7	199	266.9	
8	サイバーエージェント	2013/8/12	243,266	1,459.6	446.9	1,273.4	5,566.7	
9	パスコ	2010/3/21	201,414	9,063.6	355.0	637.1	438.6	
10	GMO インターネット	2015/2/27	188,047	1,011.3	276.5	1,444.7	1,752.7	
11	アミューズ	2009/8/10	148,680	11,597.0	307.1	187.9	1,362.6	
12	レアジョブ	2012/5/14	110,000	330.0	182.8	7.9	5.1	
13	江崎グリコ	2016/3/7	83,194	6,489.13	361.5	1,375.6	349.4	
14	椿本チエイン	2016/11/14	64,742	194.2	311.1	612.9	373.2	
15	ホットマン	2014/7/1	61,977	1,115.6	227.8	51.9	85	
16	サイバーエージェント	2014/6/23	38,280	76.6	267.7	1,852.9	3,427.2	
17	サニーサイドアップ	2015/8/28	37,006	37	184.4	145.4	228.7	
18	リブセンス	2013/2/28	3,2132	282.8	98.2	4.6	3.4	
19	良品計画	2015/1/5	22,385	405.1	165.9	716.1	495.6	
20	学研ホールディングス	2015/7/13	22,108	132.7	205.9	509.4	1,002.2	
	平均		356630.2	11,686.5	107.4	2,741.5	4,940.4	
	最大		48580000（ベネッセ）	1,603,140（ベネッセ）	2,367.6（ベネッセ）	36,953.7（東京電力）	349,630.7（東京電力）	
	最小		1	0.002	6.7	4.0	1.7	
	平均誤差			17,650.7	6,363.7	4,935.2		
	平均誤差率			4.54	2.50	1.73		

定損害額が定数倍されていた．提案モデルで，この定数の妥当性を検討する．たとえば，本人特定容易度 $x_7 = 1$ と 3 の損失額の比は

$$\begin{aligned}&\frac{f(x_1,\cdots,x_6,x_7=3,x_8,\cdots,x_{16})}{f(x_1,\cdots,x_6,x_7=1,x_8,\cdots,x_{16})}\\&=\frac{e_0^{\beta}\cdot x_1^{\beta_1}\cdots e^{\beta_6}\cdot e^{3\beta_7}\cdot e^{\beta_8}\cdots e^{\beta_{16}}}{e_0^{\beta}\cdot x_1^{\beta_1}\cdots e^{\beta_6}\cdot e^{1\beta_7}\cdot e^{\beta_8}\cdots e^{\beta_{16}}}\\&=\frac{e^{3\beta_7}}{e^{1\beta_7}}\\&=e^{2\beta_7}=1.5158<3\end{aligned}\tag{5}$$

と算出される．すなわち，損害賠償額を含まない提案モデルにおいては，JO モデルで 3 倍としているのは，高すぎで，1.5 倍の方がより適している．同様にして，提案モデルでの x_7 以外の経済的ランクなどが 1 段階上がったときの影響を計算することができる．それぞれの段階について提案モデルでの影響を計算した結果を**表 10** に示す．

提案モデルではいずれの変数についても，1 段階上がったときの影響は JO モデルよりも小さい．12 年間のインシデントデータに基づき，損害賠償額を含まずに検討すると，経済的ランク，精神的ランクはいずれも損失額を 10 倍，5 倍するほど影響を及ぼしていないことが分かる．本人特定容易度については，損失に影響しているといえるがその比は JO モデルの係数の約 1/2 であった．

表 10 提案モデルと JO モデルの係数の比較
Table 10 Comparison of coefficients of our proposed model and the JO model.

経済的ランク	JO モデル	10^0	10^1	10^2
	提案モデル	1	1.1723	1.3743
精神的ランク	JO モデル	5^0	5^1	5^2
	提案モデル	1	1.0129	1.0261
本人特定容易度	JO モデル	1	3	6
	提案モデル	1	1.5158	2.8291

経済的ランク，精神的ランク，本人特定容易度がすべて 1 のインシデントのデータにより見積もった，1 人あたりへの基礎情報価値を**表 11** に示す．経済的ランク，精神的ランク，本人特定容易度がすべて 1 ならば，式 (1) より JO モデルでの 1 人あたりへの想定損失額は 500 円となるが，提案モデルでは 212,106.1 円である．この 12 年間で漏洩インシデントが企業に及ぼす影響が大きくなっていることを表している．

表 11 基礎情報価値（$x_5 = x_6 = x_7 = 1$ の漏洩損害額）
Table 11 Mean loss for common case (cost of $x_5 = x_6 = x_7 = 1$).

件数	平均被害人数	平均漏洩損害額 [百万円]	平均漏洩損害額 [円/人]
20	5,031.3	1,067.17	212,106.1

4.2 Romanosky との比較

回帰に使用したインシデントのデータに対して，Romanosky のモデルを適用した場合と提案モデルとの比較を表 9 に示す．また，$lawsuit_{i,t}$，$FirmType_{i,t}$，λ_t，ρ_{ind}，$\mu_{i,t}$ については 0 として無視している．Romanosky のモデルでは平均が 107.4 百万円と算出額が非常に小さい．Romanosky の回帰に使用したデータでは revenue の平均が 8,031 百万ドルとなっている．このことから，アメリカと日本の市場規模の大きさの違いにより Romanosky のモデルが日本のインシデントに適応できていないと考えられる．

4.3 モデルの簡略化

式 (3) の提案線形重回帰モデルには p 値の高い変数も含められており，また，変数の数も多く複雑である．そこで，冗長な変数を削除し，損害に本質的な変数に絞りインシデント発生時の損害額をより簡単に見積もることを試みる．本研究では，モデルの簡略化の指標として，赤池情報量基準（Akaike Information Criterion：AIC）を用いる [20]．AIC は，

$$\mathrm{AIC} = -2\log(L) + 2k$$

で定められ，値が小さいほうがよいモデルとされる．ここで，L は最大尤度，k はパラメータの数をそれぞれ示す．

提案線形重回帰モデルについて，AIC を最小化する係数で定められるモデルを，

$$\hat{y} = e^{\beta_0} \cdot x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot e^{\beta_8 \cdot x_8} \cdot e^{\beta_{11} \cdot x_{11}} \tag{6}$$

と定める．ここで，係数は**表 12** に示す．提案モデルと，簡略化したモデルの比較を**表 13** に示す．変数の数は 30 から 8 になり，すべての変数において p 値が有意水準を満たすようになっている．

たとえば，2014 年に医学生物学研究所で発生したインシデントの場合を考えよう，被害人数 $x_1 = 50$，売上高 $x_2 = 7{,}172$ [百万円]，業種 $x_8 = (0, 1, 0, 0)$（製造業），電話番号 $x_{11} = 0$ となり，漏洩損害額 y は，

$$\begin{aligned} y &= e^{\beta_0} \cdot x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot e^{\beta_8 \cdot x_8} \cdot e^{\beta_{11} \cdot x_{11}} \\ &= e^{-4.959} \cdot 50^{5.146 \times 10^{-8}} \cdot 7172^{1.005} \cdot e^{-0.529} \cdot e^{0 \times -0.467} \\ &= 31.01 \text{ [百万円]} \end{aligned}$$

である．同様に提案線形重回帰モデルで漏洩損害額を計算した場合，漏洩損害額は 37.97 百万円となる．実際の漏洩損害額は 38.25 [百万円] であり，簡略化モデルの誤差が大きくなっている．簡略化モデルでは表 13 より，平均の予想損害額が小さくなり，平均誤差は提案モデルと比べて大きくなったが，平均誤差率は同等となった．

表 12 簡略化モデルの係数
Table 12 Coefficient of simple model.

係数				*Estimate*	*p.value*	
β_0				-4.959	$8.08E-11$	$***$
log(被害人数)		$\log(x_1)$	β_1	$5.146E-8$	0.061	$*$
log(売上高)		$\log(x_2)$	β_2	1.005	$2E-16$	$***$
業種	建設業	x_8	β_8	-1.206	0.001	$***$
	製造業			-0.529	0.076	$*$
	金融業，保険業			-1.344	0.048	$**$
	運輸業，郵便業			-1.013	0.081	$*$
電話番号		x_{11}	β_{11}	-0.467	0.037	$**$

表 13 線形重回帰モデルと簡略化モデルの比較
Table 13 Comparison of AIC of our proposed model and simple model.

	線形重回帰モデル	簡略化モデル
AIC	122.011	85.758
パラメーター数	30	8
平均予想損害額	2,741.46	2,068.35
平均誤差	4,935.22	5,247.10
平均誤差率	1.73	1.73

5. 考察

5.1 評価結果について

JO モデルの算出額と本研究で使用したインシデントデータとの比較の結果，損害賠償額を含むか否かで損害額が大きく変化することが明らかになった．表 9 のベネッセホールディングスの例でいうと JO モデルでは想定損害賠償額の誤差率は 60.6 であったが，提案モデルでは 0.49 となり，誤差が非常に小さい．平均誤差率を見ても，JO モデルで 4.54，提案線形重回帰モデルで 1.73 である．JO モデルは損害賠償額を推定するものであり，本研究で対象とした被害額ではないが，個人情報漏洩インシデント発生による損害額が損害賠償請求にも影響すると考えられる．文献 [23] では，見舞い費用 500 円のいきさつが述べられており，お詫びよりもセキュリティ強化に使う方が顧客の利益につながることを提言している．被害額の変化は，これらを受けて，時代に応じて個人情報の価値も大きく変わっていることを示唆している．

また，算出額を被害人数で割った 1 人あたりの損害額は JO モデルでは 33,000 円なのに対して，提案モデルでは約 273 円であった．しかし，表 11 から平均の 1 人あたりの漏洩損害額をみると提案モデルは非常に高額である．これは提案モデルが企業の売上高に大きく依存しているため，売上高が大きく被害人数の少ないケースに適応できていないためであると考えられる．

5.2 精度の向上と維持

個人情報漏洩インシデントによる損害額の予想のためのモデル（係数）は，短期間に更新することで法改正などの社会情勢を反映することができ，精度の向上と維持ができると考えられる．一方，短期間に更新を繰り返すことは，調査のためなどのコストの増加につながる．また，過去の事例についての想定損害額を算出を試みる場合精度が低下し，損害額の経年変化を観察することが困難である．そのため，損害額を推定する場合には，最新のモデルだけでなく，時代に合った適当なものを選択することが必要である．

また，提案モデルでは，本人特定容易度や，精神的ランクといった一部の変数を JNSA が定めたものをそのまま利用している．先述したとおり，JO モデルでは，損害賠償額を含む損害額を算出することを目的としているため，提案モデルで本人特定容易度などを利用する場合，損害賠償額を含まないこと考慮してランクを変動させたり，新たな指標を追加することで，精度が向上が期待できる．

6. おわりに

本研究では，個人情報漏洩の損害額の新しい数理モデルの提案を行い，(1) インシデント発生企業の情報を説明変数として重回帰を適用，(2) 2010 年から 2016 年のインシデント情報を使用，(3) 特別損失額を目的変数として利用，を行うことで，先行研究よりも実際の損害額に近い金額を算出できるモデルを作成した．提案線形重回帰モデルの重み付き平均誤差率は 1.73 であった．ベネッセ社の 1 人あたりの損害額は 273 円であり，より現実的なモデルであることを示した．

しかし，東京電力ホールディングスの 2017 年の特別損失額，349,640.7 百万円などのように，目的変数とした特別損失額には震災などの影響が含まれている可能性がある．そのため，本稿で参照した情報セキュリティ対策費のような情報漏洩事件そのものだけに関連する損失額のデータを収集する方法を検討する必要がある．さらに，日本における訴訟率や賠償額などの情報を詳しく調査し，モデルを調整することで，損害賠償額も含めて，損失額を推測できるようにすること，企業ごとのマネジメント方策実施状況による，損失額の変化の調査や，モデルへの適用を行う予定である．

謝辞 本研究を遂行するにあたり，インシデントデータを提供いただいた日本ネットワークセキュリティ協会様に感謝いたします．

参考文献

[1] 山田道洋，菊池浩明，松山直樹，乾 孝治：個人情報漏洩の損害額の新しい数理モデルの提案，情報処理学会，CSEC 研究会，CSEC80, pp.1–7 (2018).
[2] ベネッセお客様本部：事故の概要，入手先 ⟨https://www.benesse.co.jp/customer/bcinfo/01.html⟩（参照 2017-01-31）.
[3] 幻冬舎：不正アクセスによる会員情報の流出に関するご報告とお詫び，入手先 ⟨http://www.gentosha.co.jp/news/n446.html⟩（参照 2017-01-31）.
[4] 日本ネットワークセキュリティ協会：2016 年情報セキュリティインシデントに関する調査報告書〜個人情報漏えい編〜，入手先 ⟨http://www.jnsa.org/result/incident/⟩（参照 2018-02-01）.
[5] 日本ネットワークセキュリティ協会セキュリティ被害調査ワーキンググループ：情報セキュリティインシデントに関する調査報告書別紙，第 1.0 版 (2017).
[6] 情報セキュリティインシデント調査報告書（JNSA データセット）.
[7] Romanosky, S.: Examining the costs and causes of cyber incidents, *Journal of Cybersecurity*, Vol.2, No.2, pp.121–135 (2016).
[8] 本決算（連結優先）データ，株式会社 QUICK Astra Manager，入手先 ⟨http://biz.quick.co.jp/lp_astram/⟩.
[9] 日経新聞：ベネッセ HD 最終赤字 136 億円 情報漏洩で特損 260 億円，入手先 ⟨https://www.nikkei.com/article/DGXLASGD31H1G_R30C14A7EA2000/⟩（参照 2018-02-05）.
[10] セキ株式会社：平成 28 年度 3 月期決算短信，p.2, 入手先 ⟨https://www.seki.co.jp/material/dl/ir/kessan/20160506_LdfbMJKUnbPG.pdf⟩（参照 2018-02-05）.
[11] CabinetOffice, Cyber insurance market: Joint government and industry statement (2014), available from ⟨https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/371036/Cyber_Insurance_Joint_Statement_5_November_2014.pdf⟩.
[12] Franke, U.: The cyber insurance market in sweden, *Computers & Security*, Vol.68, pp.130–144 (2017).
[13] Gordon, L.A. and Loeb, M.P.: The economics of information security investment, *ACM Trans. Inf. Syst. Secur.*, Vol.5, No.4, pp.438–457 (2002).
[14] Gordon, L.A., Loeb, M.P., Lucyshyn, W. and Zhou, L.: Increasing cybersecurity investments in private sector firms, *Journal of Cybersecurity*, Vol.1, No.1, pp.3–17 (2015).
[15] Romanosky, S., Hoffman, D. and Acquisti, A.: Empirical analysis of data breach litigation, *Journal of Empirical Legal Studies*, Vol.11, No.1, pp.74–104 (2014).
[16] Romanosky, S., Telang, R. and Acquisti, A.: Do data breach disclosure laws reduce identity theft?, *Journal of Policy Analysis and Management*, Vol.30, No.2, pp.256–286 (2011).
[17] Wells, A. and Jones, S.: Growth in cyber coverage expected as underwriting evolves (2016).
[18] Wheeler, J. and Akshay, L.: Understanding when and how to use cyberinsurance effectively, PE Proctor 2015 Technical Report (2015).
[19] Edwards, B., Hofmeyr, S. and Forrest, S.: Hype and heavy tails: A closer look at data breaches, *Journal of Cybersecurity*, Vol.2, No.1, pp.3–14 (2016).
[20] 金 明哲：R によるデータサイエンス，p.146, 森北出版 (2007).
[21] 櫻井直子：情報セキュリティの価値と評価，文眞堂 (2011).
[22] 石川朝久，櫻井幸一：個人情報漏洩補償に関する一検討，Computer Security Symposium 2014, pp.1185–1191 (2014).
[23] 日本ネットワークセキュリティ協会：内部不正対策 14 の論点，インプレス R&D (2015).

山田 道洋

2017 年明治大学総合数理学部先端メディアサイエンス学科卒業．2019 年明治大学大学院博士前期課程修了．現在，日本電気株式会社所属．

菊池 浩明 （正会員）

1988 年明治大学工学部電子通信工学科卒業．1990 年同大学大学院博士前期課程修了．1994 年同博士（工学）．1990 年（株）富士通研究所入社．1994 年東海大学工学部電気工学科助手．1995 年同専任講師．1999 年同助教授，2006 年同情報理工学部情報メディア学科教授．1997 年カーネギーメロン大学計算機科学学部客員研究員．2013 年明治大学総合数理学部先端メディアサイエンス学科教授．2016 年同大学院先端数理科学研究科長．WIDE プロジェクト暗号メールシステム FJPEM の開発，認証実用化実験協議会（ICAT），IPA 独創情報技術育成事業等に従事．暗号プロトコル，ネットワークセキュリティ，ファジィ論理，プライバシー保護データマイニング等に興味を持つ．1990 年日本ファジィ学会奨励賞，1993 年情報処理学会奨励賞，1996 年 SCIS 論文賞，2010 年度，2017 年度情報処理学会 JIP Outstanding Paper Award．2013 年 IEEE AINA Best Paper Award. 2014 年情報セキュリティ文化賞．電子情報通信学会，日本知能情報ファジィ学会，IEEE，ACM 各会員．本会フェロー．

松山 直樹

1981 年大阪大学理学部数学科卒業．博士（理学）．明治安田生命保険勤務を経て，2009 年明治大学理工学部教授．2013 年同大学総合数理学部教授．アクチュアリー数理と統合的リスク管理（ERM）の研究に従事．日本アクチュアリー会正会員（理事・学術委員長）．日本保険年金リスク学会（副会長）．日本数学会会員．

乾 考治

1987 年東京工業大学工学部化学工学科卒業．1997 年筑波大学経営・政策科学研究科経営システム科学専攻修了（経営学）．2001 年明治大学博士（理学）．1987 年日本生命入社．2002 年京都大学経済学研究科・寄附講座助教授．2004 年明治大学大学院グローバルビジネス研究科・専任助教授．2010 年明治大学大学院グローバルビジネス研究科・専任教授．2013 年明治大学総合数理学部・現象数理教授．金融取引のビッグデータ分析，企業経営指標としての資本コスト，その他現象数理科学研究に従事．日本保険年金リスク学会（大会担当理事，評議員），日本オペレーションズリサーチ学会，日本ファイナンス学会，日本価値創造 ERM 学会，日本統計学会各会員．