C10　プライバシーを保護した疫学調査のための確率検定プロトコル

発表者 1BDRM015 佐藤 智貴
指導教員 菊池 浩明 教授

Privacy-Preserving Hypothesis Test Protocol for Epidemiology

Abstract: This paper proposes some new computationally efficient schemes of privacy-preserving epidemiological analysis. The performance and security of the proposed scheme are evaluated based on a trial implementation.

1　はじめに

環境因子と疾病の因果関係を明らかにする疫学調査では，疾病の原因と考えられる因子と疾病の関係性を統計的に解析する．例えば，喫煙者のがん罹患率と非喫煙者のがん罹患率を比較する事で，喫煙による健康への影響を明らかにすることができる．しかし，これらのデータは独立した組織によって管理されている事が多く，プライバシーと個人情報保護の観点から，照合が困難であった．

そこで，本研究では統計的調査の手法に暗号プロトコルを適用し，プライバシーを保護したまま安全に確率検定を行うことを試みる．本研究では，次の 3 つの方式を提案する．(1) 放射線疫学調査，(2) 相対危険度の検定，(3) 傾向性の検定.

2　要素技術

2.1　患者-対照調査

患者-対照調査とは，ある特定の疾病の患者群と非患者群について，ある要因に暴露していたか否かを調べ，因果関係を研究する調査である．表 1 の患者-対照調査のデータが与えられた時,

$$RR = \frac{a}{n_1} / \frac{c}{n_2} = \frac{a(c+d)}{(a+b)c} \approx \frac{ad}{bc} \qquad (1)$$

を相対危険度 (relative risk) と呼ぶ．相対危険度とは，特定要因へ暴露した群が，暴露しなかった群に比べて，何倍の危険度を有するかを表す指標であるこの相対危険度 RR が 1 に等しいかどうかを検定する．推定された相対危険度の有意性は，統計量 $\chi = \frac{\sqrt{N-1}\{(ad-bc)\pm N/2\}}{\sqrt{n_1 n_2 m_1 m_2}}$ が $RR = 1$ の仮定の元，標準正規分布 $N(0,1)$ に従うか否かで検定する．

2.2　秘匿内積プロトコル [1]

提案方式 (2) では，秘匿内積プロトコルを使用してプライバシーを保護したまま患者-対照調査を行う．2 つの組織がそれぞれ持つ X_A と X_B を秘匿したまま，積集合の大きさ $|X_A \cap X_B|$ のみを求める．計算結果は，$s_A + s_B = |X_A \cap X_B|$ となるような 2 つの乱数 s_A と s_B に分散されるため，計算が終わっても秘匿されている.

表 1　患者-対照調査における母集団のデータ

	死亡	生存	計
喫煙	a	b	n_1
非喫煙	c	d	n_2
計	m_1	m_2	N

2.3　Fairplay[2]

提案方式 (2) では，秘匿内積プロトコルを使用して得られた 2 つに分散された値を秘匿したまま有意水準を超えるか判定するために，任意の関数の秘匿回路評価システム Fairplay[2] を用いる．Fairplay は加算，減算，大小比較等の基本的な計算は高速にできるが，乗算，除算等の計算は極めて遅い．そのため，Fairplay においては乗算の利用を極力避けなくてはならない.

3　プライバシーを保護した相対危険度の検定

3.1　問題定義

秘匿の必要がある集合 X_A を持つ組織 A と，X_B を持つ組織 B が協力して疫学調査を行う．例えば，組織 A は特定要因である喫煙者のデータを持っている組織，組織 B は死亡者のデータを管理する組織とする．すなわち，組織 A は表 1 の n_1，n_2 の情報を持ち，組織 B は m_1, m_2 の情報を持つ．それぞれの持っているデータの合計 N は公開するが，属性毎の合計値，n_1，n_2，m_1，m_2 は秘密とする．これらの条件の元で，効率的に疫学調査を行い，対象とする特定要因の相対危険度が有意かどうかを検定する．出力結果は，その検

定結果のみとする．

これを解くナイーブな方法は，X_A と X_B をベクトル表現し，2.2 節の秘匿内積プロトコルを適用することであるが，これには，統計量 χ を求めるために大きな計算量が必要であり，Fairplay で計算するのは困難であるという問題点がある．

3.2 提案方式

問題点を解決するために，Fairplay での計算は加算や減算，比較のみに限定したい．そこで，次に示す様に，検定の条件を Fairplay で可能な等価な式に変形する．検定を行った際に有意になるか否かは，N，n_1，m_1 を固定した場合，a の大きさによって求めることができる．まず，χ を求める式を a のみを用いるように変形すると，

$$\chi = \frac{\sqrt{N-1}\{aN - n_1m_1 - N/2\}}{\sqrt{n_1n_2m_1m_2}} \qquad (2)$$

と，a の一次式になる．この値が有意水準 $Z(0.05/2) = 1.960$ を超えたかを判断すれば良い．この境界の a を a^* とおくと，式 (2) より

$$a^* = \left(\frac{\chi \cdot \sqrt{n_1n_2m_1m_2}}{\sqrt{N-1}} + n_1m_2 + \frac{N}{2}\right) \cdot \frac{1}{N}$$

$$a^*N = \frac{1.960 \cdot \sqrt{n_1n_2m_1m_2}}{\sqrt{N-1}} + n_1m_2 + \frac{N}{2} \qquad (3)$$

となる．ここで，χ と N は与えられた公開情報であり，$\sqrt{n_1n_2m_1m_2}$ は n_1 と n_2 を持つ A と m_1，m_2 を持つ B が秘匿内積プロトコルを行えば分散した値が得られる．n_1m_1 も同様である．従って，秘匿内積プロトコルで X_A と X_B の積集合 a が，a^* を上回っているか否かを Fairplay で検査するには，

$$(s_A + s_B) > (t_A + t_B) + (u_A + u_B) \qquad (4)$$

を評価し，判定結果のみを出力すればよい．ここで，s_A, s_B, t_A, t_B, u_A, u_B は，

$s_A + s_B = aN = |X_A \cap X_B|N$,

$t_A + t_B = \dfrac{\chi \cdot \sqrt{n_1n_2m_1m_2}}{\sqrt{N-1}} = \left(\dfrac{\chi\sqrt{n_1n_2}}{\sqrt{N-1}}\right) \cdot \sqrt{m_1m_2}$,

$u_A + u_B = n_1m_1 + \dfrac{N}{2}$

で定義される値であり，秘匿内積プロトコルにより効率的に求めることができる．

4 評価

4.1 パフォーマンス

Java BigInteger クラスを用いて実装したプログラムの処理時間を図 1 に示す．

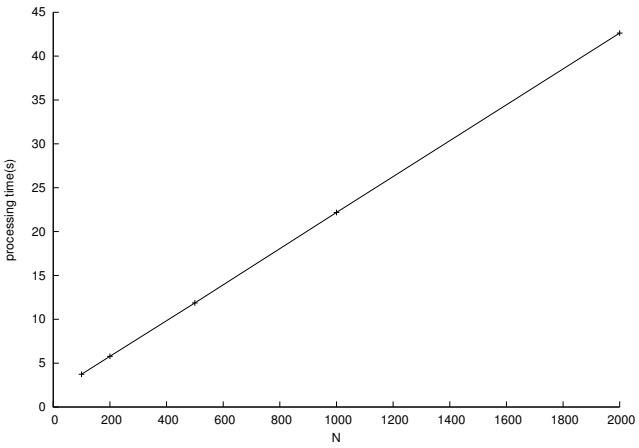

図.1 実装したプログラムの処理時間

5 おわりに

秘匿内積プロトコルと Fairplay を使用し，2 つのデータセットを秘匿したまま，確率検定を行うシステムを実装した．実装する際にネックとなる Fairplay の制約に対して，従来の式を変形し加算や減算，比較のみで計算を行うことで，要素数 14 万件に対して，約 48 分と現実的な時間で処理を行えるようになった．

参考文献

[1] Goethals, Laur, Lipmaa and Mielikainen, “On Private Scalar Product Computation for Privacy-Preserving Data Mining”, ICISC 2004, Vol. 3506 of LNCS, pp. 104-120, 2004.

[2] Dahlia, Nisan, Pinkas, and Sella, “Fairplay - A Secure Two-Party Computation System”, Usenix Security Symposium, pp. 1-17, 2004.

業績リスト

1. 佐藤, 菊池, 佐久間, “プライバシーを保護した放射線疫学調査システム”, CSEC54, Vol.2011-CSEC-54, No.25, pp. 1-6, 2011.
2. 佐藤, 菊池, 佐久間, “プライバシー保護確率検定システムの実装と評価”, ICSS2012, 信学技報 Vol. 112 No.315, pp. 61-66, 2012.
3. 佐藤, 菊池, 佐久間, “傾向性の検定における秘匿疫学調査プロトコル”, SCIS2013, 3C1-4, pp. 1-4, 2013.
4. Sato, Kikuchi, “Synthesis of Secure Password”, The 7th Asia Joint Conference on Information Security(AsiaJCIS2012), pp. 1-3, 2012.