# 線形モデルにおける安全な予測値公開メカニズムの提案とその疾患リスク予測モデルへの適用

草野 光亮[1]　竹内 一郎[2]　佐久間 淳[3]

**概要**：遺伝情報から疾患リスクを求めるサービスなど，個人情報を入力値として予測値を提供するサービスが始まりつつある．このようなサービスにおいて，ユーザーが予測値を公開する・第三者が予測値を盗み見るなどによって第三者が予測値を得た場合，予測値から個人情報である入力値を推定される可能性がある．本研究では，線形回帰モデルにおいて予測値から入力が推定されるリスクを定式化する．また，このリスクを抑えつつ有用性を保つような出力を与えるメカニズムを設計する．さらに，個人ゲノムからの疾患リスク予測の線形回帰モデルを例に取り，提案メカニズムはプライバシー保護と有用性において既存手法より優れていることを示す．

**キーワード**：PWS, プライバシー, 個人情報, リスク評価, ゲノム

## 1. はじめに

個人に関する情報の利用が進み，個人に関する情報を入力値として用い個人に特化した予測を行なうサービスも広がりつつある．例として，個人の嗜好に特化した広告を表示させるサービスや患者の遺伝情報から薬剤の投与量を予測するアプリケーションなどが存在する．

個人に関する情報を入力値として統計モデルに基づき予測を行なうサービスについて，ユーザーが予測値を公開する・第三者が予測値を盗み見るなどし，第三者が予測値を得た場合，予測値から入力値を推定されるリスクが存在する．予測値から入力値が容易に推定可能であるならば，その予測値は入力値と同等のセンシティブな情報として扱う必要がある．

例えば，遺伝情報から疾患罹患リスクを予測するサービスがあったとする．ある個人において，サービスの予測した疾患罹患リスクから入力値である遺伝情報が容易に推定できたとすると，この疾患罹患リスクは遺伝情報と同等のセンシティブな情報になりうる．一方，容易に推定できないのであれば入力値と同等にセンシティブな情報として扱う必要は必ずしもない．このように予測値をどの程度のセンシティブな情報として扱うかを判断するために，予測値から入力値がどの程度推定されるか評価する必要がある．

1　筑波大学 大学院 システム情報工学研究科
2　名古屋工業大学・情報工学専攻
3　筑波大学 大学院 システム情報工学研究科 / JST CREST

先の例に例えれば，疾患罹患リスクを web で掲載するサービスなど予測値が不特定多数に見られることを考慮しなければならないケースが存在する．このような場合には入力値が推定されるリスク値の上界が定められ，その制約下で予測器を構築する手法が求められる．

本稿においては，特に疾患罹患リスク予測モデルについて取り上げる．次節で疾患罹患リスク予測モデルに対する入力値の推定リスク評価に関する研究を紹介する．

### 1.1 関連研究

入力値の推定リスクに関しての議論は個別化医療において多く議論されている [1][3][6]．これらを紹介するために，遺伝的特徴の一つである SNP を説明する．

遺伝情報は染色体の特定の場所に記述されており，この場所を遺伝子座と呼ぶ．染色体は DNA から成り立ち，DNA は 4 種の塩基の配列により表現される．ある遺伝子座において，塩基の配列の組み合わせを遺伝子型と呼ぶ．長さが 1 の遺伝子座で，一定以上の頻度で多様性がみられるものを一塩基多型 (SNP) と呼ぶ．ある SNP における遺伝子型の分布は研究目的のため一般に公開されている．

SNP の中には疾患罹患リスクと関連がある SNP が存在すると言われている．SNP において特定の塩基はアレルと呼ばれ，疾患などに大きく関連のあるアレルはリスクアレルと呼ばれる．人間の SNP は 2 つの塩基により表現されるためリスクアレル数は 0,1,2 の 3 値をとる．これらの遺伝情報を予測器の入力値として符号化する方法として，

SNP が特定の遺伝型であるか否かの二値を入力値とする方法や SNP におけるリスクアレル数を入力値とする方法が存在する．これらの遺伝情報を用い，個人の疾患罹患リスクを予測する [3][6].

ある個人の SNP が特定の遺伝型であるという情報はセンシティブな情報になりうる．なぜならば，ある個人の SNP の遺伝型を第三者が推定できた場合，その SNP に関連のある疾患のリスクを算出できる可能性があるからである．例えば，糖尿病に関する疾患罹患リスク値から推定された SNP 情報が認知症のリスク値にも関係しているとする．その SNP と糖尿病に関する疾患罹患リスク値は高いセンシティビティを持つこととなる．ある SNP と疾患の関係は将来新たに明らかになる可能性があるため，現状センシティビティの高くない SNP が今後もそうであるとは限らない．そのため，個人がどの遺伝型を持つかという情報は一般的に秘密情報として扱われる．

Ayday ら [1] は，リスクアレル数を入力とし疾患罹患リスクを計算する暗号プロトコル上において，計算結果が第三者に漏洩しているという設定下で入力値がどの程度推定可能かを議論した．入力値が推定されるリスクが存在することを示し，また連鎖不均衡 (LD) と呼ばれる SNP 間の相関を用いることで推定リスクが更に高くなることを明らかにした．5.1 節で紹介する値域を分割した区間に変換する手法により入力値の推定が困難になることも示した．

荒井ら [6] は，リスクアレル数を入力とし疾患罹患リスクを出力するゲノム検査モデルに対し，疾患罹患リスクから入力値が推定されるリスクを評価した．また，予測値である疾患罹患リスクに対し丸め処理を行い，予測値の丸め値を公開するという設定下において，入力値の推定リスクと丸め処理の関連について調べた．丸め値から入力値を推定する問題は，整数計画問題として定式化できることを示した．また，丸め処理により漏洩を軽減できることを示し，丸め幅と漏洩がトレードオフであると結論付けた．

これらの研究の目的は，入力値推定リスクの評価が目的であった．また，入力値推定リスクを軽減する手法は丸め処理など簡便な処理であり最適性が必ずしも保証されておらず，出力の有用性についても十分な配慮されているとは言えない．

### 1.2 貢献

本稿では，入力値が離散集合である線形回帰モデルを，有限離散集合を入力とする写像としてモデル化し，有限離散集合を入力とする写像に対し $\alpha$-obscure 安全性を提案する．$\alpha$-obscure 安全性とは，予測値が得られた際の入力値の条件付き確率と入力値の事前確率の差が高々 $\alpha$ であるという安全性である．さらに，$\alpha$-obscure 安全性を保証する写像を設計することが，$\alpha$-obscure 安全性に対応する制約を満たす集合に入力値ドメインを分割することと等価であることを示す．実数を出力する予測器を元に，精度の意味で最適な区間を出力する予測器を $\alpha$-obscure 安全性の上限を保証しつつ設計するアルゴリズムを提案する．更に，実用されている疾患罹患リスク予測モデルに対し，我々の提案したアルゴリズムを適用し実験を行った．そこで，入力値推定リスクと有用性はトレードオフであることを確認し，提案手法は既存手法と比べプライバシー及び有用性の両方において優れていることを示した．

## 2. 問題設定

入力値推定問題を定義する．$\mathbb{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_d$ とし，$\mathbb{X}$ および $\mathbb{Y}$ は有限離散集合である．$\mathbb{X} \to \mathbb{Y}$ である関数集合を $\mathbb{F}$ とし，$f \in \mathbb{F}$ とする．入力 $\mathbf{x} \in \mathbb{X}$ に対し，出力を $y = f(\mathbf{x})$ とする．$\mathbf{x}$ の $i$ 番目の属性値を $x_i \in \mathcal{X}_i$ とする．

入力値推定問題には，公開者と攻撃者の 2 つのステークホルダーが存在する．公開者は，秘密情報である入力値 $\mathbf{x} \in \mathbb{X}$ を持ち，出力値 $y = f(\mathbf{x})$ を評価し $y$ を公開する．

攻撃者は，$f$ と入力 $\mathbf{x}$ の事前分布 $\Pr[\mathbf{X}]$ を事前知識として持つ．$\mathbf{X}$, $X_i$, $Y$ はそれぞれ $\mathbf{x}$, $x_i$, $y$ の確率変数である．攻撃者は任意の $\mathbf{x} \in \mathbb{X}$ について $f(\mathbf{x})$ を任意回評価可能であり，無限の計算能力をもつ．入力値推定問題における攻撃者の目的は，$y$ が与えられた際に任意の $i$ について未知の $x_i$ を推定すること，つまり条件付き分布 (事後分布)$\Pr[X_i|Y = y]$ を推定することである．ここで，$f$ は単射であるとは限らず，$y = f(\mathbf{x})$ となる $\mathbf{x}$ が複数存在しうるため，攻撃者が $y = f(\mathbf{x})$ となる $\mathbf{x}$ を一つ求めただけでは入力値推定としては不十分であることに注意されたい．

### 2.1 事後分布 $\Pr[X_i|Y = y]$ の評価

本節では攻撃者が事前分布 $\Pr[\mathbf{X}]$ と $f$ から事後分布 $\Pr[X_i|Y = y]$ を評価する方法を議論する．

$f$ は $\mathbb{X}$ から $\mathbb{Y}$ への写像であり，逆像が存在する．$f$ における $\mathcal{Y} \subseteq \mathbb{Y}$ の逆像 $f^{-1}[\mathcal{Y}]$ とは，

$$f^{-1}[\mathcal{Y}] = \{\mathbf{x} \in \mathbb{X} | f(\mathbf{x}) \in \mathcal{Y}\}$$

となる集合のことで，$f^{-1}[\mathcal{Y}]$ は常に $\mathbb{X}$ の部分集合となる．逆像は逆写像と異なり任意の写像に存在することに注意されたい．ここで，$f^{-1}[\{y\}]$ と $y$ が 1 対 1 対応である．よって，$\Pr[X_i|Y = y]$ は $\Pr\left[X_i|\mathbf{X} \in f^{-1}[\{y\}]\right]$ と同値となる．これにより $x_i$ の事後分布は式 1 の評価となる．

$$\begin{aligned}
&\Pr[X_i = a|Y = y] \\
&= \Pr\left[X_i = a|\mathbf{X} \in f^{-1}[\{y\}]\right] \\
&= \frac{\sum_{\mathbf{x} \in \{\mathbf{x}|\mathbf{x} \in f^{-1}[\{y\}], x_i = a\}} \Pr[\mathbf{X} = \mathbf{x}]}{\sum_{\mathbf{x} \in f^{-1}[\{y\}]} \Pr[\mathbf{X} = \mathbf{x}]} \qquad (1)
\end{aligned}$$

網羅的列挙により $f^{-1}[\{y\}]$ を評価するには $|\mathbb{X}|$ 回の $f(\mathbf{x})$ の評価が必要となり $|\mathbb{X}|$ は一般的に $d$ について指数オーダーである．しかし，攻撃者は $f(\mathbf{x})$ を任意回評価可能と

しているためこの評価は可能であることに注意されたい．

事後確率 $\Pr[X_i = a|Y = y]$ が 1 となる場合は，$y$ から入力値属性の値が $a$ であると一意に定まるということである．特に $f$ が単射であるならば，入力値 $\mathbf{x}$ と予測値 $y$ は一対一関係となる．このため，逆像 $f^{-1}[\{y\}] = \{\mathbf{x}\}$ となり，事後確率が常に 1 もしくは 0 となることがわかる．

## 3. 安全性指標

本章では，予測値の公開に関する入力値の推定リスクを評価する $\alpha$-obscure 安全性を定義する．この定義では攻撃者の得た事後確率が事前確率とそれほど異ならないならば，攻撃者は属性推定に成功していない，という考えを元にしている．

**定義 1.** $i$ 番目の属性 $x_i$ の事前分布を $\Pr[X_i]$ とし $y = f(\mathbf{x})$ とする．すべての $k \in \mathcal{X}_i$ とすべての $\mathbf{x} \in \mathbb{X}$ について，以下が成り立つとき，写像 $f$ は $i$ 番目の属性に対し $\alpha$-obscure 安全である．

$$|\Pr[X_i = k|Y = y] - \Pr[X_i = k]| \leq \alpha.$$

攻撃者の得た事後確率が常に事前確率と一致するとき $\alpha = 0$ であり，推定リスクが最も低いといえる．反対に，事後確率が事前確率から離れていれば離れているほど，$\alpha$ が大きくなり，推定リスクは高いと考えられる．

注意すべきこととして，この $\alpha$-obscure 安全性は属性ごとに決まるということである．すべての属性の推定リスクに配慮するためには，$f$ に対し $i$ 番目の属性についての $\alpha_i$-obscure 安全性が成り立つとすると $\alpha_1, \cdots, \alpha_d$ の $d$ 個の安全性指標に配慮する必要がある．

任意の $f$ について，属性 $i$ に関する $\alpha$-obscure 安全性の $\alpha$ の上下限は事前確率を用いて以下のようになる．

$$0 \leq \alpha \leq \max_{a \in \mathcal{X}_i} \max\{\Pr[X_i = a], 1 - \Pr[X_i = a]\} \quad (2)$$

$f$ が単射である場合つまり事後確率が 1 もしくは 0 となる場合が最も推定リスクが高いといえ，そのような $f$ においては $\alpha$ の値は上限と一致する．一方，どのような入力値でも同一の予測値を出力するような $f$ は，最も推定リスクが小さいといえ，この $f$ において $\alpha$ は 0 となる．

$\alpha$-obscure 安全性の $\alpha$ の値を属性間で比較し，数値的に入力値の推定のしやすさを議論することは適切でない．また，$\alpha$-obscure 安全性の $\alpha$ の値を見る際に事前分布も参照すべきである．なぜならば，同一の事後分布であっても事前分布が異なるならば $\alpha$-obscure 安全性の $\alpha$ の値が異なるためである．例えれば $f$ が単射であるならば，すべての属性について $f$ の $\alpha$-obscure 安全性は式 2 の上限に一致する．$j$ 番目の属性に関して $\mathcal{X}_j = \{0, 1\}$ とし $\Pr[X_j = 0] = 0.1$ とした際の $\alpha$-obscure 安全性の $\alpha$ の上限は 0.9 となる．同様に，$k$ 番目の属性に関して $\mathcal{X}_k = \{0, 1\}$ とし $\Pr[X_k = 0] = 0.5$ とした際の $\alpha$-obscure 安全性の $\alpha$ の上限は 0.5 となる．$f$ は単射であるため予測値が得られた際に，$j$ 番目の属性も $j$ 番目の属性も同等に一意に特定できてしまうにもかかわらず，0.9 と 0.5 のような異なる値を示すが，属性 $j$ が属性 $k$ の 1.8 倍推定リスクが高いとは言い難い．このように $\alpha$-obscure 安全性の $\alpha$ の値を属性間で比較することは適切ではなく，同時に $\alpha$-obscure 安全性の $\alpha$ の値を見る際に事前分布も参照すべきである．

次章でこの安全性指標を用い疾患罹患リスク予測モデルに対する入力推定リスクの評価を行なう．

## 4. 疾患罹患リスク予測モデルに対する入力値推定リスクの評価

本稿で用いる疾患罹患リスク予測モデルは遺伝情報を入力値として用いており，1.1 節のような入力値の推定リスクが存在する．本章では疾患罹患リスク予測モデルに対し，入力値推定リスク指標 $\alpha$-obscure 安全性の評価を行う．

### 4.1 疾患罹患リスク予測モデル

まず, 疾患罹患リスク予測モデルを導入する．$\mathbf{x} \in \{0, 1\}^d$ を遺伝情報と，$\mathbf{x}_c \in \mathbb{R}^{d'}$ を背景情報とする．$\mathbf{x}$ の $i$ 番目の属性 $x_i$ は，SNP$i$ について特定の遺伝型を持つか否かの遺伝的特徴を表す二値の情報である．$\mathbf{x}_c$ は高血圧であるかや喫煙者であるかなどの背景情報である．$\mathbf{x}, \mathbf{x}_c$ からロジスティック回帰により疾患罹患リスク値 $r \in \mathbb{R}$ を予測する．モデルのパラメータは $\mathbf{w} \in \mathbb{R}^d, \mathbf{w}_c \in \mathbb{R}^{d'}$ である．

$$r = \sigma\left(\mathbf{w}^\intercal \mathbf{x} + \mathbf{w}_c^\intercal \mathbf{x}_c\right)$$

次に，疾患罹患リスク予測モデルにおける入力値推定問題を考える．公開者は，$\mathbf{x}$ と $\mathbf{x}_c$ から $r$ を算出し $r$ を公開するが，背景情報 $\mathbf{x}_c$ は標的の周囲の人間に知られやすい情報である．そのため，ここでは $\mathbf{x}$ を秘密情報とし $\mathbf{x}_c$ を公開情報として扱うこととする．また，人間の SNP における遺伝型の分布は研究目的で公開されているため，攻撃者は $\Pr[\mathbf{X}]$ を用いることができる．さらに，$\mathbf{w}, \mathbf{w}_c$ は公開されており，攻撃者は利用可能であることを仮定する．以上より，攻撃者は $\Pr[\mathbf{X}], \mathbf{x}_c, \mathbf{w}, \mathbf{w}_c, r$ を知識として持つ．

このとき攻撃者は，すべての属性 $i$ について $x_i$ の事後分布を求めたいものとする．ここで $\mathbf{w}^\intercal \mathbf{x}$ は $\mathbf{w}^\intercal \mathbf{x} = \sigma^{-1}(r) - \mathbf{w}_c^\intercal \mathbf{x}_c$ であることから既知の情報から評価でき，この問題は $\sigma^{-1}(r) - \mathbf{w}_c^\intercal \mathbf{x}_c$ から $\mathbf{x}$ を求める問題に帰着する．つまり，このモデルにおける入力値推定問題は

$$g(\mathbf{x}) = \mathbf{w}^\intercal \mathbf{x} \quad (3)$$

とすると，攻撃者は $\Pr\left[X_i \middle| \mathbf{X} \in g^{-1}\left[\left\{\sigma^{-1}(r) - \mathbf{w}_c^\intercal \mathbf{x}_c\right\}\right]\right]$ を推定する問題となる．よって，この疾患罹患リスク予測モデルの入力値推定リスクを評価する際は，$g$ に対し入力値推定リスク評価を行えばよい．

以降の実験では 3600 から 6500 人の日本人被験者を対象に収集した SNP のプロファイル情報と 10 種の背景情報 [*1] における疫学調査に基づいて構築された，肥満及び脳出血の発症・罹患リスク (以降，罹患リスク) を評価するモデルを用いて入力値推定リスクを評価する．肥満の罹患リスクモデルは 10 個の SNP を，脳出血は 13 個の SNP を予測に用いた．高次元の予測モデリングでは，まず，marginal screening によって変数を絞込み [2]，選ばれた変数のみを用いてロジスティック回帰のような予測モデルを作成することが多い [4][5]．今回用いたリスクモデルも，同様の方法でロジスティック回帰モデルを用いている．前述よりロジスティック回帰を用いた疾患罹患リスク予測モデルに対する入力値推定リスクを評価することは，$g$ に対する入力値推定リスク評価と同等であることを述べた．本稿では，線形回帰関数 $g$ に対し入力値推定リスクの評価を行なう．

### 4.2 実値公開における推定リスク評価

まず，疾患罹患リスク推定モデルの予測値をそのまま公開した場合の入力値推定リスクの評価を行なう．ここでは，実際の肥満リスクを推定するモデルを用い，入力値推定リスク値として $\alpha$-obscure 安全性を用いる．このモデルは 10 個の SNP から肥満リスクを推定しており $d = 10$ であった．また，$\mathbf{w}$ の有効数字は 10 桁であった．

$i$ 番目の属性に関する $\alpha$-obscure 安全性を評価するためには，式 1 に従い $\Pr[X_i|Y = y]$ を評価する必要がある．この評価には $f^{-1}[\{y\}]$ についての和を必要とする．この評価のために，すべての $\mathbf{x} \in \mathbb{X}$ について $f$ の関係 $(\mathbf{x}, y)$ を評価し，この関係からすべての $y$ について $f^{-1}[\{y\}]$ を得た．

同様に式 1 の評価には入力値分布 $\Pr[\mathbf{X}]$ を必要とする．これを評価するには入力値の分布推定を行う必要があり，限られた数のサンプルから直接 $\Pr[\mathbf{X}]$ を求めることは困難である．そのため本稿では入力値属性間は独立であると仮定し入力値分布を近似する [*2]．

$$\Pr[\mathbf{X} = \mathbf{x}] = \prod_{1 \le i \le d} \Pr[X_i = x_i] \qquad (4)$$

疾患罹患リスク推定モデルに対し式 4 の近似において式 1 を評価した結果，すべての $i, \mathbf{x} \in \mathbb{X}, a \in \mathcal{X}_i$ において，事後確率 $\Pr[X_i = a|Y = y]$ は 0,1 のどちらかになった．つまり，この実験設定において $g$ は単射となり，このモデルにおいて予測結果をそのまま公開することは入力値をそのまま公開することと常に等価であることがわかった．

[*1] 年齢，性別，BMI，喫煙歴，血中クレアチニン濃度，糖尿病・高トリグリセリド血症・低 HDL コレステロール血症・高 LDL コレステロール血症・高血圧症の既往歴

[*2] 入力値属性間が独立でない場合も本稿における議論は成り立つ．連鎖不均衡など入力値属性の相関 $\Pr[X_i, X_j]$ を用い $\Pr[\mathbf{X}]$ を推定し我々の手法を利用することも可能である．

## 5. 等分割メカニズム

入力値推定リスクの高い予測モデルを加工し，入力値を推定されにくい予測モデルを設計する必要がある．ここでは線形回帰による予測値を公開する代わりに，予測値を含む区間を公開することにより，推定リスクが軽減された予測モデルを設計することを考える．

線形回帰モデル $g : \mathbb{X} \to \mathbb{R}$ が与えられ，区間を出力するメカニズムを $\mathcal{M} : \mathbb{X} \to \mathbb{I}$ とする．$\mathbb{I}$ は連続な実数区間集合である．メカニズムに対する入力値推定リスク $\alpha$-obscure 安全性を評価する際は，定義 1 における $f$ を $\mathcal{M}$ として評価することとなる．

### 5.1 等分割メカニズムの定式化

本節では，予測値の区間化に関する先行研究を紹介し，既存手法を用いた等分割メカニズムを定義する．

Ayday らは，multiplicative model と呼ばれる疾患罹患リスク予測モデルに対し区間化処理を適用し，入力値推定リスクを評価した [1]．同様に荒井らも multiplicative model に対し類似の区間化処理を適用し，入力値推定リスクを評価した [6]．両者は実質的には等価であり，本稿では Ayday らの区間化処理を取り上げる．

Ayday らの区間化処理は，区間 $[t_{\min}, t_{\max}]$ を分割数 $n$ で等分割し実数値 $t \in [t_{\min}, t_{\max}]$ を含む区間を出力する．等分割した区間集合 $\mathbf{I}_n$ は $k = 1, \cdots, n$ を用いて

$$\mathbf{I}_n = \begin{cases} \left[\frac{(n-k+1)t_{\min}+(k-1)t_{\max}}{n}, \frac{(n-k)t_{\min}+kt_{\max}}{n}\right) \\ \qquad\qquad (1 \le k < n) \\ \left[\frac{t_{\min}+(n-1)t_{\max}}{n}, t_{\max}\right] \qquad (k = n) \end{cases}$$

のように表せ，区間化関数 $\mathrm{round}_n : [t_{\min}, t_{\max}] \to \mathbb{I}$ は，

$$\mathrm{round}_n(t) = I \text{ where } t \in I, I \in \mathbf{I}_n$$

と定義される．この区間化関数は分割数が大きいほど区間が小さくなり，より正確な予測値を出力できる．

例えば $[t_{\min}, t_{\max}] = [0, 1], n = 4$ の時，区間化関数は

$$\left[0, \frac{1}{4}\right), \left[\frac{1}{4}, \frac{2}{4}\right), \left[\frac{2}{4}, \frac{3}{4}\right), \left[\frac{3}{4}, 1\right]$$

の 4 通りの出力のうち入力を含む区間を出力する．0.3 を与えられると $\left[\frac{1}{4}, \frac{2}{4}\right)$ を出力する．別の例で $[t_{\min}, t_{\max}] = [0, 1], n = 1$ の時は，どの入力値に対しても区間 $[0, 1]$ を返す関数となる．

Ayday らの区間化処理を用い，等分割メカニズム $\mathcal{M}_{\mathrm{round}_n} : \mathbb{X} \to \mathbb{I}$ を式 5 のように定義する．

$$\mathcal{M}_{\mathrm{round}_n}(\mathbf{x}) = \mathrm{round}_n \circ g(\mathbf{x}) \qquad (5)$$

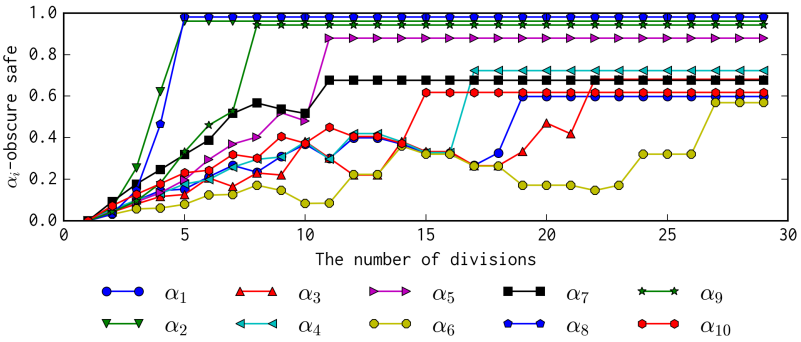



**図 1** 疾患罹患リスク予測モデルに対応した等分割メカニズムを用い，実数値の代わりに実数区間を公開した際の入力値推定リスク ($\alpha$-obscure 安全性) を表している．縦軸が $\alpha$-obscure 安全性を示し値が大きいほど入力値推定リスクが高い．横軸は分割数を示し分割数が大きいほど有用性が高い．

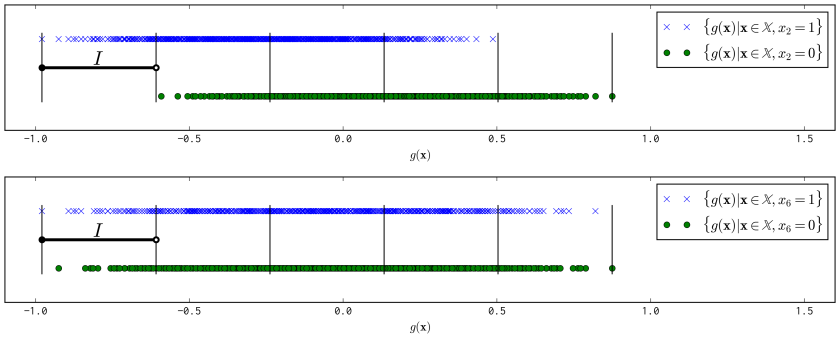



**図 2** 線形回帰における係数 $|w_i|$ の大きさと推定リスクの関係を示すために用いた図である．係数 $|w_i|$ が大きい場合，線形回帰による出力が入力値属性に大きく影響を受けるため入力値推定が容易であることを示している．上図が $|w_i|$ が $\mathbf{w}$ の中で最も大きかった 2 番目の属性に注目し，下図は $|w_i|$ が最も小さかった 6 番目の属性に注目している．

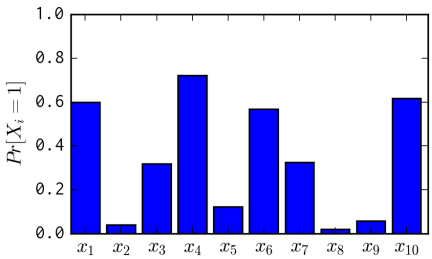



**図 3** 本稿で使用した疾患罹患リスク予測モデルにおける入力値属性の事前分布

### 5.2 等分割メカニズムの推定リスク評価

本節では，4.2 節で用いた疾患罹患リスク予測モデル $g$ を使用し等分割メカニズム $\mathcal{M}_{\mathrm{round}_n}$ の $\alpha$-obscure 安全性を評価する．$\alpha$ の評価方法は 4.2 節と同じである．

分割数 $n$ を変化させながら $i$ 番目の SNP について $\alpha_i$-obscure 安全性を評価した結果を図 1 に示す．

分割数が 1 である場合，$\alpha_i$ は常に 0 となる．これはどの入力値でも同一の出力値 $y$ を出力するため，$f^{-1}[\{y\}]=\mathbb{X}$ となり，事後分布が事前分布と一致するためである．分割数を増やすにつれ，$\alpha$-obscure 安全性の $\alpha$ が単調ではないが増加しているのがわかる [*3]．

分割数が大きければより正確な出力値公開ができるため，入力値推定リスクが分割数について単調ではないが増加しているということは，有用性と安全性にトレードオフの関係があることを示している．$\alpha$ の値に対し上限があり，属性により異なることがわかる．これは式 2 にて解析した通り，$\alpha$ の上限は事前分布に依存しているためである．事前分布を示した図 3 を参照すると，$2,8,9$ 番目の属性の事前確率 $\Pr[X_i=1]$ が 0 に近く，図 1 において $\alpha$ の上限が 1 に近いことがわかる．一方で $1,6,10$ 番目の属性の事前確率は 0.5 に近いため，$\alpha$ の上限が 0.5 に近いことがわかる．

分割数を増やした際の $\alpha$ の上昇の仕方について注目する．図 1 において $\alpha_2,\alpha_8$ から先に $\alpha_i$ が上限に到達し，上限への到達が最も遅いのは $\alpha_6$ であることがわかる．ここで $\mathbf{w}$ の

[*3] 単調でない理由は $\{f(\mathbf{x})|\mathbf{x}\in\mathbb{X}\}$ が離散値を取ることと，分割数 $n$ が自然数であり，2 以上の整数 $n$ について任意の $I\in\mathbf{I}_n$ と $I'\in\mathbf{I}_{n+1}$ において $I'\subset I$ が成り立たないことに起因する．分割数を $2^k(k$ は自然数) とすると，$I\in\mathbf{I}_{2^k}$ と $I'\in\mathbf{I}_{2^{k+1}}$ において $I'\subset I$ が常に成り立ち，$\alpha$ の値は単調となる．

要素の絶対値は2, 8番目の属性が最も大きく$w_2 = -0.38$で$w_8 = 0.38$であり，6番目の属性が最も小さく$w_6 = -0.05$であった．$\alpha$の上昇のしやすさは$w_i$の絶対値に依存していると考えられ，このことについて分析を行う．

図2は線形回帰における係数$w_i$の大きさと推定リスクの関係を示しており，注目した属性$i$に関して入力ドメインを$x_i = 0$と$x_i = 1$となる2つの集合に分割しそれぞれにおける$g$の像を示している．縦線は分割数5により分割される線を示している．注目した属性は，$\mathbf{w}$の要素の中で最も大きかった2番目の属性(上図)と最も小さかった6番目の属性(下図)の2つである．上図中の$I$に注目すると区間に含まれる要素は$x_2 = 1$となる$\mathbf{x}$のみからなり，上図の左端の集合$\mathcal{S} = \{\mathbf{x} \in \mathbb{X} | f(\mathbf{x}) \in I\}$の事後確率$\Pr[x_2 = 1 | \mathbf{X} \in \mathcal{S}] = 1$となる．下図の区間$I$において6番目の属性が1となる$\mathbf{x}$も0となる$\mathbf{x}$も両方混在しており，$\Pr[x_6 = 1 | \mathbf{X} \in \mathcal{S}]$は1にも0にもならないことがわかる．以上の解析から$w_i$の絶対値の大きさが大きい属性は小さい属性と比べ，少ない分割数で事後確率が1ないし0となると推測される．このため，$w_i$の絶対値の大きさが大きい属性から$\alpha$-obscure安全性の$\alpha$が上昇したと考えられる．

図1において，分割数5で2, 8番目の属性における$\alpha$-obscure安全性が上限に達したということは，分割数5において入力属性値が一意に特定できる入力値が存在することを示している．よってこのモデルにおいて，これらの属性に関しどのような入力値であっても一意に特定できないようにするには，分割数を5未満にしなければならない．分割数5とは出力が5通りしかないということを意味し，本来の予測器$g$の出力は$|\mathbb{Y}| = 2^{10}$通りであることと比べると低い精度の予測値しか公開できないことがわかる．

## 6. 有用性の最適性を保証するメカニズム

実際のサービスにおいては，許容できる入力値推定リスクの中で最も精度の高い出力をする予測器を用いたいと考えられる．そのため，公開者は許容できる入力値推定リスク$\{\alpha_i\}_{1 \le i \le d}$を定め，$i$番目の属性に関して$\alpha_i$-obscure安全を満たすことを制約とし，そのとき精度の意味において有用性を最大とするメカニズム$\mathcal{M} : \mathbb{X} \to \mathbb{I}$を設計するべきである．本章では，メカニズムの有用性や満たすべき性質・制約を議論し，メカニズム設計を最適化問題へ帰着させ，これを解くアルゴリズムを提案する．

### 6.1 最適化問題への定式化

まず，メカニズムの有用性を定める．メカニズムの出力する区間は，線形回帰の出力である実数値を含む．そのため，メカニズムの出力する区間の長さが短いほうが，線形回帰の出力により近く有用であると考えられる．このため，メカニズムの有用性を式6のように定める．この有用性が大きければより有用であるといえる．$|\cdot|$は区間長を示す．

$$\text{utility}(\mathcal{M}) = -\sum_{\mathbf{x} \in \mathbb{X}} |\mathcal{M}(\mathbf{x})| \tag{6}$$

次に，メカニズムの満たすべき制約について議論する．メカニズムが異なる区間を出力するならばその区間は重ならないほうが実用上有用である．たとえば，$g(\mathbf{x}_1) = 0.1, g(\mathbf{x}_2) = 0.2, g(\mathbf{x}_2) = 0.2$だとし，$\mathbf{x}_1, \mathbf{x}_3$のメカニズムの出力が$\mathcal{M}(\mathbf{x}_1) = \mathcal{M}(\mathbf{x}_3) = [0.1, 0.2]$だと仮定する．このとき，$\mathbf{x}_2$のメカニズム出力は$\mathcal{M}(\mathbf{x}_2) = [0.1, 0.2]$であるほうが有用である．

以上から，安全性を満たしつつ有用性を最大となるメカニズムを求める最適化問題Aが考えられる．$\mathbb{M}$は$\mathbb{X} \to \mathbb{I}$となるメカニズムの集合である．

$$\begin{aligned} \underset{\mathcal{M} \in \mathbb{M}}{\text{maximize}} \quad & \text{utility}(\mathcal{M}) \qquad \text{(A)} \\ \text{subject to} \quad & \forall_{1 \le i \le d}, \mathcal{M} \text{ is } \alpha_i\text{-obscure-safe} \\ & \forall_{\mathbf{x}, \mathbf{x}' \in \mathbb{X}, \mathcal{M}(\mathbf{x}) \neq \mathcal{M}(\mathbf{x}')}, \mathcal{M}(\mathbf{x}) \cap \mathcal{M}(\mathbf{x}') = \emptyset \\ & \forall_{\mathbf{x} \in \mathbb{X}}, g(\mathbf{x}) \in \mathcal{M}(\mathbf{x}). \end{aligned}$$

1行目の制約は安全性の上限を，2行目の制約は異なる出力は区間が重ならないことを，3行目の制約はメカニズムの出力区間に線形回帰の出力が含まれることを保証する．

この問題における解空間が$\mathbb{M}$となっており，写像の集合となっているため，このままでは解くことができない．解空間をより具体的にするために，$\alpha$-obscure安全である写像の満たすべき性質について述べる．

準備のために$\alpha$-obscure集合を定義する．

**定義 2.** $\mathbb{X}$の部分集合$\mathcal{S}$が，すべての$a \in \mathcal{X}_i$について以下が成り立つとき，$\mathcal{S}$を$i$番目の属性についての$\alpha$-obscure集合と呼ぶ．

$$|\Pr[X_i = a | \mathbf{X} \in \mathcal{S}] - \Pr[X_i = a]| \le \alpha.$$

次に$\mathbb{X}$の分割$\pi$を用い$\phi_\pi : \mathbb{X} \to \pi$を定義する．

$$\phi_\pi(\mathbf{x}) = \mathcal{S} \text{ where } \mathbf{x} \in \mathcal{S}, \mathcal{S} \in \pi$$

このとき，$\phi_\pi$と$\alpha$-obscure集合を用い次のことがいえる．

**定理 1.** $\pi$を$\mathbb{X}$の分割とする．すべての$\pi$の要素が$i$番目の属性について$\alpha$-obscure集合であるとき，かつそのときに限り，$\phi_\pi$は$i$番目の属性について$\alpha$-obscure安全である．

*Proof.* $\mathbb{X}$の分割$\pi$，任意の$\mathcal{S} \in \pi$について$\phi_\pi$の逆像は

$$\phi_\pi^{-1}[\{\mathcal{S}\}] = \mathcal{S}$$

となるのは明らかである．よって，

$\phi_\pi$ is $\alpha$-obscure-safe with respect to $i$-th attribute.

$\Leftrightarrow \forall_{a \in \mathcal{X}_i} \forall_{\mathbf{x} \in \mathbb{X}}, |\Pr[X_i = a | Y = \phi_\pi(\mathbf{x})] - \Pr[X_i = a]| \le \alpha$

$\Leftrightarrow \forall_{a \in \mathcal{X}_i} \forall_{\mathcal{S} \in \pi}, \left| \Pr\left[X_i = a \middle| \mathbf{X} \in \phi_\pi^{-1}[\{\mathcal{S}\}]\right] - \Pr[X_i = a] \right| \le \alpha$

$\Leftrightarrow \forall_{a \in \mathcal{X}_i} \forall_{\mathcal{S} \in \pi}, |\Pr[X_i = a | \mathbf{X} \in \mathcal{S}] - \Pr[X_i = a]| \leq \alpha$

$\Leftrightarrow \forall_{\mathcal{S} \in \pi}, \mathcal{S}$ is $\alpha$-obscure-set with respect to $i$-th attribute. □

よって，$i$ 番目の属性について $\alpha$-obscure 安全を満たす写像 $\phi_\pi$ を設計するには，すべての要素が $i$ 番目の属性について $\alpha$-obscure 集合となる $\mathbb{X}$ の分割 $\pi$ を求めればよい．

しかしながら $\phi_\pi$ の出力は区間ではなく $\mathbb{X}$ の部分集合であるためメカニズムではない．ここで $\mathbb{X}$ の部分集合 $\mathcal{S}$ を区間に変換するために，$\psi : \mathcal{P}(\mathbb{X}) \to \mathbb{I}$ を以下のように定義する．$\mathcal{P}(\mathbb{X})$ は $\mathbb{X}$ 上の集合族である．

$$\psi(\mathcal{S}) = \left[ \min_{\mathbf{x} \in \mathcal{S}} g(\mathbf{x}), \max_{\mathbf{x} \in \mathcal{S}} g(\mathbf{x}) \right]$$

$\mathbb{X}$ の部分集合 $\mathcal{S}$ における $g$ の像 $\{g(\mathbf{x}) | \mathbf{x} \in \mathcal{S}\}$ は実数集合となる．この集合の要素をすべて含む連続な実数区間のうち，$\psi$ は常に最も短い区間を出力するのは明らかであり，写像 $\psi$ は $\mathcal{P}(\mathbb{X}) \to \mathbb{I}$ の中で最も有用である区間を出力しているといえる．

これまでに定義した $\phi_\pi$ と $\psi$ を用い，$\mathbb{X}$ の分割 $\pi$ により定まるメカニズム $\mathcal{M}_\pi$ は以下のように書ける．

$$\mathcal{M}_\pi(\mathbf{x}) = \psi \circ \phi_\pi(\mathbf{x})$$

以上により最適化問題 A を最適化問題 B に変換することができ，安全性を保証しつつ有用性を最大化する問題は $\mathbb{X}$ の分割問題に帰着された．$\Pi$ は $\mathbb{X}$ の取りうる分割である．

$$\begin{aligned} &\underset{\pi \in \Pi}{\text{maximize}} && \text{utility}(\mathcal{M}_\pi) && \text{(B)} \\ &\text{subject to} && \forall_i \forall_{\mathcal{S} \in \pi}, \mathcal{S} \text{ is } \alpha_i\text{-obscure set} \\ & && \forall_{\mathcal{S}, \mathcal{S}' \in \pi, \mathcal{S} \neq \mathcal{S}'}, \psi(\mathcal{S}) \cap \psi(\mathcal{S}') = \emptyset. \end{aligned}$$

### 6.2 分割最適化アルゴリズム

問題 B を解く分割最適化アルゴリズム (Alg.1) を提案する．$\mathbb{T}$ を $\mathbb{T} = \{g(\mathbf{x}) | \mathbf{x} \in \mathbb{X}\}$ とし，$t_i$ は $\mathbb{T}$ の中で $i$ 番目に大きい値である．このアルゴリズムは，入力次元 $d$，線形回帰関数 $g$，公開者の定める安全性基準 $\{\alpha_i\}_{1 \leq i \leq d}$ を入力とし，$\mathbb{X}$ の最適な分割 $\pi^*$ を出力する．このアルゴリズムは，問題 B が $\mathbb{T}$ の分割問題に帰着されることを利用している．アルゴリズムの 8 行目により問題 B の 1 行目の制約が保証され，$\mathbb{T}$ の列の分割と解くことにより問題 B の 2 行目の制約が保証される．また，このアルゴリズムは動的計画法を用いており，$m$ が dynamic programming table である．$\pi$ は $\mathbb{X}$ の部分集合における分割，$\sigma$ は $\mathbb{T}$ の部分集合における分割を示している．$\{t_j\}_{1 \leq j \leq i}$ において安全性を満たす分割が存在するならば，$m_\sigma[i]$ に最適な分割を格納し $m_{\text{safe}}[i]$ に true を格納する．$i$ を 1 つずつ増加させながら，最終的には $\mathbb{T}$ における最適な分割 $\sigma^*$ を探索する．最後に $\sigma^*$ の要素に対し $g$ の逆像を評価し $\mathbb{X}$ における最適な分割を出力する．このアルゴリズムのオーダーは $O\left(|\mathbb{T}|^2(|\mathbb{T}| + d)\right)$ である．$|\mathbb{T}|$ は一般に入力次元 $d$ に対し指数オーダーであるため，このアルゴリズムは $d$ に対し指数オーダーとなる．

**Algorithm 1** 分割最適化アルゴリズム

**Require:** 入力次元数 $d$, 線形回帰関数 $g$, 安全性基準 $\{\alpha_i\}_{1 \leq i \leq d}$
**Ensure:** $\mathbb{X}$ の分割 $\pi^*$

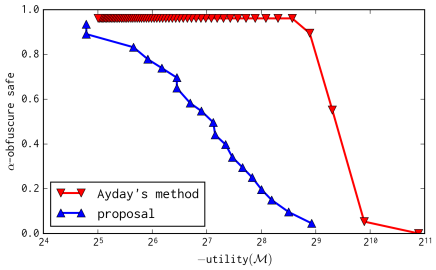
1: $m_\sigma[0] \leftarrow \emptyset$
2: $m_{\text{safe}}[0] \leftarrow$ true
3: **for** $i : 1, \cdots, |\mathbb{T}|$ **do**
4: $\quad \sigma_i^* \leftarrow \emptyset$
5: $\quad$ **for** $j : 0, \cdots, i-1$ **do**
6: $\quad\quad T_{i,j} \leftarrow \{t_k\}_{j+1 \leq k \leq i}$
7: $\quad\quad S_{i,j} \leftarrow g^{-1}[T_{i,j}]$
8: $\quad\quad$ **if** $\forall_{1 \leq k \leq d}, S_{i,j}$ is $\alpha_k$-obscure set on $k$-th attribute $\wedge m_{\text{safe}}[j]$ **then**
9: $\quad\quad\quad \sigma_{i,j} \leftarrow \{T_{i,j}\} \cup m_\sigma[j]$
10: $\quad\quad\quad \pi_{i,j} \leftarrow \{g^{-1}[T] | T \in \sigma_{i,j}\}$
11: $\quad\quad\quad \pi_i^* \leftarrow \{g^{-1}[T] | T \in \sigma_i^*\}$
12: $\quad\quad\quad \mathrm{u}_i^* \leftarrow \sum_{\mathcal{S} \in \pi_i^*} -|\mathcal{S}||\psi(\mathcal{S})|$ ▷ partial utility
13: $\quad\quad\quad \mathrm{u}_{i,j} \leftarrow \sum_{\mathcal{S} \in \pi_{i,j}} -|\mathcal{S}||\psi(\mathcal{S})|$
14: $\quad\quad\quad$ **if** $\sigma_i^* = \emptyset \vee \mathrm{u}_i^* < \mathrm{u}_{i,j}$ **then**
15: $\quad\quad\quad\quad \sigma_i^* \leftarrow \sigma_{i,j}$
16: $\quad\quad\quad$ **end if**
17: $\quad\quad$ **end if**
18: $\quad$ **end for**
19: $\quad m_\sigma[i] \leftarrow \sigma_i^*$
20: **end for**
21: $\pi^* \leftarrow \{g^{-1}[T] | T \in m_\sigma[\,|\mathbb{T}|\,]\}$

## 7. メカニズム比較実験

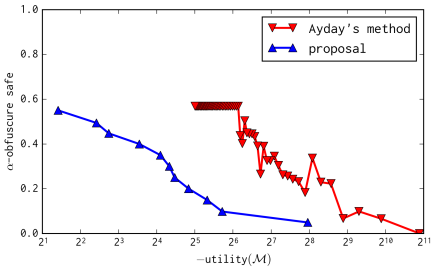
提案法と従来の等分割メカニズムと比較する．有用性と推定リスクの 2 つの指標を最適化するため，多目的関数最適化となる．5.2 節で注目した肥満リスク予測モデルの 2 と 6 番目の属性に注目し実験を行う．また，13SNP を入力値とする脳出血リスク予測モデルに対しても同様に実験を行う．$\alpha$-obscure 安全性の評価方法は 5.2 節と同じである．有用性指標は，提案法・既存手法共に式 6 を使用した．

提案法を用いメカニズムを設計する際に，$\alpha_i$ を与える必要がある．注目した属性 $i$ のみ $\alpha_i$ を 0 から 1 まで 0.05 ずつ変化させ，その他の属性 $j$ の $\alpha_j$ は 1 とした．

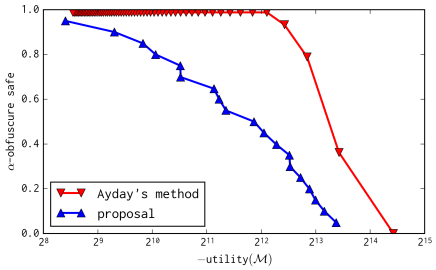
結果を図 4 に示す．提案法により設計されたメカニズムは従来法に比べ，有用性・プライバシーの両方において優れていることがわかる．同一の $\alpha$-obscure 安全性を保証する提案法のメカニズムと従来法のメカニズムにおいて有用性を比較すると，提案法は従来法と比べ双方の属性において少なくとも 2 倍程度の有用性を持っていることがわかる．これは有用性が式 6 により評価されることから，提案法により設計されたメカニズムは従来法に比べ出力される区間の長さが平均的に半分となったことを示している．



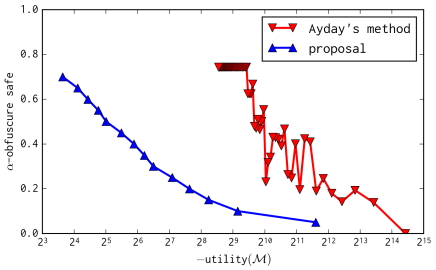

(a) 肥満リスク予測モデルの 2 番目の入力値属性における入力値推定リスクと有用性




(b) 肥満リスク予測モデルの 6 番目の入力値属性における入力値推定リスクと有用性




(c) 脳出血リスク予測モデルの 12 番目の入力値属性における入力値推定リスクと有用性




(d) 脳出血リスク予測モデルの 7 番目の入力値属性における入力値推定リスクと有用性

**図 4** 肥満 (10SNP) と脳出血 (13SNP) の疾患罹患リスク予測モデルに対し，等間隔メカニズム (Ayday's method) と提案メカニズム (proposal) を適用し，メカニズムの有用性と入力値の特定の属性に関する推定リスクを示している．等分割メカニズムにおいては分割数を，提案手法においては保証する $\alpha$-obscure 安全性の上界を変化させている．横軸が負の有用性であり，原点に近づくほど有用性が高い．縦軸は入力値推定リスクを示しており，原点に近づくほど入力値推定リスクが低い．

## 8. 結論

本稿では有限離散集合を入力値とする予測器に対し入力値推定問題を定義した．この問題において，入力値推定リスク $\alpha$-obscure 安全性を提案した．実際に用いられている疾患罹患リスク予測モデルに対し $\alpha$-obscure 安全性を評価し線形回帰による結果をそのまま公開することは入力値をそのまま公開することと等価であることを示した．既に提案されている入力値推定リスクを軽減する手法をこのモデルに適用し，十分な安全性を確保すると有用性は実用に耐えないことを示した．その後，入力値推定リスク $\alpha$ の上界が与えられている設定において，$\alpha$-obscure 安全性を保証する予測器を設計することが入力値ドメインの制約付き集合分割問題に帰着することを示した．制約付き集合分割問題を求解するアルゴリズムを提案し，疾患罹患リスク予測モデルに適用し提案法が有用であることを示した．

今後の発展として，提案アルゴリズムの改善がある．提案アルゴリズムは入力次元に対し指数オーダーであり，入力次元が大きい場合実用的な時間で解を求めることができない．そのためアルゴリズムの改善が課題である．

## 謝辞

本研究は，JST CREST「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」領域におけるプロジェクトおよび科学研究費 24680015,16H02864 の助成を受けました．

## 参考文献

[1] Erman Ayday, Jean Louis Raisaro, Jean-Pierre Hubaux, and Jacques Rougemont. Protecting and evaluating genomic privacy in medical tests and personalized medicine. In *Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society*, WPES '13, pp. 95–106, New York, NY, USA, 2013. ACM.

[2] Jianqing Fan and Jinchi Lv. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 70, No. 5, pp. 849–911, 2008.

[3] Matthew Fredrikson, Eric Lantz, Somesh Jha, Simon Lin, David Page, and Thomas Ristenpart. Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing. In *23rd USENIX Security Symposium (USENIX Security 14)*, pp. 17–32, San Diego, CA, August 2014. USENIX Association.

[4] David W Hosmer Jr and Stanley Lemeshow. *Applied logistic regression*. John Wiley & Sons, 2004.

[5] Yoshiji Yamada. Identification of genetic factors and development of genetic risk diagnosis systems for cardiovascular diseases and stroke. *Circulation Journal*, Vol. 70, No. 10, pp. 1240–1248, 2006.

[6] 荒井ひろみ, 津田宏治, 佐久間淳. ゲノム検査結果の開示によるプライバシ侵害の評価. コンピュータセキュリティシンポジウム 2015 論文集, 第 2015 巻, pp. 1258–1265, oct 2015.