SCIS 2020 2020 Symposium on
Cryptography and Information Security
Kochi, Japan, Jan. 28 – 31, 2020
The Institute of Electronics,
Information and Communication Engineers

企業のサイバーインシデントの予測 ～あなたの会社は何年後にサイバーインシデントを受けるか？～

池上 和輝 * 菊池 浩明 †
Kazuki Ikegami Hiroaki Kikuchi

あらまし 今日企業は多くのサイバー攻撃を受けている．例えば，2018 年には 443 件の個人情報漏洩インシデントが発生し，約 561 万件の個人情報が漏洩した．この脅威に対して，企業は各種のセキュリティ対策を導入して，セキュリティリスクアセスメントを行なっている．しかしながら，認証取得にかかるコストとインシデント削減効果の関係は明らかではなく，信頼できるリスクの基準が求められている．

そこで，本研究では、JNSA が収集した 2005-2016 年のインシデント情報を用いて，企業規模や業種などの異なる組織のリスクを算出する確率モデルを提案する。提案モデルの特徴は，負の二項分布に基づき，与えられた任意の条件の企業におけるインシデントの発生確率を定式化することである．提案モデルに基づき，セキュリティ対策の異なる企業間でインシデント発生確率の比を示し，その原因について考察を与える．

キーワード 個人情報漏洩，負の二項分布

1 はじめに

1.1 背景

近年，不正アクセスや内部犯行といった悪意のある攻撃による個人情報漏洩が増加している．日本ネットワークセキュリティ協会（JNSA）の調査では，2018 年 443 件のインシデントが発生し，561 万件の個人情報が漏洩した [1]．インシデント 1 件あたりの平均想定損害賠償額は 6 億 3,767 万円に上り，前年と比較しても 9,000 万円の増加であった．

こうした背景から，経済産業省は「サイバーセキュリティ経営ガイドライン」[2] を 2015 年に策定した．サイバー攻撃から企業を守る観点で，経営者が認識する必要のある「3 原則」，及び経営者が情報セキュリティ対策を実施する上での責任者となる担当幹部（CISO 等）に指示すべき「重要 10 項目」をまとめている．その中にはサイバーセキュリティリスクの認識や，リスクに対応するための仕組み構築などが含まれている．

対策の一つとして，セキュリティ保険が日本でも 2015 年から大手保険会社で取り扱われ始めた．しかし，諸外国に比べてセキュリティ保険の加入率が 17.2 %（2017 年，ICDJapan 社調査）と低く，その理由として「情報漏洩の可能性を感じていない」,「費用対効果が見えない」が挙げられていた [3]．また，小梶らは金融機関においてサイバーセキュリティアセスメント手法だけでは，リスクの特定，発生確率，リスク評価ができないため，一般的なリスクアセスメントを組み合わせることが有効であると示している．[4]

1.2 関連研究・研究目的

リスク評価の先行研究として，Edwards らによる [5] がある．彼らは，PRC(Privvacy Right Clearinghouse) 公開データセットの 2005-2015 年データのうち 2,234 件のインシデントを使用して，アメリカの個人情報漏洩の傾向を調査した．彼らは，インシデントによる被害人数と発生頻度をモデル化するためにベイズ一般化線形モデル (Bayesian Generalized Linear Models) を用いた．悪意のある攻撃により発生した 1 インシデントの被害人数に対数正規分布，人的ミス等による被害人数に対数歪曲正規分布 (Log SkewNormal)，インシデントの発生頻度に負の二項分布を使用した．その結果，インシデントの被害人数と頻度をモデル化することで，2005-2015 年の間にインシデントの傾向が変化していないことを示した．また，提案モデルにより 95%予測区間の中でインシデントの発生を予測した．

山田らは，セキュリティマネジメント方策のによるインシデント生起確率への影響を分析した [6]．彼らは，業

* 明治大学大学院 先端数理科学研究科 Graduate School of Advanced Mathematical Sciences, Meiji University, cs192021@meiji.ac.jp
† 明治大学 総合数理学部 先端メディアサイエンス学科 Department of Frontier Media Science, School of Interdisciplinary Mathematical Sciences, Meiji University, kikn@meiji.ac.jp

種や企業規模といった交絡因子を除いたマネジメント効果を測るために，ロジスティック回帰を使用し．その結果，オッズ比から CIO 設置企業では，インシデントの生起確率が 0.3 倍に抑えられていることを明らかにした．

Maochao らは 2005-2017 年のサイバーハッキングインシデントの到着間隔と被害の大きさを確率過程によりモデル化した [7]．また，提案もモデルにより到着間隔と被害規模を予測できることを示した．

しかしながら，Edwards らの研究では，業種，企業規模，ISMS などのセキュリティ対策もまちまちの多くの組織の集合体について，インシデントの発生リスクをモデル化しており，その結果から特定の組織のリスクを測ることはできない．

そこで，本研究では組織ごとのインシデント発生リスク，マネジメントによる効果を定量的に示すことを目的とする．次の問いにより動機付けられている．

- ある組織が一年後にインシデントを起こす確率?
- ある組織が次にインシデントを起こすまでの日数?
- セキュリティマネジメント対策がインシデント発生を抑制す効果があるか？

上記を明らかにすために，過去の組織ごとのインシデント到着間隔を，負の二項分布でモデル化することで，リスクの定量化を試みる,，到着間隔を組織ごとのマネジメント情報を説明変数とした一般化線形モデルにかけることでマネジメント効果を定量化する．表 1 に関連研究 [5] と [6]，本研究の違いを整理する．先行研究 [5] では，マネジメントの効果やインシデントの傾向を全域的に調査していたため，組織ごとの過去のインシデント傾向やそれらを考慮したマネジメント効果が分析できていなかった．

2　使用するデータ

2.1　CSR データセット

東洋経済新報社は，毎年全上場企業と主要未上場企業に対して CSR(corporate social responsibility) に関するアンケート調査を行っている [8]．データベースは，従業員数，平均年齢等に関する「雇用・人材活用」，ISMS 取得の有無や CIO の設置に関する「CSR 全般・社会貢献・内部統制等」，CO2 排出量，環境保全コスト等に関する「環境」の三部から構成される．質問項目は多様な形式を含んでいる．例えば,「内部監査を行っているか」という質問に対し「1．定期的に行っている　2．不定期で行っている…」など複数の選択肢がある．本研究では，それらの質問の回答を Yes，No に分類し直し調査を行った．2017 年調査による CSR データの統計量を表 2 に示す．

本稿では，約 800 の質問項目のうち，情報セキュリティに関係する表 3 に示した 17 に絞り，調査結果を報告する．

3　分析手法

3.1　確率分布

3.1.1　負の二項分布

発生確率 p の事象（インシデント）が r 回発生するまでに，かかる日数の確率変数 X（到着間隔）は次式で定められる，が負の二項分布に従う

$$Pr(X = x) = \binom{x + r - 1}{x} p^r (1 - p)^x$$

負の二項分布の期待値，到着間隔 μ が以下の式で表される．

$$\mu = \frac{(1 - p)r}{p}$$

本研究では，最尤推定により当てはめを行い，パラメータを推定する．最尤推定は R の fitdistr 関数を使い計算した．

3.1.2　ポアソン分布

定数 λ>0 に対し，0 以上の整数を値にとる確率変数 X が

$$Pr(X = x) = \frac{\lambda^k e^{-\lambda}}{k!}$$

を満たすとき、確率変数 X は母数 λ のポアソン分布に従うという。

3.1.3　正規分布

平均 μ, 分散を $\sigma^2 > 0$ の確率変数 X が正規分布に従う時，確率密度関数は

$$Pr(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x - \mu)^2}{2\sigma^2})$$

で与えられる．

3.2　KS 検定

複数の確率分布が同一かどうかを判別する検定として KS 検定がよく知られている．到着間隔がある確率分布に従うかを確認するため，コルモゴロフ-スミルノフ検定 (Kolmogorov-Smirnov)[9] KS 検定は，データに基づいて求めた経験分布関数と，推測した分布関数との違いを検定する．d を未知な分布関数 $F(D)$ に従う，大きさ m の確率標本，標本が従う確率分布を $F_0(D)$ とする．帰無仮説 H_0 は

$$H_0 : F(x) = F_0(x)$$

とする．

表 1: 関連研究との比較

	本調査	山田 [6]	Edword [5]
目的	組織ごとのリスク評価	マネジメント効果の定量化	インシデント傾向調査
手法	負の二項分布	ロジスティック回帰	ベイズ一般化線形モデル
インシデントデータ	JNSA(2005-2018)	JNSA(2012-2016) Security Next	CSR(2005-2015)
分析対象	日本の 391 組織	日本の 17 業種の全企業，企業規模	全米の全企業

表 2: CSR データセットの統計量

期間	対象企業数	質問項目数	JNSA との共通企業
2017	1,414	840	223

表 3: 主な CSR 企業属性項目

項目 ID	質問項目	略称
C122	内部告発者の権利保護に関する規定制定	告発保護
C139	内部統制委員会の設置	内統委員
C147	ＣＩＯ（最高情報責任者）の有無	CIO
C150	ＣＦＯ（最高財務責任者）の有無	CFO
C161	プライバシー・ポリシーの制定	PP
C153	情報システムに関するセキュリティポリシー	SP
C155	情報システムのセキュリティに関する内部監査	内部監査
C157	情報システムのセキュリティに関する外部監査	外部監査
C159	ＩＳＭＳ（情報セキュリティマネジメントシステム）認証	ISMS
C120	内部告発窓口（社内）の設置	内部窓口
C202	内部告発窓口（社外）の設置	外部窓口
C207	業務部門から独立した内部監査部門の有無	独立監査
C227	リスクマネジメント・クライシスマネジメントの体制の構築	RM・CM
C229	リスクマネジメント・クライシスマネジメントの基本方針の有無	RM・CMP
E082	環境監査の実施状況	環境監査
E087	環境マネジメントシステムの構築	環境 M
K136	労働安全衛生マネジメントシステムの構築の有無	労働 M

経験分布関数 F_m は

$$F_m(D) = \begin{cases} 0 & D < d_1 \\ \dfrac{i}{m} & d_{(i)} \le D < d_{(i+1)}, \quad i = 1, \ldots, m-1 \\ 1 & D \ge d_{(m)} \end{cases}$$

と表す．m が大きくなるにつれて経験分布関数 $F_m(D)$ は真の分布に $F_0(D)$ に近づくので，経験分布 $F_m(D)$ は真の分布 $F_0(D)$ の推定量となる．

仮定された分布と $F_0(D)$ と経験分布 $F(D)$ の検定統計量 K_m は

$$K_m = \sup_D |F_m(D) - F_0(D)|$$

で与える．

3.3 一般化線形モデル

一般化線形モデル (Generalized Linear Models) は，正規分布を拡張した指数型分布族に対応させ，非線形の減少を線形モデルの場合と同じく簡単に扱えるように，かつ不自然な尺度で解釈しないように工夫したデータ解析手法である [10].

表 4: 1 組織当たりの 13 年間のインシデント数 (N=9,358)

平均	分散	最大	最小	合計
1.5	12	195	1	16,392

表 5: 1 組織によるインシデント数（13 年間）

インシデント数	組織数 n（割合）	インシデント数 B（割合）
1	7,972 (0.85)	7,972 (0.57)
2	757 (0.08)	1,514 (0.11)
3	238 (0.03)	714 (0.05)
4 以上	391 (0.04)	3,789 (0.27)
計	9,358	13,989

4 提案リスクモデル

4.1 JNSA データセット

JNSA セキュリティ被害調査ワーキンググループは，企業のプレスリリースやニュースサイトなどでの報道から個人情報漏洩インシデントを 2005 年から毎年収集している [1]．企業経営者がセキュリティ対策投資を行う際の参考を目的の一つとして，それらのインシデント情報を被害人数，漏洩原因 (紛失 ・置忘れ，不正アクセス，誤操作等)，漏洩経路（紙媒体，インターネット等）に分類し，評価を行っている．

JNSA データセットには，2005-2018 年の 16,392 インシデント，計 9,358 組織のデータが含まれる．1 組織におけるインシデント数 B の統計量を表 4 に示す．ここで，195 件のインシデントを起こした組織は大阪市であった．2005-2018 年のインシデント数の異なる組織の分布を図 1 と表 5 に示す．1 組織によるインシデント数は，日付の異なるインシデントのみを使用して集計したため，表 5 の総インシデント数は上記と異なる．13 年間でインシデントを 1 件だけ起こした組織は 7,972 で全体の 85%である．その一方で，全インシデントのうち 40%以上は，インシデントを 2 回以上起こした 15%(1,386) の組織が起こしたものであった．

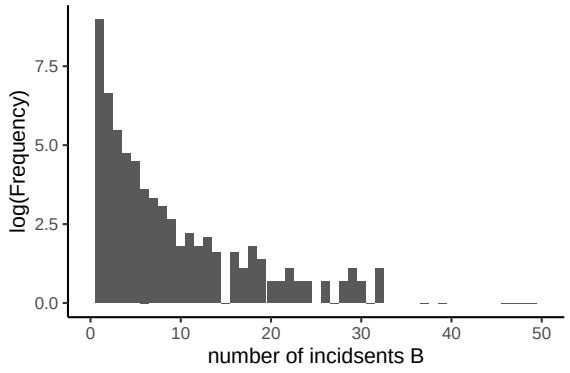


図 1: 1 組織当たりの 13 年間のインシデント数の分布

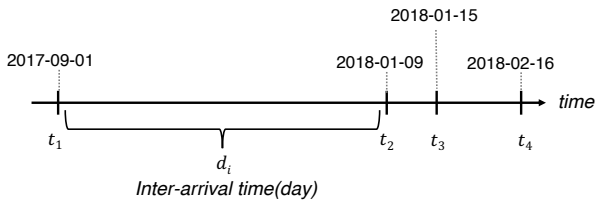


図 2: インシデント到着間隔

4.2 提案モデル

本稿では，ある組織 i にインシデントが起きてから次にインシデントが発生するまでの間隔をインシデント到着間隔（到着間隔 d_i）(日) とする．例えば，図 2 のようにある組織でインシデントが 4 回起きた時，インシデント到着間隔は $d_1 = 130, d_2 = 6, d_3 = 32$ となる．確率モデルで到着間隔を表現するとき，d_i が学習用に最低 2 つ，評価用に 1 つを使用する．図 2 の例では d_1, d_2 が学習用，d_3 が評価用である．この条件の下，インシデントを 4 回以上起こした 391 組織を最尤推定により確率分布へ当てはめる．

ある組織のインシデント発生間隔 $d_1, d_2, \ldots, d_m$ が与えられた時，その組織のインシデント発生間隔 D は次のモデルに従う

$$D \sim F_{NB}(\mu, r)$$

ただしここで，F_{NB} は負の二項分布に従って発生する（累積）確率分布であり，そのパラメータである平均到着間隔 μ は

$$\mu = \mathrm{e}^{\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_{19} x_{19}}$$

に従う．x_1 を業種，x_2 を従業員数，$x_3, \cdots, x_{19}$ を表 3 に示した各マネジメントとする．ここで，x_1 はダミー変数，従業員数は Log をとった値，マネジメントは bool 値とする．D に対して最尤推定により，μ, r を定める．推定には R の fitdistr を用いた．

表 6: 1 年後のインシデント発生確率 $Pr[D \leq 365]$
n=391

平均	最大	最小	標準偏差
0.11	1	0	0.27

4.3 負の二項分布への当てはめ結果

391 モデルのうち 4 つの例を図 3 に，推定したパラメータ r，μ の分布をそれぞれ図 4，5 に示す．図 3 で，黒線は経験分布関数を示しており，赤線が推定された負の二項分布である．

負の二項分布の解釈は，例えば，東京電力の図 3 から，$d = 365$，$\Pr[D \leq 365] = 0.74$ なので，365 日以内にインシデントを起こす確率が 75%であること表す．表 6 は $\Pr[D \leq 365]$ の統計量を表す．ここで $\Pr[D \leq 365]$ の平均算出時には，インシデントを起こしていない組織数を以下のようにする．

CSR(2017) 対象組織のうち，13 年間でインシデントを起こした組織は 17%(223/1351) であった．そこで，本モデルではインシデントを起こさなかった組織数 n は，インシデントを起こした n 組織に対して，

$$n' = n * 1351/223 = 1955$$

と仮定する．この時，ある企業にインシデントが生じる平均間隔 (day) の確率変数 D^* が，1 年以内となる確率は，

$$\begin{aligned} Pr[D^* < 365] &= E[Pr[D_1 < 365 OR D_0 < 365]] \\ &= E[Pr[D_1 < 365]Pr[Z = 1] \\ &\quad + Pr[D_0 < 365]Pr[Z = 0] \\ &= Pr[Z = 1]0.55 + Pr[Z = 0]0 \\ &= 0.11 \end{aligned}$$

と与えられる．ただし，ここで，Z を企業が 13 年間で 4 件以上インシデントが生じる (0,1) の確率変数, D_1 と D_0 を，4 件以上と未満の企業の平均間隔の確率変数とし，また，$Pr[D_1 < 365]$ は 391 企業の到着間隔が 1 年未満となる平均確率である．

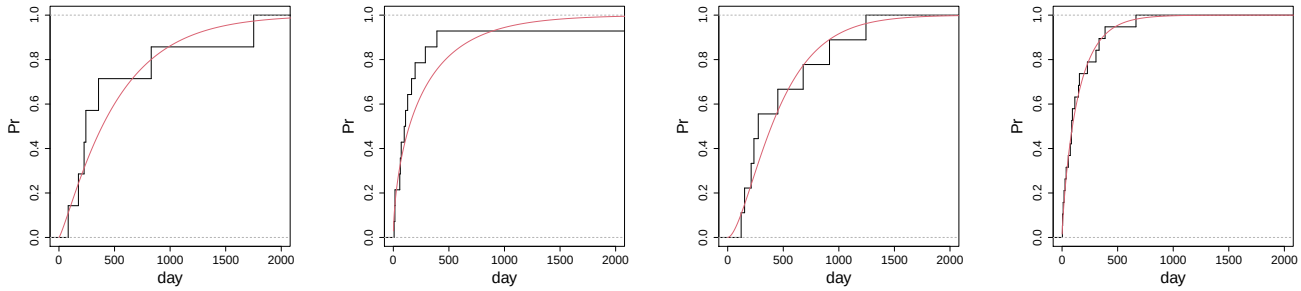
また，最尤推定による負の二項分布のパラメータ μ, r のうち，μ でソートした上位下位 3 件を表 7 に示す．パラメータの中央値は，それぞれ 1.07，257 であった．インシデント発生の平均発生間隔は，組織間で最大 64 倍変わる．

4.4 モデルの信頼性

表 8 は，平均到着間隔を負の二項分布 (nbinom)，ポアソン分布 (pois)，正規分布 (norm) にそれぞれ当てはめた時ときの KS 検定結果の一部の例を表 **??**し，その中で東京ガスに当てはめた結果を図 6 に示す．表 8 の例

図 3: 負の二項分布への当てはめ結果

りそな銀行　　東京電力　　三重県　　高島屋



r =1.21, μ = 523　　r = 0.56, μ = 284　　r = 1.8, μ =476　　r =0.97, μ =149

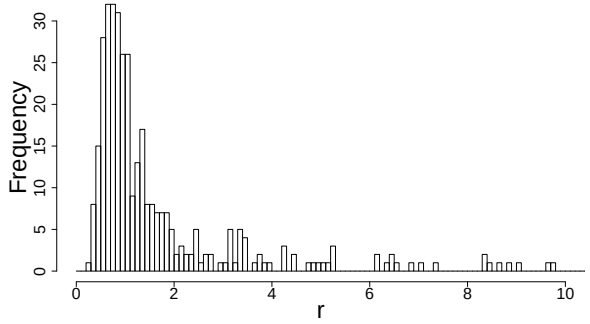


図 4: 負の二項分布のサイズ r の分布

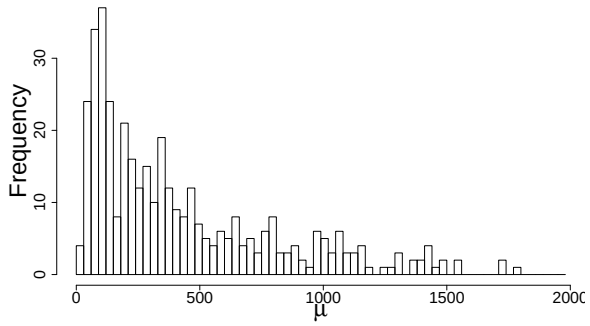


図 5: 負の二項分布の平均到着間隔 μ の分布

表 7: 最尤推定によるパラメータの推定

組織名	r	mu
三井不動産レジデンシャル株式会社	2,436	19
旭川信用金庫	18	20
大阪市	0.59	25
⋮	⋮	⋮
三重銀行	0.64	1719
北越銀行	0.3	1738
新潟大学	38	1789

表 8: KS 検定 (p-value) 結果の確率分布比較

確率分布	nbinom	poisson	norm
東京ガス	0.0963	0.0000	0.0008
NTT 西日本	0.0892	4.26E-14	0.0065
都市再生機構	0.088	8.82E-14	0.0004

表 9: 5%水準で帰無仮説を否定された組織数の割合

nbinom	poisson	norm
0.02 (9/391)	0.39 (155/391)	0.08 (31/391)

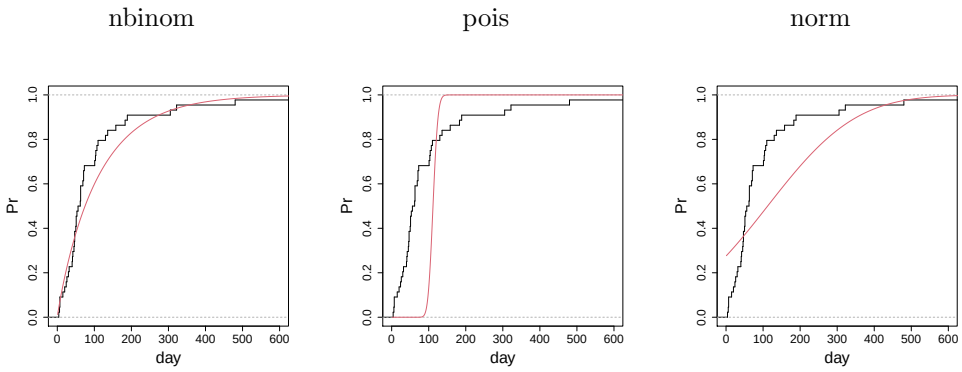
では，負の二項分布以外の確率分布は 5%水準で有意差が見られ，組織の到着間隔がそれぞれの確率分布従うことを否定された．同様に 391 モデルに当てはめたとき，5%水準で否定される組織数を表 **??**に示す．負の二項分布で帰無仮説を否定される組織数が全体の 2%で最も少なかった．

4.5　予測評価

本実験では，インシデント到着間隔を予測するために Pr=0.7 を閾値として，その時の d の値を予測到着間隔として評価を行った．評価の定義を図 8 のように予測到着間隔 $\hat{d}_i$ が，実際の到着間隔 d_i よりも長い場合は正しい ($\hat{d}_i \geq d_i : correct$)，逆の場合は誤り ($\hat{d}_i < d_i : miss$) とした．また，その時に発生する誤差 (error)$|\hat{d}_i - d_i|$ で定義する．はじめに，全体と業種ごとの Pr=0.7 の平均到着間隔と正解率（正解組織数/全組織数）を表 10 に示す．この時の平均は，インシデントを 4 件以上起こしている 391 組織による平均である．

391 組織での，予測到着間隔は 426 日後であり，正解率は 54%であった．また，組織数が二桁である業種においては，正解は最大で 64%(公務 (他に分類されるもの

図 6: 異なる確率分布による当てはめ (東京ガス)



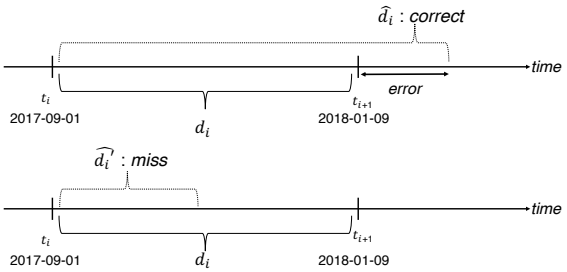


図 7: 予測評価の説明

表 10: 業種ごとの予測精度

業種名	平均到着間隔	正解率 (正解組織数 / 全組織数)
複合サービス事業	364	1.00 (2/2)
林業	177	1.00 (1/1)
公務 (他に分類されるものを除く)	439	0.64 (103/162)
情報通信業	641	0.61 (19/31)
教育，学習支援業	777	0.57 (20/35)
金融業，保険業	614	0.53 (31/58)
不動産業，物品賃貸業	244	0.50 (6/12)
建設業	297	0.50 (3/6)
製造業	349	0.50 (2/4)
卸売業，小売業	513	0.44 (4/9)
医療，福祉	341	0.39 (12/31)
電気・ガス・熱供給・水道業	507	0.35 (8/23)
運輸業，郵便業	388	0.33 (1/3)
サービス業 (他に分類されないもの)	394	0.17 (2/12)
生活関連サービス業，娯楽業	348	0.00 (0/2)
全体	426	0.55 (214/391)

を除く))，最低で 17%(サービス業 (他に分類されないもの)) であった．

表 11 に，組織ごとの予測到着間隔 $\hat{d}_i$（Pr=0.7）の上位下位 3 件を示す．予測到着間隔は，組織間で最大 1,938 日差があった．

次に，閾値別の誤差の大きさを図 8 で表した．60–90%の間で正解した組織の半数は，365 日以内の誤差であったが，1000 日以上の誤差がでる組織 (25 組織) も存在した．また閾値 70%の時が，365 日以内誤差で正しく予測できる組織数が最も多く 135 件であった..

表 11: 組織ごとの予測到着間隔 $\hat{d}_i$ （Pr=0.7）

組織名	$\hat{d}_i$
三井不動産レジデンシャル株式会社	23
旭川信用金庫	24
大阪市	28
⋮	⋮
松山市	1,767
新潟大学	1,930
三重銀行	1,961

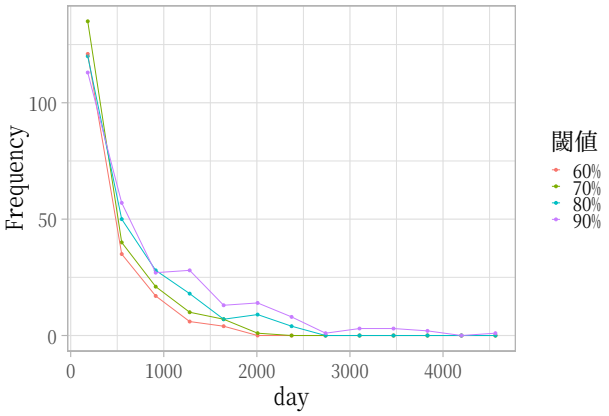


図 8: 誤差の分布

表 12: 一般化線形モデルの結果

	Estimate	Std. Error	P-value	
(Intercept)	8.96	0.12	< 2e−16	***
医薬品	−0.14	0.13	0.26	
運輸，物流	−0.07	0.12	0.55	
機械	−0.04	0.12	0.72	
金融 (除く銀行)	−0.31	0.13	0.01	*
銀行	−1.06	0.14	0.00	***
建設，資材	−0.37	0.12	0.00	**
自動車，輸送機	0.00	0.12	0.98	
商社，卸売	−0.12	0.12	0.31	
小売	−0.30	0.12	0.01	*
情報通信，サービスその他	−0.18	0.12	0.12	
食品	−0.02	0.12	0.87	
素材，化学	−0.05	0.12	0.66	
鉄鋼，非鉄	−0.03	0.13	0.84	
電機，精密	−0.08	0.12	0.49	
電気，ガス	−2.24	0.29	0.00	***
不動産	−0.46	0.13	0.00	***
LOG(従業員数)	−0.07	0.01	¡ 2e−16	***
ISMS	0.04	0.03	0.18	
CIO	−0.07	0.03	0.01	**
CFO	0.01	0.03	0.64	
外部窓口	0.01	0.02	0.70	
内部窓口	−0.07	0.05	0.14	
告発保護	0.06	0.05	0.24	
内統委員	−0.01	0.02	0.65	
PP	0.00	0.03	0.98	
SP	−0.01	0.04	0.79	
内部監査	0.01	0.03	0.75	
外部監査	−0.07	0.02	0.00	**
独立監査	0.02	0.04	0.61	
RM_CM	0.03	0.04	0.42	
RM_CMP	−0.08	0.04	0.03	*
環境監査	−0.03	0.04	0.33	
環境 M	0.10	0.03	0.01	**
労働 M	0.00	0.02	0.98	

4.6 セキュリティ対策の効果

マネジメント x_l を実施した時の到着間隔を μ_l^+，実施していない時の到着間隔 μ_l^- とすると，マネジメント x_l による効果を

$$\frac{\mu_l^+}{\mu_l^-} = \frac{e^{\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_l x_l}}{e^{\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_{l-1} x_{l-1}}} = e^{\beta_l}$$

と表す．一般化線形モデル glm による各係数を表 4.6 に示す．本実験では，インシデントを起こしていない組織と 1 件のみしかインシデントを起こしていない組織の影響を加味するために，インシデントを起こしていない企業の到着間隔 d_1 を JNSA の観測期間である 13 年間の 4,745 日，インシデントが 1 件だけの場合は，観測した到着間隔の最大である $d_1 = 4,270$ 日を仮定した．Estimate は係数であり，これが正の場合，業種に当てはまるとき，該当マネジメントを行っているときに推定到着間隔 $\hat{d}$ が長くなる．逆に Estimate が負の場合は，推定到着間隔 $\hat{d}$ を短くする．例えば外部監査を設置することで，インシデント到着間隔が 0.9 倍になることがわかる．調査した 17 マネジメントのうち，9 つのマネジメントにインシデント到着間隔を長くする効果があった．

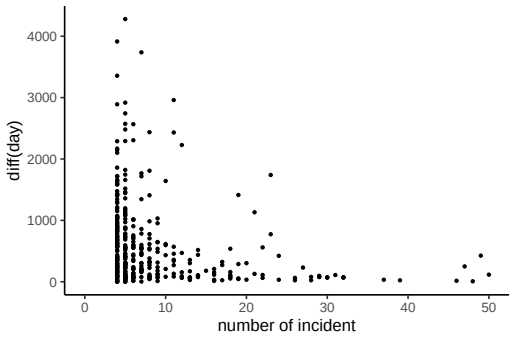


図 9: モデルに用いたインシデント数誤差の分布

5 考察

本調査では，2005-2018 年にインシデント ($B \geq 4$) を 4 件以上起こした組織を対象として，負の二項分布によるモデル化を行った．図 9 に，学習と評価に用いたインシデント数と，70%時の予測到着間隔と実際の到着間隔との誤差の分布を示す．学習に用いるインシデントが少ない場合は，誤差が大きくなりやすく最大で 4.279 日の誤差であったが，最小で 1 日の組織も存在したことから，組織によっては周期的な到着間隔が存在すると主張する．

また，負の二項分布のパラメータ μ と r の分布を図 10 に示す．パラメータのばらつきから，組織によって到着間隔の振る舞いが異なることがわかる．

また，Pr=0.7 時の業種ごとの平均到着間隔で，電気・ガス業種が上位から 5 番目に長い結果であったのに対して glm の結果では，到着間隔を短くする効果に有意差があった．これはインシデントが発生していない組織の到着時間間隔が表 10 では，考慮しなかったことが原因だと考えられる．例えば CSR データセットで電気・ガス業種のうち JNSA(2005-2018) に含まれる組織は 83%(10/12) であったが，運輸業，郵便業では 6%(4/64) であった．運輸業は，表 10 では到着間隔の業種平均が 388 日であったが，これは偏った一部の組織の傾向であったことがわかる．したがって，表 4.6 で到着間隔が長くなる業種であっても組織ごとにみると到着間隔が短くリスクが高い組織が存在することが表 10 から示される．

6 結論

本研究目的の問いに対して，次に示す．

- ある組織が一年後にインシデントを起こす確率は，インシデントを起こしていな組織を CSR データセットを使用して近似した場合，全体平均は **0.11** である．
- ある組織が次にインシデントを起こすまでの日数は，Pr=0.7 の時の到着間隔をインシデント発生と

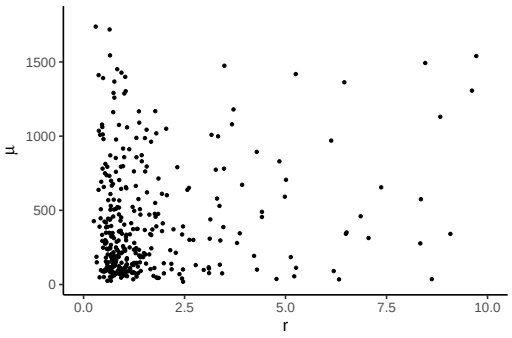


図 10: 負の二項分布パラメータ μ と r の分布

したとき 391 組織の最短は **23** 日，391 組織の平均は，**426** 日である．

- 情報セキュリティマネジメントがインシデント発生間隔に及ぼす効果は，外部監査を設置することで，インシデント到着間隔が **0.9** 倍に短くする．

しかし，モデルに用いたインシデント数が組織数によって異なるため，予測精度にもばらつきがあり，マネジメント方策情報が 2017 年の調査結果を 13 年間のインシデントデータに当てはめているなどの問題点がある．そこで今後は，各組織ごとのインシデントを増加させ，マネジメントの経年変化の考慮したモデルを検討する．

謝辞

本研究に有益なご助言を賜りました．明治大学の乾孝治教授，松山直樹教授に深く感謝申し上げます．本研究を遂行するにあたり，インシデントデータを提供して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致します．

参考文献

[1] 日本ネットワークセキュリティ協会， 2018 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～(速報版).

[2] 経済産業省，”サイバーセキュリティ経営ガイドライン Ver2.0”，（meti.go.jp/press/2017/11/20171116003/20171116003-1.pdf，2019.12.12 参照）.

[3] 佐久間樹里，猪俣敦夫，サイバー保険の調査・分析による加入率向上への提案，研究報告インターネットと運用技術 (IOT)(IPSJ)，pp. 1-8，2019

[4] 小梶 顯義，原田 要之助，後藤 厚宏，金融機関におけるサイバーセキュリティのアセスメントに関する考察，研究報告電子化知的財産・社会基盤 (IPSJ)，pp. 1-5，2019

[5] B. Edwards, S. Hofmeyr, and S. Forrest, Hype and heavy tails: A closer look at data breaches, Journal of Cybersecurity, 2(1):3–14, 2016.

[6] 山田道洋, 池上和輝, 菊池浩明, 乾考治, 経営マネジメント状況による情報漏洩インシデント削減効果の評価 (2), Computer Security Symposium 2018(IPSJ), pp. 376-384, 2018.

[7] M Xu, K Schweitzer, R Bateman, S.Xu , Modeling and Predicting Cyber Hacking Breaches, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, pp.2856-2871, 2018.

[8] 東洋経済データベース，CSR データ

[9] 村上秀俊,「ノンパラメトリック法」，朝倉書店，2015，pp. 23-26

[10] 金明哲,「R によるデータサイエンス」，森北出版，2017，pp. 153-156