推論の信頼性を考慮した不確実な知識の表現法と推論法について

鈴木 誠[†,☆] 松嶋敏泰[††,☆☆] 平澤茂一[†]

近年，知識工学の分野において，さまざまなエキスパートシステムが活発に開発されている．数多くのエキスパートシステムの中に，不確実な知識を確信度という概念を用いて記述したものが存在する．しかし，従来の確信度を用いた推論方式では推論結果の信頼性について言及されていない．このため，その確信度を用いたエキスパートシステムによって推論された結果下される判断は，どの程度妥当なのか，どのくらいの信頼性があるのかということについては，明確にされていない．そこで本研究は，システムの信頼性を保持しながら，確信度を持つ推論を行う方法について考察を行っている．すなわち，確率世界論理における確信度の推定に統計学の区間推定の理論を用い，確信度の上限と下限を明示する推論方式を提案する．その結果，導出された推論結果の信頼性を確率 $1-\alpha$（α は適当に小さい定数）で保証することができる．また，この幅を持つ確信度を用いて帰納・演繹の推論が一連の推論として考察可能となる．本研究は，エキスパートシステムなどの推論結果の信頼性を保証するという見地から有効な手段である．

On Knowledge Representation and Reasoning System with a Range of Certainty Factors

Makoto Suzuki,[†,☆] Toshiyasu Matsushima[††,☆☆] and Shigeichi Hirasawa[†]

We discuss the method for representing uncertain knowledge information and also on a reasoning system that guarantees a given reliability based on information theory and probability theory. In this paper, we propose a new reasoning system which consists of inductive and deductive inferences, by using a knowledge representation method with a range of certainty factors. In inductive inference, a new atom is computed from the observed data or facts, together with the upperbound and lowerbound between which its probability must fall. Also in deductive inference, which includes negation, conjunction, disjunction, chain and modus ponens, a new atom is deduced with a range of certainty factors computed from the results of inductive inference. Properties and behaviors of the system are clarified. It can be concluded that the new reasoning system is effectively applicable to inference engines of conventional expert systems.

1. まえがき

近年，人工知能の分野では，不確実な知識の処理が重要な問題の1つとなっている．この不確実な知識の処理の基礎理論として，従来の第一階述語論理のみでは不十分であるため，それを補う方法としてさまざまな提案がなされている．これらの従来法は大きく2つに分類される．1つは，様相論理や高階の述語論理に属する方法[3]で，論理記号や推論規則を付加する方法であり，他の1つは，各命題に補助情報としてその命題の確実さを数値で付加する方法[1),2)]である．後者の立場の中に属する研究で確率的計算法を用いた有効な推論法は，命題の確実さを確率値で表現し，通常，その値を確信度と呼ぶ．従来のほとんどの研究は確信度が専門家により主観的に与えられるとしていた．しかし，最近，専門家へのインタビューや事実を観測することによって知識獲得を行おうとする研究も活発になされている．本研究では何らかの事実を観測することにより論理式の確信度を求める状況を考える．つまり，統計学的見地に立って，この確信度をある母集団からデータをサンプリングした結果得られる客観的な尺度であると考えている．従来の研究のほとんどは，

† 早稲田大学理工学部工業経営学科
School of Science and Engineering, Waseda University
†† 横浜商科大学経営情報学科
Department of Management Information, Yokohama Collegeof Commerce
☆ 現在，(株) 東芝 研究開発センター システム・ソフトウェア生産技術研究所
Presently with Systems & Software Engineering Laboratory, R & D Center, Toshiba Corporation
☆☆ 現在，早稲田大学理工学部工業経営学科

その確信度の値を1個のパラメータで一意に求めていた．これは上述の統計学的見地に立つと統計学における点推定ともみなせる．真の確信度が存在するという仮定をおくと，従来の点推定ともみなせる推論法の多くは，その真の確信度に近い値を求めているが，その導出された確信度がどの程度信用できるものなのか（確信度の信頼性）については議論されていなかった．すなわち，サンプル数が多い場合と少ない場合では，導出された確信度の信頼性はおのずと異なるにもかかわらず，同じ推論結果を導出していた．一方，確信度を2個のパラメータで表現しようという試みもいくつか提案された．その代表的な手法の1つとして Dempster-Shafer の確率理論[4)~6)]（以下，D-S 理論）に基づく推論法がある．この D-S 理論に基づく推論法[7)~11)]は，確信度に幅を持たせ，確信度の上界，下界確率によって不確実性を扱おうとしている．ここで扱おうとしている不確実性とは，情報の欠如に起因している．また，最近，岡本らによって，信念推論システム[12)]が提案された．これは，知識の不確実性を確率的な現象の不確実性に起因するものと認識主体の側の情報の不足に起因するものに分け，確信度と主観確率という2個のパラメータで不確実性を扱うことを試みている．いずれの手法においても，不確実性の扱いについて，ある事象（論理式）に関する観測情報を基に決定される「論理式が真である度合い」を扱っている本手法とは立場を異にしている．

本研究では，確率世界論理[13)]における確信度の推定に統計学の区間推定の理論を用い，確信度の上限と下限を示す確信度区間推論方式を提案する．そして，確信度に幅を持たせることにより，その確信度区間内に先に仮定した真の確信度が存在することを $1-\alpha$（α は適当に小さい定数☆）の信頼性で保証し，この確信度を用いて帰納・演繹の一連の推論が可能となる論理体系を構築することを目的としている[16),17)]．ここで，本研究でいう帰納推論とは，ある事実をもとにある論理式の確信度を導いてくる推論を，演繹推論とはその帰納推論によって導出された論理式の確信度に従って新たな論理式の確信度を求める推論を意味している．

また，確信度区間推論方式において，論理式の確信度が導出される過程は以下のとおりである．

1) 現実世界から，ある論理式に関するサンプルを採る．
2) 1) のサンプルをもとに帰納推論することにより，その論理式の確信度区間を求める．
3) 2) で求められた確信度区間を利用し演繹推論することにより，導出された新しい論理式の確信度区間を求める．

このため，本研究では，なぜ確信度に幅を持たせる必要があるかという点について，以下のような立場をとっている．すなわち，本論文では，上記のように，ある論理式の確信度が何らかの帰納推論によって求められるという観点から，その推論結果の信頼性を保証しようとすることを目的としている．この問題を統計学におけるパラメータの区間推定に対応させて考えた場合，その推論結果である確信度に対して幅を持たせた表現がおのずと必要とされるのである．このように，本手法では，確信度に幅を持つ知識表現法に対して，数理的に明確なモデルを提供している．

本論文では3章で確信度区間推論方式を提案する．その手順として，第一に，確信度に幅を持つ知識表現法を定義する．第二に，システムの知識ベースを構築するための帰納推論法について論じる．第三に，演繹推論法として，否定，連言，選言，連鎖，modus ponens の5つの推論法に対する確信度区間の算出方法を示す．さらに4章で，確信度区間推論方式について考察を行い，いくつかの重要な性質を明らかにし，その有効性を示す．

2. 準　　備

本章では，従来の研究として，確率世界論理[13)]について述べる．

2.1 確率世界論理

論理学における古典論理上の解釈のパターン☆を確率変数とみなすことにより，第一階述語論理に確率構造を導入した理論に確率世界論理がある．そして，論理学の推論を情報の変換の一種としてとらえ，帰納推論を情報源符号化の観点から，演繹推論を復号化の観点から考察し，高次知識情報処理の統一した基礎理論を情報理論の立場から構築しようとする研究がなされている[13),14)]．

［定義1］ 確率世界論理の定義

可算な第一階述語論理言語 L に対する古典論理的解釈のパターンを要素とした集合である解釈空間 Ω を標本空間とし，その部分集合の σ 集合体とその上の確

☆ α とは推論結果の誤り率を意味しており，以下，α は本推論方式が使われるシステムの不信頼性として固定されているものとする．

☆ 第一階言語 L のすべてのグランドアトムに対する真理値の付値を1つの解釈とする古典論理的解釈をいう．

率測度 P を定義することにより確率空間を構成する．このような確率空間を持つ述語論理を確率世界論理と定義する． □

［定義2］ 論理式 W が真である確率の定義

論理式 W が真である確率 $P(w=1)$ は

$$P(w=1)=\sum_{m} M(W|m)P(m) \quad (1)$$

で示される．ここで，$P(m)$ は解釈空間 Ω 上で解釈のパターン m の確率であり，$M(W|m)$ は以下のような関数である．

$$M(W|m)=\begin{cases}1, & \text{解釈 } m \text{ において論理式 } W \text{ は真} \\ 0, & \text{解釈 } m \text{ において論理式 } W \text{ は偽}\end{cases} \quad (2)$$ □

確率世界論理では，2種類の様相（全称限定子・必然性様相）を対象とし，その様相記号の拡張を行っている．

［定義3］ 確率全称限定子記号 Γ の定義

数量化された全称限定子を確率全称限定子記号 Γ で次のように表現する．

$$\Gamma xQ(x): \sum_{x} M(Q(x)|m)P(x|m) \in [0, 1] \quad (3)$$

ここで，x は変数，$P(x|m)$ は世界 m における定数空間（個体領域）上での確率分布である． □

［定義4］ 確率必然様相記号 Δ の定義

数量化された必然性様相を確率必然様相記号 Δ で次のように定義する．

$$\Delta Q(a): \sum_{m} M(Q(a)|m)P(m) \in [0, 1] \quad (4)$$ □

定義3，4で定義された確率全称限定子記号 Γ と確率必然様相記号 Δ についての説明を以下に加える．確率全称限定子記号 Γ の定義3では，その定数の定義域で $Q(x)$ が成り立つ割合を論理式の後に付加して表現している．確率必然様相記号 Δ の定義4では，考えられるすべての世界で $Q(a)$ が真である割合を論理式の後に付加して表現している．以上の2つの確率論理記号のイメージを図1に示す．

後述の帰納推論におけるサンプリングの直感的なイメージの例を示すと，確率全称限定子記号 Γ は，ある時点での固定された解釈の下で，その定義域内の各定数についてサンプリングを行い，確率必然様相記号 Δ☆ は，ある同一の定数について時間的推移を経てサンプリングを行うと考えられる．

☆ 1つの世界だけを対象としている場合は確率必然様相記号 Δ を外して考える．

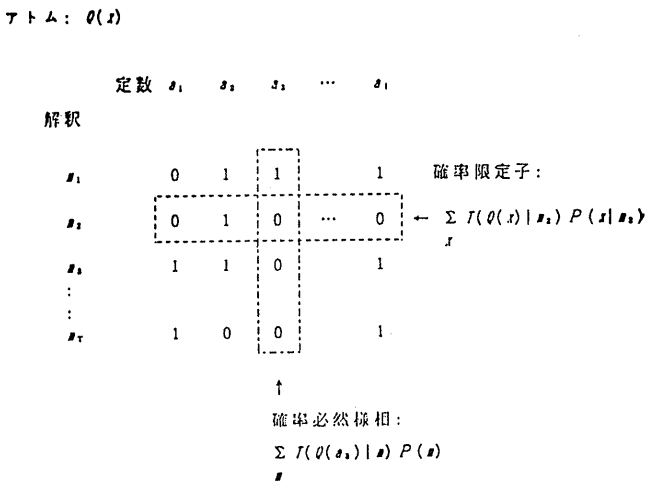


図1 確率限定子と確率必然様相

Fig. 1 Probabilistic universal quantifier and probabilistic necessity operator.

例えば，$\Gamma xQ(x)$ では，$Q(x)$ を「x は学生である」を意味すると仮定すると，その時点で一郎，次郎，三郎，…について学生か否かのサンプリングを行うこととする．$\Delta Q(a)$ では，$Q(a)$ を「a は朝食をとる」を意味すると仮定すると，定数 a（一郎ならば一郎一人）について，昨日，今日，明日，…と時間的推移を経ながらサンプリングを行うことになる．特に，確率全称限定子記号 Γ と確率必然様相記号 Δ を同時に考える場合は，$\Delta\Gamma xQ(x)$ と表現する．

3. 確信度区間推論方式

本章では，以下の手順で確信度区間推論方式を提案する．

1) 確信度に幅を持たせた知識の表現法を定義する．
2) n 個の事実をランダムにサンプリングし，ある論理式に関してその真偽を観測した結果，k 個が真であった場合，その論理式が真である確率は二項分布に従うことに着目し，区間推定より，幅を持たせた確信度を求める帰納推論法について述べる．
3) 2)によって求められた確信度の上限，下限を用いて，新たな論理式の確信度の上限と下限を求める演繹推論法について論じる．

これにより，演繹推論を何段か重ねて行っても，そのシステムの信頼性を近似的に $1-\alpha$ に保証しながら推論結果を導くことができる推論方式を構築することができる．

3.1 知識表現法

本節では，確信度に幅を持たせた知識の表現法を以下のように定義する．

［定義5］ 知識表現法の定義

ある個体領域 x におけるある定数 a に対して $Q(a)$ が真である確率が P_L 以上 P_U 以下であることが $1-\alpha$ の信頼性でいえるとき，このことを次のように表現する．

$$\Delta\Gamma\, xQ(x):[P_L, P_U](1-\alpha) \tag{5}$$

ここで，P_U を上限確信度，P_L を下限確信度，P_U と P_L を総称して確信度区間と定義する．また，α は一連の帰納・演繹推論において本推論方式が使われるシステムの不信頼性として固定された適当に小さいある定数とする． □

ここで，特記しておきたい点は，上で定義した知識表現法では，その個体領域 x におけるある定数 a に対して $Q(a)$ が真である確率に幅を持たせ，その確信度区間（$[P_L, P_U]$）とその確信度区間の信頼性を信頼係数（$1-\alpha$）として論理式の後に付加していることである．

（例 3.1） $\Delta\Gamma\, xJAPANESE(x):[0.7,\ 0.8]\ (1-0.05)$ は対象としている個体領域における日本人の割合が 70～80% の間にあることが 0.95 の確率でいえることを示している．

以後，本論文では，特に断りのない限り，確信度区間の付いている論理式（例えば，$\Delta\Gamma\, xQ(x):[P_L, P_U](1-\alpha)$）のことを単に論理式と呼ぶこととする．

3.2 帰納推論法

先にも述べたように，本研究でいう帰納推論とは，ある論理式に関する事実をサンプリングすることにより，その論理式の確信度区間を導出し，知識ベースを構築する推論を意味している．

n 個の個体 $a_1, a_2, \cdots, a_n$ をランダムにサンプリングし，$Q(a_1), Q(a_2), \cdots, Q(a_n)$ の真偽を観測した結果，k 個が真であったとすると，一般的にこの事象はベルヌーイの系列とみなせて，$Q(x)$ が真である個数（確率変数）k は $Q(x)$ が真である確率 p^* をパラメータとした二項分布 $B(k:n, p^*)$ に従うと仮定できる[18]．この仮定の下で，以下の補題1のような帰納推論法が導かれる．

［補題1］ $\Delta\Gamma\, xQ(x):[P_{LI}, P_{UI}](1-\alpha)$ についての帰納推論

n 個のサンプルを観測したとき，Q という属性に関して k 個が真であったとすると，$\Delta\Gamma\, xQ(x):[P_{LI}, P_{UI}](1-\alpha)$ の上限確信度 P_{UI} と下限確信度 P_{LI} は次式のように求まる．

$$P_{UI}=\frac{k}{n}+Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{n}\frac{k}{n}\left(1-\frac{k}{n}\right)}$$

$$P_{LI}=\frac{k}{n}-Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{n}\frac{k}{n}\left(1-\frac{k}{n}\right)} \tag{6}$$

ここで，二項分布はサンプル数 n が大で，k/n の値が極端に0や1に近くない場合，正規分布に近似できることを用いており，$Z(\alpha/2)$ は標準正規分布のパーセント点である． □

3.3 演繹推論法

本研究でいう演繹推論とは，補題1に示した帰納推論によって導出された論理式の確信度区間をもとに新たな論理式の確信度区間を導出する推論を意味している．

ここで扱う演繹推論は，否定，連言，選言，連鎖，modus ponens の5つの演繹推論法である．この連鎖と modus ponens は，近年の人工知能や数理論理学などの分野でよく用いられる推論規則の三段論法の中でも最も基本的なものとして位置づけられている．また，従来の確信度を用いる方法では，確信度が付加された論理式における推論規則について導出原理から考察している方法[13]と，含意ルールの確信度を条件付き確率とみなし考察している方法[1),13]の2通りがある．ここでは，後者の立場から考察を行っている．

以下，上記の5つの演繹推論法のいずれも次の手順で導かれる．

1) 導出したい論理式の確信度区間の中心となる値を求める．
2) その論理式の確信度の推定値の分布の分散を求める．
3) その論理式の確信度区間の上限，下限確信度を，サンプル数（n_i）と真であるものの個数（k_i）で表した式を示す．

すなわち，本研究の各演繹推論は，結論として導かれる論理式の確信度の確率分布を直接考えることにより，その分布の平均と分散を計算し比較的精度の良い近似で推論結果を導いてくる立場をとっている．結果として，各演繹推論法は次の定理1～5で与えられる．

3.3.1 否定の演繹推論

否定の演繹推論とは，$\Delta\Gamma\, xQ(x):[P_{LI}, P_{UI}](1-\alpha)$ の確信度区間がわかっているとき，$\Delta\Gamma\, x\neg Q(x):[P_{LD}, P_{UD}](1-\alpha)$ の確信度区間を求める推論である．

［定理１］ 否定の演繹推論

帰納推論の結果，$\Delta\Gamma\, xQ(x):[P_{LI},P_{UI}](1-\alpha)$ が得られたとする．このとき，$\Delta\Gamma\, x\neg Q(x):[P_{LD},P_{UD}](1-\alpha)$ の上限確信度 P_{UD} と下限確信度 P_{LD} は，次式で示される．

$$P_{UD}=1-A+B$$
$$P_{LD}=1-A-B \tag{7}$$

ここで，

$$A=\frac{P_{UI}+P_{LI}}{2},\quad B=\frac{P_{UI}-P_{LI}}{2} \tag{8}$$

［証明］ 3.2 節の帰納推論法の一般的な仮定で述べたように $\Delta\Gamma\, xQ(x):[P_{LD},P_{UD}](1-\alpha)$ の確信度の分布は二項分布 $B(k:n,p^*)$ に従うので，その否定の論理式 $\Delta\Gamma\, x\neg Q(x):[P_{LD},P_{UD}](1-\alpha)$ の確信度の分布は二項分布 $B(k:n,1-p^*)$ に従う．よって，$\Delta\Gamma\, x\neg Q(x):[P_{LD},P_{UD}](1-\alpha)$ の上限確信度 P_{UD} と下限確信度 P_{LD} は，先に述べた補題１の場合に帰着する．

今，A は $\Delta\Gamma\, xQ(x):[P_{LI},P_{UI}](1-\alpha)$ の確信度区間の中央値に，B は $\Delta\Gamma\, xQ(x):[P_{LI},P_{UI}](1-\alpha)$ の確信度区間の幅の 1/2 に相当している．□

例えば，$Q(x)$ を「x は学生である」を意味すると仮定すると，否定の演繹推論とは，「x は学生である」の確信度区間が［0.79, 0.81］とわかっているとき，「x は学生でない」の確信度区間を以下のように求める推論である．

$$\frac{\Delta\Gamma\, xQ(x):[0.79,\ 0.81](0.95)}{\Delta\Gamma\, x\neg Q(x):[0.19,\ 0.21](0.95)}$$

3.3.2 連言の演繹推論

連言の演繹推論とは，２つの論理式 $\Delta\Gamma\, xQ_1(x):[P_{L1},P_{U1}](1-\alpha)$ と $\Delta\Gamma\, xQ_2(x):[P_{L2},P_{U2}](1-\alpha)$ の確信度区間がおのおのわかっているとき，$\Delta\Gamma\, x(Q_1(x)\wedge Q_2(x)):[P_{LD},P_{UD}](1-\alpha)$ の確信度区間を求める推論である．

まず最初に，帰納推論時のサンプリングの条件について述べる．n_1, n_2 個の個体をランダムにサンプリングし，互いに独立な２つの属性 Q_1, Q_2 に関してその真偽を観測した結果，おのおの k_1, k_2 個が真であったとする．このとき，次の定理２が得られる．

［定理２］ 連言の演繹推論

帰納推論の結果，独立な２つの論理式 $\Delta\Gamma\, xQ_1(x):[P_{L1},P_{U1}](1-\alpha)$, $\Delta\Gamma\, xQ_2(x):[P_{L2},P_{U2}](1-\alpha)$ が得られたとする．このとき，$\Delta\Gamma\, x(Q_1(x)\wedge Q_2(x)):[P_{LD},P_{UD}](1-\alpha)$ の上限確信度 P_{UD} と下限確信度 P_{LD} は，次式で示される．

$$P_{UD}=A_1A_2+\frac{1}{\left\{Z\left(\frac{\alpha}{2}\right)\right\}\alpha^{1/2}}\left[\frac{B_1^2A_1A_2(1-A_2)}{1-A_1}+\frac{B_2^2A_1A_2(1-A_1)}{1-A_2}+\frac{B_1^2B_2^2}{\left\{Z\left(\frac{\alpha}{2}\right)\right\}^2}\right]^{1/2}$$

$$P_{LD}=A_1A_2-\frac{1}{\left\{Z\left(\frac{\alpha}{2}\right)\right\}\alpha^{1/2}}\left[\frac{B_1^2A_1A_2(1-A_2)}{1-A_1}+\frac{B_2^2A_1A_2(1-A_1)}{1-A_2}+\frac{B_1^2B_2^2}{\left\{Z\left(\frac{\alpha}{2}\right)\right\}^2}\right]^{1/2} \tag{9}$$

ここで，

$$A_i=\frac{P_{Ui}+P_{Li}}{2},\ B_i=\frac{P_{Ui}-P_{Li}}{2}\quad(i=1,2) \tag{10}$$

［証明］ 付録Ａ参照．□

例えば，$Q_1(x)$ が「x は学生である」を意味し，$Q_2(x)$ が「x は下宿している」を意味すると仮定すると，連言の演繹推論とは，「x は学生である」の確信度区間が［0.79, 0.81］と「x は下宿している」の確信度区間が［0.59, 0.61］とわかっているとき，「x は学生であり，かつ下宿している」の確信度区間を以下のように求める推論である．

$$\frac{\begin{array}{c}\Delta\Gamma\, xQ_1(x):[0.79,0.81](0.95)\\ \Delta\Gamma\, xQ_2(x):[0.59,0.61](0.95)\end{array}}{\Delta\Gamma\, x(Q_1(x)\wedge Q_2(x)):[0.46,0.50](0.95)}$$

3.3.3 選言の演繹推論

選言の演繹推論とは，２つの論理式 $\Delta\Gamma\, xQ_1(x):[P_{L1},P_{U1}](1-\alpha)$ と $\Delta\Gamma\, xQ_2(x):[P_{L2},P_{U2}](1-\alpha)$ の確信度区間がおのおのわかっているとき，$\Delta\Gamma\, x(Q_1(x)\vee Q_2(x)):[P_{LD},P_{UD}](1-\alpha)$ の確信度区間を求める推論である．

また，帰納推論時のサンプリングの条件については，連言の演繹推論の場合と全く同じである．このとき，次の定理３が得られる．

［定理３］ 選言の演繹推論

帰納推論の結果，独立な２つの論理式 $\Delta\Gamma\, xQ_1(x):[P_{L1},P_{U1}](1-\alpha)$, $\Delta\Gamma\, xQ_2(x):[P_{L2},P_{U2}](1-\alpha)$ が得られたとする．このとき，$\Delta\Gamma\, x(Q_1(x)\vee Q_2(x)):[P_{LD},P_{UD}](1-\alpha)$ の上限確信度 P_{UD} と下限確信度 P_{LD} は，次式で示される．

$$P_{UD}=A_1+A_2-A_1A_2+\frac{1}{\left\{Z\left(\frac{\alpha}{2}\right)\right\}\alpha^{1/2}}$$

$$\left[B_1^2(1-A_2)^2+B_2^2(1-A_1)^2+\frac{B_1^2B_2^2}{\left\{Z\left(\frac{\alpha}{2}\right)\right\}^2}\right]^{1/2}$$

$$P_{LD}=A_1+A_2-A_1A_2-\frac{1}{\left\{Z\left(\frac{\alpha}{2}\right)\right\}\alpha^{1/2}}$$

$$\left[B_1^2(1-A_2)^2+B_2^2(1-A_1)^2+\frac{B_1^2B_2^2}{\left\{Z\left(\frac{\alpha}{2}\right)\right\}^2}\right]^{1/2} \tag{11}$$

ここで,

$$A_i=\frac{P_{Ui}+P_{Li}}{2},\ B_i=\frac{P_{Ui}-P_{Li}}{2}\quad(i=1,2) \tag{12}$$

[証明] 定理2とほぼ同様に証明される. □

例えば, $Q_1(x)$ が「x は学生である」を意味し, $Q_2(x)$ が「x は下宿している」を意味すると仮定すると, 選言の演繹推論とは,「x は学生である」の確信度区間が [0.79, 0.81] と「x は下宿している」の確信度区間が [0.59, 0.61] とわかっているとき,「x は学生であるか, または下宿している」の確信度区間を以下のように求める推論である.

$$\begin{array}{c} \Delta\Gamma\,xQ_1(x):[0.79,0.81](0.95) \\ \underline{\Delta\Gamma\,xQ_2(x):[0.59,0.61](0.95)} \\ \Delta\Gamma\,x(Q_1(x)\vee Q_2(x)):[0.91,0.93](0.95) \end{array}$$

3.3.4 連 鎖

連鎖とは, 2つの論理式 $\Delta\Gamma\,x(Q_1(x)\rightarrow Q_2(x)):[P_{L1},P_{U1}](1-\alpha)$ と $\Delta\Gamma\,x(Q_2(x)\rightarrow Q_3(x)):[P_{L2},P_{U2}](1-\alpha)$ の確信度区間がおのおのわかっているとき, $\Delta\Gamma\,x(Q_1(x)\rightarrow Q_3(x)):[P_{LD},P_{UD}](1-\alpha)$ の確信度区間を導く推論である.

まず, 帰納推論時のサンプリング条件とその帰納推論結果を述べておく. n_1 個の個体をランダムにサンプリングし, 属性 Q_1, Q_2 に関してその真偽を観測した結果, Q_1 に関して真だった下で Q_2 に関しても真である個体の数が k_1 個であったとする. この条件の下で帰納推論を行うと $\Delta\Gamma\,x(Q_1(x)\rightarrow Q_2(x)):[P_{L1},P_{U1}](1-\alpha)$ が得られる. また, 同様に n_2 個の個体をランダムにサンプリングし, 属性 Q_2, Q_3 に関してその真偽を観測した結果, Q_2 に関して真だった下で Q_3 に関しても真である個体の数が k_2 個, $\neg Q_2$ に関して真だった下で Q_3 に関しても真である個体の数が k_3 個であったとする. この条件の下で帰納推論を行うと $\Delta\Gamma\,x(Q_2(x)\rightarrow Q_3(x)):[P_{L2},P_{U2}](1-\alpha)$, $\Delta\Gamma\,x(\neg Q_2(x)\rightarrow Q_3(x)):[P_{L3},P_{U3}](1-\alpha)$ がおのおの得られる.

また, 各属性間の関係は, Q_1 と Q_3 が Q_2 の中で互いに独立であると仮定する. このとき, 次式が成り立つ.

$$p^*_{1,2,3}=\frac{p^*_{1,2}p^*_{2,3}}{p^*_2} \tag{13}$$

ここで $p^*_{1,2,3}$ は属性 Q_1, Q_2, Q_3 のすべてが真となる3変数の同時確率を表す. $p^*_{1,2}$, $p^*_{2,3}$ は, $p^*_{1,2,3}$ に準じた2変数の同時確率とする.

この条件と仮定の下で, 以下の定理4が得られる.

[定理4] 連鎖

帰納推論の結果, 3つの論理式 $\Delta\Gamma\,x(Q_1(x)\rightarrow Q_2(x)):[P_{L1},P_{U1}](1-\alpha)$, $\Delta\Gamma\,x(Q_2(x)\rightarrow Q_3(x)):[P_{L2},P_{U2}](1-\alpha)$, $\Delta\Gamma\,x(\neg Q_2(x)\rightarrow Q_3(x)):[P_{L3},P_{U3}](1-\alpha)$ が得られたとする. このとき, 属性 Q_2 の中で Q_1 と Q_3 が互いに独立であると仮定すると, $\Delta\Gamma\,x(Q_1(x)\rightarrow Q_3(x)):[P_{LD},P_{UD}](1-\alpha)$ の上限確信度 P_{UD}, 下限確信度 P_{LD} は, 次式で表される.

$$P_{UD}=A_1A_2+(1-A_1)A_3+\frac{1}{\left\{Z\left(\frac{\alpha}{2}\right)\right\}\alpha^{1/2}}$$

$$\left[B_2^2\left\{A_1^2+\frac{(1-A_1)^2A_3(1-A_3)}{A_2(1-A_2)}\right\}+B_1^2(A_2^2+A_3^2)\right.$$

$$\left.+\frac{B_1^2B_2^2}{\left\{Z\left(\frac{\alpha}{2}\right)\right\}^2}\left\{1+\frac{A_3(1-A_3)}{A_2(1-A_2)}\right\}-2B_1^2A_2A_3\right]^{1/2}$$

$$P_{LD}=A_1A_2+(1-A_1)A_3-\frac{1}{\left\{Z\left(\frac{\alpha}{2}\right)\right\}\alpha^{1/2}}$$

$$\left[B_2^2\left\{A_1^2+\frac{(1-A_1)^2A_3(1-A_3)}{A_2(1-A_2)}\right\}+B_1^2(A_2^2+A_3^2)\right.$$

$$\left.+\frac{B_1^2B_2^2}{\left\{Z\left(\frac{\alpha}{2}\right)\right\}^2}\left\{1+\frac{A_3(1-A_3)}{A_2(1-A_2)}\right\}-2B_1^2A_2A_3\right]^{1/2} \tag{14}$$

ここで,

$$A_i=\frac{P_{Ui}+P_{Li}}{2},\ B_i=\frac{P_{Ui}-P_{Li}}{2}\quad(i=1,2,3) \tag{15}$$

[証明] 付録B参照. □

例えば, $Q_1(x)$ が「x は学生である」を, $Q_2(x)$ が「x は下宿している」を, $Q_3(x)$ が「x は朝食をとらない」を意味すると仮定すると, 連鎖とは,「x は学生であるならば x は下宿している」の確信度区間が [0.79, 0.81] と「x は下宿しているならば x は朝食をとらない」の確信度区間が [0.59, 0.61] とわかっているとき,「x は学生であるならば x は朝食をとら

ない」の確信度区間を以下のように求める推論である．

$\Delta\Gamma\, x(Q_1(x)\rightarrow Q_2(x)):[0.79, 0.81](0.95)$

$\underline{\Delta\Gamma\, x(Q_2(x)\rightarrow Q_3(x)):[0.59, 0.61](0.95)}$

$\Delta\Gamma\, x(Q_1(x)\rightarrow Q_3(x)):[0.56, 0.60](0.95)$

定理４において，裏の情報[15] $\Delta\Gamma\, x(\neg Q_2(x)\rightarrow Q_3(x)):[P_{L3}, P_{U3}](1-\alpha)$ がわかっている場合を考えた．しかし一般的には，不確実な知識による推論では，推論を行う事前の情報が必ずしも完全に揃っているとは限らない条件の下で推論が行われることが多い．もし，$\Delta\Gamma\, x(\neg Q_2(x)\rightarrow Q_3(x)):[P_{L3}, P_{U3}](1-\alpha)$ がわからない場合は，$A_3=0.5$ とすることによって推論が可能である*.

3.3.5 modus ponens

modus ponens とは，２つの論理式 $\Delta Q(a):[P_{L1}, P_{U1}](1-\alpha)$, $\Delta\Gamma\, x(Q(x)\rightarrow R(x)):[P_{L2}, P_{U2}](1-\alpha)$ の確信度区間がおのおのわかっているとき，$\Delta R(a):[P_{LD}, P_{UD}](1-\alpha)$ の確信度区間を求める推論である．

［定理５］ modus ponens

帰納推論の結果，３つの論理式 $\Delta Q(a):[P_{L1}, P_{U1}](1-\alpha)$, $\Delta\Gamma\, x(Q(x)\rightarrow R(x)):[P_{L2}, P_{U2}](1-\alpha)$, $\Delta\Gamma\, x(\neg Q(x)\rightarrow R(x)):[P_{L3}, P_{U3}](1-\alpha)$ であるとする．このとき，$\Delta R(a):[P_{LD}, P_{UD}](1-\alpha)$ の上限確信度 P_{UD}, 下限確信度 P_{LD} は定理４の式(14)と同様に表される．

［証明］ 確率世界論理[15]の単一化の概念を用いると，定理４と全く同様に証明される． □

例えば，$Q(x)$ が「x は朝食をとらない」を，$R(x)$ が「x は下宿している」を意味すると仮定すると，modus ponens とは，「一郎は朝食をとらない」の確信度区間が [0.79, 0.81] と「x は朝食はとらないならば x は下宿している」の確信度区間が [0.59, 0.61] とわかっているとき，「一郎は下宿している」の確信度区間を以下のように求める推論である．

$\Delta Q(a):[0.79, 0.81](0.95)$

$\underline{\Delta\Gamma\, x(Q(x)\rightarrow R(x)):[0.59, 0.61](0.95)}$

$\Delta R(a):[0.56, 0.60](0.95)$

定理５においても定理４と同様に，$\Delta\Gamma\, x(\neg Q(x)\rightarrow R(x)):[P_{L3}, P_{U3}](1-\alpha)$ がわからない場合は，$A_3=0.5$ とすることによって推論が可能である．

4. 確信度区間推論方式の特性

前記のように，確信度区間推論方式においては，まず，ある論理式に関する事例をもとに帰納推論を行うことによりその論理式の確信度区間を導出する．そして，演繹推論の連言，選言，連鎖，modus ponens では，その帰納推論によって導出された２つの論理式の確信度区間をもとに演繹推論を行うことにより，新たな論理式の確信度区間を導いてきている．このとき，２つの論理式に関する帰納推論時のサンプル数(n_1, n_2)が等しい場合に導出された確信度区間をもとに演繹推論された推論結果と，サンプル数 (n_1, n_2) が極端に異なる場合に導出された演繹推論結果との間には違いがあると考えられる．そこで，第一に，複数の論理式の帰納推論時におけるサンプル数が異なる場合についての考察を行う．

また，エキスパートシステムなどの実用的な分野では，演繹推論を何段か重ねることによりある推論結果を導出してくることが多い．そこで，第二に，確信度区間推論方式により演繹推論を何段か重ねた場合についての考察を行う．

以上のように，本章では，確信度区間推論方式の特性について，２つの見地から考察を行う．

4.1 複数の論理式の帰納推論時におけるサンプル数が異なる場合について

前記のように，本節では複数の論理式の帰納推論時におけるサンプル数が異なる場合についての考察を行う．まず第一に，帰納推論，連言，選言の演繹推論，連鎖または modus ponens の場合の確信度区間の幅と推論の誤り率とサンプル数の関数関係について補題 2, 3, 4, 5 で明らかにする．第二に，補題 2, 3, 4, 5をもとに，本節の本題である複数の論理式の帰納推論時のサンプル数が等しい場合と異なる場合に行われた演繹推論の結果の確信度区間の幅の差についてを系 1, 2, 3 で示す．

4.1.1 確信度区間の幅と推論の誤り率とサンプル数について

サンプル数が一定の下で，確信度区間の幅と推論の誤り率は二律背反の関係にあるといえる．すなわち，ある論理式の確信度の二項分布を考えたとき，確信度区間の幅を広くすれば広くするほど推論の誤り率は小さくなるし，確信度区間の幅を狭くすれば狭くするほど推論の誤り率は大きくなるといえる．

また，推論の誤り率αが一定の下では，その論理式の確信度の分布の分散により，確信度区間の幅の大小が決定される．一方，分散はサンプル数と関数関係があり，サンプル数を多く採ると分散は小さくなる．このように，確信度区間の幅は推論の誤り率αとサンプ

* 確率世界論理[13]においてその正当性は示されている．

ル数の関数として表現可能である．

以下に，帰納推論の結果の確信度区間の幅と推論の誤り率αとサンプル数の関係を補題2で，そして，その帰納推論の結果をそのまま用いた場合の演繹推論の確信度区間の幅とそれらの関係を補題3，4，5で明らかにする．

［補題2］ 帰納推論の結果の確信度区間の幅

n個のサンプルを観測したときQという属性に関してk個が真であったとすると，$\Gamma xQ(x):[P_{LI}, P_{UI}](1-\alpha)$の確信度区間の幅は次式で示される．

$$I_I=2Z\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{n}\frac{k}{n}\left(1-\frac{k}{n}\right)} \tag{16}$$

ここで，$Z(\alpha/2)$は標準正規分布のパーセント点である．

［証明］ 補題1の式(6)より

$I_I=P_{UI}-P_{LI}$ □

補題2に関する数値計算の結果より，確信度区間の中央値（k/nの値）が0.5に近いほど，確信度区間の幅は広くなることが明らかになった．これは，確信度が0.5に近いほど，真でもなく偽でもないという意味で曖昧さが大きいことを意味していると考えられる．

［補題3］ 連言の演繹推論の結果の確信度区間の幅

n_1, n_2個の個体をランダムにサンプリングし，互いに独立な2つの属性Q_1, Q_2に関してその真偽をおのおの観測した結果，k_1, k_2個が真であったとする．このとき，$\Delta\Gamma x(Q_1(x)\wedge Q_2(x)):[P_{LD}, P_{UD}](1-\alpha)$の確信度区間の幅は次式で示される．

$$I_D=\frac{2}{\alpha^{1/2}}\left\{\frac{p_1^2p_2(1-p_2)}{n_2}+\frac{p_1(1-p_1)p_2^2}{n_1}+\frac{p_1(1-p_1)p_2(1-p_2)}{n_1n_2}\right\}^{1/2} \tag{17}$$

ここで，

$$p_1=\frac{k_1}{n_1},\ p_2=\frac{k_2}{n_2}$$

［証明］ 補題7の式(A.3)より

$I_D=P_{UD}-P_{LD}$ □

［補題4］ 選言の演繹推論の結果の確信度区間の幅

n_1, n_2個の個体をランダムにサンプリングし，互いに独立な2つの属性Q_1, Q_2に関してその真偽をおのおの観測した結果，k_1, k_2個が真であったとする．このとき，$\Delta\Gamma x(Q_1(x)\vee Q_2(x)):[P_{LD}, P_{UD}](1-\alpha)$の確信度区間の幅は次式で示される．

$$I_D=\frac{2\{(1-p_1)(1-p_2)\}^{1/2}}{\alpha^{1/2}}\left\{\frac{p_1(1-p_2)}{n_1}+\frac{p_2(1-p_1)}{n_2}+\frac{p_1p_2}{n_1n_2}\right\}^{1/2} \tag{18}$$

ここで，

$$p_1=\frac{k_1}{n_1},\ p_2=\frac{k_2}{n_2}$$

［証明］ 補題3とほぼ同様に証明される． □

［補題5］ 連鎖（または，modus ponens）の結果の確信度区間の幅

n_1個の個体をランダムにサンプリングし，属性Q_1, Q_2に関してその真偽を観測した結果，Q_1に関して真だった下でQ_2に関しても真である個体の数がk_1個であったとする．また，同様にn_2個の個体をランダムにサンプリングし，属性Q_2, Q_3に関してその真偽を観測した結果，Q_2に関して真だった下でQ_3に関しても真である個体の数がk_2個，$\neg Q_2$に関して真だった下でQ_3に関しても真である個体の数がk_3個であったとする．このとき得られる演繹推論結果$\Delta\Gamma x(Q_1(x)\rightarrow Q_3(x)):[P_{LD}, P_{UD}](1-\alpha)$の確信度区間の幅は次式で示される．

$$I_D=\frac{2}{\alpha^{1/2}}\left\{\frac{p_1^2p_2(1-p_2)+(1-p_1)^2p_3(1-p_3)}{n_2}+\frac{p_1(1-p_1)p_2^2+p_1(1-p_1)p_3^2}{n_1}+\frac{p_1(1-p_1)p_2(1-p_2)+p_1(1-p_1)p_3(1-p_3)}{n_1n_2}-\frac{2p_1(1-p_1)p_2p_3}{n_1}\right\}^{1/2} \tag{19}$$

ここで，

$$p_1=\frac{k_1}{n_1},\ p_2=\frac{k_2}{n_2},\ p_3=\frac{k_3}{n_3}$$

modus ponens に関しても同様に導かれる．

［証明］ 補題10の式(A.9)より

$I_D=P_{UD}-P_{LD}$ □

なお，演繹推論を2回以上重ねて行う場合は，より複雑な関数となることに注意されたい．

4.1.2 帰納推論時におけるサンプル数が等しい場合と異なる場合の演繹推論の確信度区間の幅の差について

それでは，本節の本題であるサンプル数が等しい場合と異なる場合の確信度区間の幅の差を数式で示す．以下に，定理2の連言の場合を系1で，定理3の選言の場合を系2で，定理4と5の連鎖とmodus ponensの場合を系3で述べる．

［系1］ サンプル数が等しい場合と異なる場合の連言の演繹推論の確信度区間の幅の差

帰納推論時に2つの論理式 $\Delta\Gamma\, xQ_1(x):[P_{L1}, P_{U1}](1-\alpha)$, $\Delta\Gamma\, xQ_2(x):[P_{L2}, P_{U2}](1-\alpha)$ を導出するもととなるサンプル数が等しい場合 ($n_{1e}=a$, $n_{2e}=a$, $p_1=k_1/a$, $p_2=k_2/a$) と異なる場合 ($n_{1d}=a$, $n_{2d}=b$, $p_1=k_1/a$, $p_2=k_2'/b$ または, $n_{1d}=b$, $n_{2d}=a$, $p_1=k_1'/b$, $p_2=k_2/a$) との下で導出された連言の演繹推論結果 $\Delta\Gamma\, x(Q_1(x)\wedge Q_2(x)):[P_{LD}, P_{UD}](1-\alpha)$ の確信度区間の幅の差は式(20)で示される．ここで, n_1, n_2 は属性 Q_1, Q_2 おのおのに関するサンプル数を, p_1, p_2 は論理式 $\Delta\Gamma\, xQ_1(x):[P_{L1}, P_{U1}](1-\alpha)$, $\Delta\Gamma\, xQ_2(x):[P_{L2}, P_{U2}](1-\alpha)$ おのおのの確信度の推定量を表している．

$$I_{dis}=\frac{2}{\alpha^{1/2}}\left|\left\{\frac{p_1^2p_2(1-p_2)}{n_{2e}}+\frac{p_1(1-p_1)p_2^2}{n_{1e}}+\frac{p_1(1-p_1)p_2(1-p_2)}{n_{1e}n_{2e}}\right\}^{1/2}-\left\{\frac{p_1^2p_2(1-p_2)}{n_{2d}}+\frac{p_1(1-p_1)p_2^2}{n_{1d}}+\frac{p_1(1-p_1)p_2(1-p_2)}{n_{1d}n_{2d}}\right\}^{1/2}\right| \tag{20}$$

特に, $n_{1d}\to\infty$ のとき,

$$I_{dis}\leq\frac{2}{\alpha^{1/2}}\left|\left\{\frac{p_1^2p_2(1-p_2)+p_1(1-p_1)p_2^2}{a}+\frac{p_1(1-p_1)p_2(1-p_2)}{a^2}\right\}^{1/2}-\left\{\frac{p_1^2p_2(1-p_2)}{a}\right\}^{1/2}\right|$$
$$(=I_{dismax}(n_{1d})) \tag{21}$$

また, $n_{2d}\to\infty$ のとき,

$$I_{dis}\leq\frac{2}{\alpha^{1/2}}\left|\left\{\frac{p_1^2p_2(1-p_2)+p_1(1-p_1)p_2^2}{a}+\frac{p_1(1-p_1)p_2(1-p_2)}{a^2}\right\}^{1/2}-\left\{\frac{p_1p_2^2(1-p_2)}{a}\right\}^{1/2}\right|$$
$$(=I_{dismax}(n_{2d})) \tag{22}$$

[証明]　補題3の式(17)より式(20)が導ける．　□

[系2]　サンプル数が等しい場合と異なる場合の選言の演繹推論の確信度区間の幅の差

帰納推論時に2つの論理式 $\Delta\Gamma\, xQ_1(x):[P_{L1}, P_{U1}](1-\alpha)$, $\Delta\Gamma\, xQ_2(x):[P_{L2}, P_{U2}](1-\alpha)$ を導出するもととなるサンプル数が等しい場合 ($n_{1e}=a$, $n_{2e}=a$, $p_1=k_1/a$, $p_2=k_2/a$) と異なる場合 ($n_{1d}=a$, $n_{2d}=b$, $p_1=k_1/a$, $p_2=k_2'/b$ または, $n_{1d}=b$, $n_{2d}=a$, $p_1=k_1'/b$, $p_2=k_2/a$) との下で導出された選言の演繹推論結果 $\Delta\Gamma\, x(Q_1(x)\vee Q_2(x)):[P_{LD}, P_{UD}](1-\alpha)$ の確信度区間の幅の差は式(23)で示される．ここで, n_1, n_2 は属性 Q_1, Q_2 おのおのに関するサンプル数を, p_1, p_2 は論理式 $\Delta\Gamma\, xQ_1(x):[P_{L1}, P_{U1}](1-\alpha)$, $\Delta\Gamma\, xQ_2(x):[P_{L2}, P_{U2}](1-\alpha)$ おのおのの確信度の推定量を表している．

$$I_{dis}=\frac{2\{(1-p_1)(1-p_2)\}^{1/2}}{\alpha^{1/2}}\left|\left\{\frac{p_1(1-p_2)}{n_{1e}}+\frac{p_2(1-p_1)}{n_{2e}}+\frac{p_1p_2}{n_{1e}n_{2e}}\right\}^{1/2}-\left\{\frac{p_1(1-p_2)}{n_{1d}}+\frac{p_2(1-p_1)}{n_{2d}}+\frac{p_1p_2}{n_{1d}n_{2d}}\right\}^{1/2}\right| \tag{23}$$

特に, $n_{1d}\to\infty$ のとき,

$$I_{dis}\leq\frac{2\{(1-p_1)(1-p_2)\}^{1/2}}{\alpha^{1/2}}\left|\left\{\frac{p_1(1-p_2)+p_2(1-p_1)}{a}+\frac{p_1p_2}{a^2}\right\}^{1/2}-\left\{\frac{p_2(1-p_1)}{a}\right\}^{1/2}\right|$$
$$(=I_{dismax}(n_{1d})) \tag{24}$$

また, $n_{2d}\to\infty$ のとき,

$$I_{dis}\leq\frac{2\{(1-p_1)(1-p_2)\}^{1/2}}{\alpha^{1/2}}\left|\left\{\frac{p_1(1-p_2)+p_2(1-p_1)}{a}+\frac{p_1p_2}{a^2}\right\}^{1/2}-\left\{\frac{p_1(1-p_2)}{a}\right\}^{1/2}\right|$$
$$(=I_{dismax}(n_{2d})) \tag{25}$$

[証明]　補題4の式(18)より式(23)が導ける．　□

系1と系2の数値計算の結果より，2つの論理式に関するサンプル数 (n_1, n_2) が異なる連言や選言の演繹推論を行う場合，演繹推論結果の確信度区間の幅は，サンプル数の少ない（確信度区間の幅の広い）論理式の影響を受け，もう一方の（確信度区間の幅の狭い）論理式のサンプル数をいくら多く採ってもある値までにしか縮まらないことがわかる．

また，系1と系2の結果からも明らかなように，連言と選言の演繹推論の場合，前提となる2つの論理式 $\Delta\Gamma\, xQ_1(x):[P_{L1}, P_{U1}](1-\alpha)$, $\Delta\Gamma\, xQ_2(x):[P_{L2}, P_{U2}](1-\alpha)$ の間には対称性があり，式(21)と式(22)は，$\Delta\Gamma\, xQ_1(x):[P_{L1}, P_{U1}](1-\alpha)$ の確信度の推定量 p_1 と $\Delta\Gamma\, xQ_2(x):[P_{L2}, P_{U2}](1-\alpha)$ の確信度の推定量 p_2 を入れ換えただけである．

[系3]　サンプル数が等しい場合と異なる場合の連鎖（または，modus ponens）の確信度区間の幅の差

帰納推論時に2つの論理式 $\Delta\Gamma\, x(Q_1(x)\to Q_2(x)):[P_{L1}, P_{U1}](1-\alpha)$ と $\Delta\Gamma\, x(Q_2(x)\to Q_3(x)):[P_{L2}, P_{U2}](1-\alpha)$ を導出するもととなるサンプル数が等しい場合 ($n_{1e}=a$, $n_{2e}=a$, $p_1=k_1/a$, $p_2=k_2/a$, $p_3=k_3/a$) と異なる場合 ($n_{1d}=a$, $n_{2d}=b$, $p_1=k_1/a$, $p_2=k_2'/b$, $p_3=k_3'/b$ または, $n_{1d}=b$, $n_{2d}=a$, $p_1=k_1'/b$, $p_2=k_2/a$, $p_3=k_3/a$) との下で導出された連鎖の結果 $\Delta\Gamma\, x(Q_1(x)\to Q_3(x)):[P_{LD}, P_{UD}](1-\alpha)$ の確信度区間の幅の差は式(26)で示される．ここで, n_1, n_2 は含意 $Q_1(x)\to Q_2(x)$, $Q_2(x)\to Q_3(x)$ おのおのに関するサンプル数

を，p_1, p_2 は論理式 $\Delta\Gamma\, x(Q_1(x)\to Q_2(x))$: $[P_{L1}, P_{U1}](1-\alpha)$ と $\Delta\Gamma\, x(Q_2(x)\to Q_3(x))$: $[P_{L2}, P_{U2}](1-\alpha)$ おのおのの確信度の推定量を表している．また，p_3 は裏の情報の論理式 $\Delta\Gamma\, x(\neg Q_2(x)\to Q_3(x))$: $[P_{L1}, P_{U1}](1-\alpha)$ の確信度の推定量を表している．

$$
\begin{aligned}
I_{dis}=\frac{2}{\alpha^{1/2}}\Bigg|&\left\{\frac{p_1^2p_2(1-p_2)+(1-p_1)^2p_3(1-p_3)}{n_{2e}}\right.\\
&+\frac{p_1(1-p_1)p_2^2+p_1(1-p_1)p_3^2}{n_{1e}}\\
&+\frac{p_1(1-p_1)p_2(1-p_2)+p_1(1-p_1)p_3(1-p_3)}{n_{1e}n_{2e}}\\
&\left.-\frac{2p_1(1-p_1)p_2p_3}{n_{1e}}\right\}^{1/2}\\
&-\left\{\frac{p_1^2p_2(1-p_2)+(1-p_1)^2p_3(1-p_3)}{n_{2d}}\right.\\
&+\frac{p_1(1-p_1)p_2^2+p_1(1-p_1)p_3^2}{n_{1d}}\\
&+\frac{p_1(1-p_1)p_2(1-p_2)+p_1(1-p_1)p_3(1-p_3)}{n_{1d}n_{2d}}\\
&\left.-\frac{2p_1(1-p_1)p_2p_3}{n_{1d}}\right\}^{1/2}\Bigg|
\end{aligned}
\tag{26}
$$

特に，$n_{1d}\to\infty$ のとき，

$$
\begin{aligned}
I_{dis}\le\frac{2}{\alpha^{1/2}}\Bigg|&\left\{\frac{p_1^2p_2(1-p_2)+(1-p_1)^2p_3(1-p_3)}{a}\right.\\
&+\frac{p_1(1-p_1)p_2^2+p_1(1-p_1)p_3^2}{a}\\
&+\frac{p_1(1-p_1)p_2(1-p_2)+p_1(1-p_1)p_3(1-p_3)}{a^2}\\
&\left.-\frac{2p_1(1-p_1)p_2p_3}{a}\right\}^{1/2}\\
&-\left\{\frac{p_1^2p_2(1-p_2)+(1-p_1)^2p_3(1-p_3)}{a}\right\}^{1/2}\Bigg|\\
&(=I_{dismax}(n_{1d}))
\end{aligned}
\tag{27}
$$

また，$n_{2d}\to\infty$ のとき，

$$
\begin{aligned}
I_{dis}\le\frac{2}{\alpha^{1/2}}\Bigg|&\left\{\frac{p_1^2p_2(1-p_2)+(1-p_1)^2p_3(1-p_3)}{a}\right.\\
&+\frac{p_1(1-p_1)p_2^2+p_1(1-p_1)p_3^2}{a}\\
&+\frac{p_1(1-p_1)p_2(1-p_2)+p_1(1-p_1)p_3(1-p_3)}{a^2}\\
&\left.-\frac{2p_1(1-p_1)p_2p_3}{a}\right\}^{1/2}\\
&+\left\{\frac{p_1(1-p_1)p_2^2+p_1(1-p_1)p_3^2}{a}\right.\\
&\left.-\frac{2p_1(1-p_1)p_2p_3}{a}\right\}^{1/2}\Bigg|\\
&(=I_{dismax}(n_{2d}))
\end{aligned}
\tag{28}
$$

modus ponens に関しても同様に導かれる．

[証明] 補題5の式(19)より式(26)が導ける． □

連鎖の場合は，連言や選言の場合とは異なり，演繹推論の前提の2つの論理式 $\Delta\Gamma\, x(Q_1(x)\to Q_2(x))$: $[P_{L1}, P_{U1}](1-\alpha)$ と $\Delta\Gamma\, x(Q_2(x)\to Q_3(x))$: $[P_{L2}, P_{U2}](1-\alpha)$ は対称ではない．すなわち，帰納推論時に $\Delta\Gamma\, x(Q_1(x)\to Q_2(x)$:$[P_{L1}, P_{U1}](1-\alpha)$, $\Delta\Gamma\, x(Q_2(x)\to Q_3(x))$: $[P_{L2}, P_{U2}](1-\alpha)$ を導出するもととなるサンプル数をおのおの n_1, n_2 としたとき，数値計算の結果より，n_1 は100位あればそれ以上採ってもあまり意味がなく，また，n_2 が少ない場合には，いくら n_1 を多く採っても，確信度区間の幅は狭まらないことがわかる．つまり，2つの論理式 $\Delta\Gamma\, x(Q_1(x)\to Q_2(x))$: $[P_{L1}, P_{U1}](1-\alpha)$ と $\Delta\Gamma\, x(Q_2(x)\to Q_3(x))$: $[P_{L2}, P_{U2}](1-\alpha)$ から $\Delta\Gamma\, x(Q_1(x)\to Q_3(x))$: $[P_{LD}, P_{UD}](1-\alpha)$ を導く連鎖の場合，後者の $\Delta\Gamma\, x(Q_2(x)\to Q_3(x))$:$[P_{L2}, P_{U2}](1-\alpha)$ の論理式の方が重要である．

この系1，2，3より保証されることを例を用いて説明する．例えば，前提の2つの論理式の一方のサンプル数が 100，他方のサンプル数が 10,000 存在し，もうそれ以上サンプルを採ることができないとする．ユーザの立場からするとサンプル数が両方とも10,000存在する場合に比べて推論結果の確信度区間の幅にどれだけの影響が出るのか不安に思われる．しかし，この系1，2，3によって，上述のそのままのサンプル数を使った場合に導出される確信度区間の幅は，両方のサンプル数が 10,000 で等しい場合に導出された確信度区間の幅に比べて，最大でも式 (20)，(23)，(26) に $a=10,000$，$b=100$ を代入することにより計算された I_{dis} の差異しかでないことを保証できる．また，逆に，両方のサンプル数が100で等しい場合に比べても最大 I_{dismax} までしか差異がでることはない．このように，実用的な面で，たとえサンプル数 n_1, n_2 に大きな差があったとしても，サンプル数 n_1, n_2 が同じ場合に導かれる結果との差異を I_{dis} まで許容すればユーザは安心してそのシステムを使用することができる．さらに，一方のサンプル数が100しかなく，もうそれ以上サンプルを採ることができず他方のサンプルはいくらでも採ることができる場合でも，他方のサンプルを 10,000 採った場合も 100,000 採った場合もあまり推論結果に差異が出ないことを述べており，サンプリングの無駄を省くことができる．

4.2 演繹推論を何段か重ねる場合について

推論を多段に重ねて論理式を導出する連鎖の推論は実用上も重要である．今回は，各論理式がその一つ前の論理式と関係があり，それ以前の論理式とは確率的に独立であるとして，連鎖の推論を何段か重ねた場合の推論法に関して考察し，数値計算を行い，連鎖の推論を一段行った場合と二段，三段と重ねた場合とを比較する．

ここで，数値計算の条件を以下に記す．

1) 推論の誤り率 α は，すべて 0.05 に固定する．
2) 定理 4 における裏の情報である $\Delta\Gamma\ x(\neg Q_{i+1}(x)\rightarrow Q_{i+2}(x)):[P_{L(2i+1)}, P_{U(2i+1)}](1-\alpha)\quad(i=1,2\cdots)$ の確信度の推定量は，すべて 0.5 とする．
3) 演繹推論の前提となる各論理式の帰納推論時におけるサンプル数は，すべて等しいものとする．

数値計算の結果の例を図 2 に示す．

ここで，演繹推論の前提となる論理式の確信度の上限，下限は，その論理式についてのサンプル数によって異なるため，論理式の[ ]内の確信度は，サンプル数によらない確信度区間の中央値で表している．また，演繹推論の結果となる論理式の[ ]内の確信度も確信度区間の中央値で示している．

〈数値計算の結果の例〉

(1) 一段の場合

$\Delta\Gamma\ x(Q_1(x)\rightarrow Q_2(x)):[0.9](0.95)$

$\underline{\Delta\Gamma\ x(Q_2(x)\rightarrow Q_3(x)):[0.8](0.95)}$

$\Delta\Gamma\ x(Q_1(x)\rightarrow Q_3(x)):[0.770](0.95)$

(2) 二段の場合

$\Delta\Gamma\ x(Q_1(x)\rightarrow Q_2(x)):[0.9](0.95)$

$\Delta\Gamma\ x(Q_2(x)\rightarrow Q_3(x)):[0.8](0.95)$

$\underline{\Delta\Gamma\ x(Q_3(x)\rightarrow Q_4(x)):[0.1](0.95)}$

$\Delta\Gamma\ x(Q_1(x)\rightarrow Q_4(x)):[0.192](0.95)$

(3) 三段の場合

$\Delta\Gamma\ x(Q_1(x)\rightarrow Q_2(x)):[0.9](0.95)$

$\Delta\Gamma\ x(Q_2(x)\rightarrow Q_3(x)):[0.8](0.95)$

$\Delta\Gamma\ x(Q_3(x)\rightarrow Q_4(x)):[0.1](0.95)$

$\underline{\Delta\Gamma\ x(Q_4(x)\rightarrow Q_5(x)):[0.2](0.95)}$

$\Delta\Gamma\ x(Q_1(x)\rightarrow Q_5(x)):[0.442](0.95)$

この結果より以下の 3 点が明らかになった．

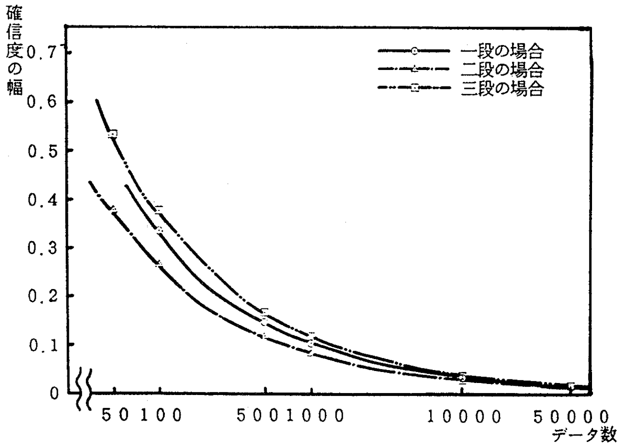


図 2 連鎖の推論を重ねる場合
Fig. 2 The range of certainty factors when the chain reasoning is used repeatedly.

1) 推論を重ねた結果得られる論理式の確信度区間の中央値が 0.5 に近いほど確信度区間の幅が広くなる傾向がある．
2) 前提となる複数の論理式の確信度の推定量が近い値の場合ほど推論を重ねた結果得られる論理式の確信度区間の幅は狭まる傾向がある．
3) $\Delta\Gamma\ x(Q_1(x)\rightarrow Q_2(x))$: $[P_{L1}, P_{U1}]\,(1-\alpha)$, $\Delta\Gamma\ x(Q_2(x)\rightarrow Q_3(x)):[P_{L2}, P_{U2}](1-\alpha), \cdots, \Delta\Gamma\ x(Q_{i+1}(x)\rightarrow Q_{i+2}(x)):[P_{L(2i+1)}, P_{U(2i+1)}](1-\alpha)\ (i=1,2,\cdots)$ から，$\Delta\Gamma\ x(Q_1(x)\rightarrow Q_{i+2}(x)):[P_{LD}, P_{UD}](1-\alpha)$ を導く場合，$\Delta\Gamma\ x(Q_1(x)\rightarrow Q_{i+2}(x)):[P_{LD}, P_{UD}](1-\alpha)$ の確信度は，連鎖の推論の前提となる論理式の中の最後の論理式 $\Delta\Gamma\ x(Q_{i+1}(x)\rightarrow Q_{i+2}(x)):[P_{L(2i+1)}, P_{U(2i+1)}](1-\alpha)$ の確信度の影響を強く受ける．

1)，2)の原因を考察すると，1)は，確信度 0.5 という値が最も真偽を決めにくく曖昧性を含んでいるためと考えられる．2)は，前提となる複数の論理式の確信度の推定量が近い値の場合ほど，推論を重ねた結果得られる論理式の確信度の分布の分散が小さくなるためと考えられる．また，3)については，$Q_1(x)\rightarrow Q_2(x)$, $Q_2(x)\rightarrow Q_3(x), \cdots, Q_{i+1}(x)\rightarrow Q_{i+2}(x)\quad(i=1,2,\cdots)$ から，$Q_1(x)\rightarrow Q_{i+2}(x)$ を導く場合，$Q_{i+1}(x)\rightarrow Q_{i+2}(x)$ が重要であるという直感と一致している．

5. む す び

本研究では，演繹推論を何段か重ねて行っても，そのシステムの信頼性を近似的に $1-\alpha$ に保証しながら推論結果を導くことができる確信度区間推論方式を提

案することができた.

また,先述の4章の2つの点について考察を行い,確信度区間推論方式のいくつかの重要な性質を明らかにした.

確信度区間推論方式は,確率世界論理という情報理論と第一階述語論理を融合した統一的な基礎理論上で論理が展開されているため,帰納・演繹推論の両方を統一的に論じることができる.さらに,本手法は,統計学の基礎的な推定量を用いているため,サンプル数を増加させることによって,優れた推論を行うことができる.

また,従来の手法では,推論結果の確信度を点推定量だけを用いて示しており,導出された確信度の信頼性については何も述べられていなかった.しかし,本手法では,目標の論理式の確信度の分布の分散に着目することにより,適当な仮定と近似の下で,演繹推論を何段か重ねて行っても導出される確信度の信頼性を $1-\alpha$ の確率で保証することができる.すなわち,本手法によりサンプル数を考慮した確信度区間の上限,下限確信度の計算方式を人工知能や数理論理学などの分野で基礎とされている汎用性のある推論規則に対し確立することができた.

さらに,上述のように,本手法は確信度に幅を持たせることにより導出される確信度の信頼性を保証したが,確信度に幅を持たせることは,知識表現を繁雑にし,かえってユーザを混乱させる結果になるのではないかという心配が生じるかもしれない.確信度は従来のように一意に示される方が望ましい場合には,推論の誤り率 α を固定して推論を行った結果,確信度区間の幅がある値 δ 以内に抑えられたときに限り確信度区間の中央値のみを出力するようなシステムにすれば良い.これは,確信度区間の幅は必ずしもユーザが知る必要はなく,そのシステム内で記憶されていれば十分だからである.例えば,推論の誤り率 $\alpha=0.05$,確信度区間の幅のしきい値 $\delta=0.1$ として,属性 Q に関してサンプル数を1,000個用意して推論を行った結果,$\Delta\Gamma xQ(x):[0.8, 0.9](0.95)$ という結果が得られたとする.この場合は,その結果はシステム内に記憶させておき,$\Delta\Gamma xQ(x):[0.85](0.95)$ のようにその確信度区間の中央値のみを出力するようなシステムを設計すれば良い.ここで,α と δ は,そのユーザにあらかじめ入力してもらうこととする.もちろん,幅を持たせた確信度の出力表現上には δ という値は存在しない.しかし,ユーザは,その出力された確信度区間の幅を見たとき,「幅が0.1くらいならば,この推論結果は有効だ.」とか,「幅が0.5もあったのでは,この推論結果では役に立たない.」というような判断を下すためのしきい値を設けていることになる.すなわち,一意に確信度を出力する場合には,α のほかに δ もあらかじめ入力しておき,その条件を満たしている場合にはその確信度区間の中央値のみを出力し,そうでない場合にはサンプル数をさらに増やしたり,その α と δ のもとでの結果はあきらめて α と δ の制限を緩めたりすれば良い.本研究の目的は,確信度に幅を持たせることではなく,そうすることによってその推論結果の信頼性を保証することである.すなわち,確信度に幅を持たせることは推論結果の信頼性を保証するための手段にすぎない.直感的な言い方をすると,確信度区間の幅の中に帰納推論時におけるサンプル数を保存しているのである.

近年,多くのエキスパートシステムが開発されるに至って,その信頼性についての疑問がますます高まってきている.本研究は,エキスパートシステムなどの推論結果の信頼性を保証するという見地からは有効な手段であると思われる.

最後に今後の課題として,以下の2つが挙げられる.本研究では,5つの演繹推論法を提案した.この5つの演繹推論法については,かなり良い近似精度で推論結果を導出することが可能である.しかし,この5つの演繹推論法が組み合わされたような推論を行う場合は大きな近似誤差を覚悟しなければならない.そこで,Horn 節形式で表現できる論理式だけに制限をした場合のすべての推論を良い近似精度で実現する推論規則を構築したい.また,その近似の精度をサンプル数 n のオーダーで示したいと考えている.

謝辞 本研究を行うにあたり,ご討論,ご助言いただいた青山学院大学・稲積宏誠先生,法政大学・西島利尚先生,大阪大学・鈴木 讓先生,早稲田大学・新家稔央氏および平澤研究室の各氏に心より感謝申し上げます.本研究の一部は文部省科学研究費(一般C No. 03832042)と早稲田大学特定課題研究(93A186)の補助による.

参考文献

1) Nilsson, N. J.: Probabilistic Logic, *Artif. Intell.*, Vol. 28, pp. 71-87 (1986).
2) Cheeseman, P.: A Method of Computing Generalized Bayesian Probability Value for Expert Systems, *Proc. 8th Int. Joint Conf. of Artif. Intell.*, pp. 198-202 (1983).

3) McCarthy, J.: Circumscription—A Form of Non-monotonic Reasoning, *Artif. Intell.*, Vol. 13, pp. 27-39 (1986).
4) Dempster, A.P.: A Generalization of Bayesian Inference, *J. of Royal Statistical Society, B*, Vol. 30, pp. 205-247 (1968).
5) Shafer, G.: *A Mathematical Theory of Evidence*, Princeton Univ. (1976).
6) 石塚 満：D-S の確率論理，電子通信学会誌，Vol. 66, No. 9, pp. 900-903 (1983).
7) Laskey, K. B. and Lehner, P. E.: Assumptions. Beliefs, and Probabilities, *Artif. Intell.*, Vol. 41, pp. 65-77 (1989/90).
8) 松山隆司：D-S の確率モデルに基づく Evidential Reasoning の論理的意味に関する考察，人工知能学会誌，Vol. 4, No. 3, pp. 340-350 (1989).
9) 松山隆司：Dempster-Shafer の確率モデルに基づくパターン分類，電子情報通信学会論文誌，Vol. J 76-D-II, No. 4, pp. 843-853 (1993).
10) 溝口，井上：D-S 理論に基づくあいまいさの処理，知識プログラミング，古川康一，溝口文雄(編)，淵 一博(監修)，7章，共立出版，pp. 157-182 (1988).
11) 大和田，溝口：不確実性を伴う証拠推論の計算方式，コンピュータソフトウェア，Vol. 8, No. 4, pp. 17-29 (1991).
12) 岡本，中島，大沢：確信度と主観確率を持つ信念推論システム，人工知能学会誌，Vol. 7, No. 2, pp. 79-86 (1992).
13) 松嶋敏泰：情報理論に基づく知識情報処理に関する研究，早稲田大学理工学研究科博士論文 (1991).
14) 松嶋，稲積，平澤：不確実性をもつ論理式の帰納推論に関する一考察，情報処理学会論文誌，Vol. 33, No. 12, pp. 1461-1475 (1992).
15) 松嶋，鈴木，稲積，平澤：不確実性様相のための確率世界論理，第 12 回情報理論とその応用シンポジウム (SITA '89) pp. 701-706 (1989).
16) 鈴木，野村，松嶋，平澤：不確実な知識の推論に関する一考察，信学技報，IT 91-50 (1991).
17) 鈴木，松嶋，平澤：確信度に幅を持つ推論方式について，信学技報，IT 92-5 (1992).
18) Feller, W.: *An Introduction to Probability Theory and Its Applications*, VOLUME 1, John Wiley & Sons (1957)；卜部，矢部，池守，大平，阿部：確率論とその応用 I，河田龍夫(監訳)，紀伊国屋書店，pp. 193-197 (1960).

付　録

以下に，連言の演繹推論についての計算方式を示した定理2と連鎖について示した定理4の導出過程を述べる．本研究では，5つの演繹推論法について論じたが，この5つの演繹推論法のいずれも3.3節で示したような手順で導かれている．

付 録 A

定理2について考える．まず，確率変数 k_1, k_2 の分布の平均と分散を求める．

［補題6］ 確率変数 k_1k_2 の分布の平均と分散

互いに独立な2つの属性 Q_1, Q_2 に関する確率変数 k_1, k_2 が，二項分布 $B(k_1:n_1, p_1^*)$, $B(k_2:n_2, p_2^*)$ に従うとき，その平均はおのおの $n_1p_1^*$, $n_2p_2^*$ であり，その分散はおのおの $n_1p_1^*(1-p_1^*)$, $n_2p_2^*(1-p_2^*)$ であるので，k_1k_2 の分布の平均と分散は次式で表される．

$$E[k_1k_2]=n_1n_2p_1^*p_2^* \tag{A.1}$$

$$V[k_1k_2]=n_1^2n_2p_1^{*2}p_2^*(1-p_2^*)+n_1n_2^2p_1^*(1-p_1^*)p_2^{*2}+n_1n_2p_1^*(1-p_1^*)p_2^*(1-p_2^*) \tag{A.2}$$

よって，k_1k_2/n_1n_2 は $Q_1(x)\wedge Q_2(x)$ が真となる確率 $p_1^*p_2^*$ の不偏推定量になっている． □

さらに，これをもとに補題7により連言の演繹推論の結果の確信度区間の上限，下限確信度が求められる．

［補題7］ 連言の演繹推論の結果の確信度区間の上限，下限確信度

n_1, n_2 個のサンプルを観測し，k_1, k_2 個が互いに独立な2つの属性 Q_1, Q_2 に関して真であったとする．このとき，$\varDelta\Gamma\ x(Q_1(x)\wedge Q_2(x):[P_{LD}, P_{UD}](1-\alpha)$ の上限確信度 P_{UD} と下限確信度 P_{LD} は次式で示される．

$$P_{UD}=\frac{k_1}{n_1}\frac{k_2}{n_2}+\frac{1}{n_1n_2}\sqrt{\frac{V[k_1k_2]}{\alpha}}$$

$$P_{LD}=\frac{k_1}{n_1}\frac{k_2}{n_2}-\frac{1}{n_1n_2}\sqrt{\frac{V[k_1k_2]}{\alpha}} \tag{A.3}$$

［証明］ チェビシェフの不等式より

$$Pr[|k_1k_2-n_1n_2p_1^*p_2^*|\leq\epsilon]\geq 1-\frac{V[k_1k_2]}{\epsilon^2} \tag{A.4}$$

ここで，α を推論の誤り率とすると

$$1-\frac{V[k_1k_2]}{\epsilon^2}=1-\alpha$$

であり，これを ϵ について解くと

$$\epsilon=\sqrt{\frac{V[k_1k_2]}{\alpha}}$$

であるので，式(A.4)の $Pr[\ ]$ の$[\ ]$内の式は

$$|k_1k_2-n_1n_2p_1^*p_2^*|\leq\sqrt{\frac{V[k_1k_2]}{\alpha}} \tag{A.5}$$

と変形することができ，これを真のパラメータ $p_1^*p_2^*$ について解くと，式(A.3)が得られる． □

補題1の式(6)を k, n の連立方程式として解き，その k, n について解かれた式を補題7の式(A.3)に

代入する．こうして，k, n が消去され，帰納推論の結果だけを用いて定理2の連言の演繹推論法を導くことができる．ここで，式(A.3)の真の分散 $V[k_1k_2]$ は実際には求められないので，式(A.2)の p_1^*, p_2^* を推定したパラメータで書き換えた統計量で真の分散 $V[k_1k_2]$ を近似している．

付　録　B

定理4についても付録Aと同様に考えられる．

[補題8] $\Delta\Gamma x(Q_1(x)\to Q_3(x)):[P_{LD}, P_{UD}](1-\alpha)$ の確信度の条件付き確率的表現

3つの論理式 $\Delta\Gamma x(Q_1(x)\to Q_2(x)):[P_{L1}, P_{U1}](1-\alpha)$, $\Delta\Gamma x(Q_2(x)\to Q_3(x)):[P_{L2}, P_{U2}](1-\alpha)$, $\Delta\Gamma x(\neg Q_2(x)\to Q_3(x)):[P_{L3}, P_{U3}](1-\alpha)$ の確信度を条件付き確率で表現するとおのおの $p_{2|1}^*, p_{3|2}^*, p_{3|\neg1}^*$ となる．このとき，$\Delta\Gamma x(Q_1(x)\to Q_3(x)):[P_{LD}, P_{UD}](1-\alpha)$ の確信度の条件付き確率的表現 $p_{3|1}^*$ は，前記の式(13)の仮定の下で，次式で表される．

$$p_{3|1}^*=p_{2|1}^*p_{3|2}^*+\{1-p_{2|1}^*\}p_{3|\neg2}^* \tag{A.6}$$

[証明] 仮定の式(13)より明らか． □

[補題9] 確率変数 $(k_1k_2+(n_1-k_1)k_3)$ の分布の平均と分散

n_1 個の個体をランダムにサンプリングし，属性 Q_1, Q_2 に関してその真偽を観測した結果，Q_1 に関して真だった下で Q_2 に関しても真である個体の数が k_1 個であったとする．また，同様に n_2 個の個体をランダムにサンプリングし，属性 Q_2, Q_3 に関してその真偽を観測した結果，Q_2 に関して真だった下で Q_3 に関しても真である個体の数が k_2 個，$\neg Q_2$ に関して真だった下で Q_3 に関しても真である個体の数が k_3 個であったとする．このとき，確率変数 k_1, k_2, k_3 は，二項分布 $B(k_1:n_1, p_1^*)$, $B(k_2:n_2, p_2^*)$, $B(k_3:n_2, p_3^*)$ に従うので，$k_1k_2+(n_1-k_1)k_3$ の平均と分散は次式で表される．

$$\begin{aligned}&E[k_1k_2+(n_1-k_1)k_3]\\&\quad=n_1n_2p_1^*p_2^*+n_1n_2(1-p_1^*)p_3^*\end{aligned} \tag{A.7}$$

$$\begin{aligned}&V[k_1k_2+(n_1-k_1)k_3]\\&\quad=n_1^2n_2\{p_1^{*2}p_2^*(1-p_2^*)+(1-p_1^*)^2p_3^*(1-p_3^*)\}\\&\qquad+n_1n_2^2\{p_1^*(1-p_1^*)p_2^{*2}+p_1^*(1-p_1^*)+p_3^{*2}\}\\&\qquad+n_1n_2\{p_1^*(1-p_1^*)p_2^*(1-p_2^*)\\&\qquad+p_1^*(1-p_1^*)p_3^*(1-p_3^*)\}\\&\qquad-2n_1n_2^2p_1^*(1-p_1^*)p_2^*p_3^*\end{aligned} \tag{A.8}$$

よって，$\frac{k_1}{n_1}\frac{k_2}{n_2}+\left(1-\frac{k_1}{n_1}\right)\frac{k_3}{n_3}$ は $Q_1(x)\to Q_3(x)$ が真となる確率 $p_{2|1}^*p_{3|2}^*+\{1-p_{2|1}^*\}p_{3|\neg2}^*$ の不偏推定量になっている．

[証明] 補題8の式(A.6)より明らか． □

[補題10] 連鎖の結果の確信度区間の上限，下限確信度

先の条件と式(13)の仮定の下で，$\Gamma x(Q_1(x)\to Q_3(x)):[P_{LD}, P_{UD}](1-\alpha)$ の上限確信度 P_{UD}, 下限確信度 P_{LD} は，推論の誤り率 α を用いて，次式で与えられる．

$$\begin{aligned}P_{UD}&=\frac{k_1}{n_1}\frac{k_2}{n_2}+\left(1-\frac{k_1}{n_1}\right)\frac{k_3}{n_3}\\&\quad+\frac{1}{n_1n_2}\sqrt{\frac{V[k_1k_2+(n_1-k_1)k_3]}{\alpha}}\\P_{LD}&=\frac{k_1}{n_1}\frac{k_2}{n_2}+\left(1-\frac{k_1}{n_1}\right)\frac{k_3}{n_3}\\&\quad-\frac{1}{n_1n_2}\sqrt{\frac{V[k_1k_2+(n_1-k_1)k_3]}{\alpha}}\end{aligned} \tag{A.9}$$

[証明] チェビシェフの不等式より

$$\begin{aligned}&Pr[|\{k_1k_2+(n_1-k_1)k_3\}\\&\qquad-n_1n_2\{p_1^*p_2^*+(1-p_1^*)p_3^*\}|\le\epsilon]\\&\quad\ge1-\frac{V[k_1k_2+(n_1-k_1)k_3]}{\epsilon^2}\end{aligned} \tag{A.10}$$

ここで，α を推論の誤り率とすると

$$1-\frac{V[k_1k_2+(n_1-k_1)k_3]}{\epsilon^2}=1-\alpha$$

であり，これを ϵ について解くと

$$\epsilon=\sqrt{\frac{V[k_1k_2+(n_1-k_1)k_3]}{\alpha}}$$

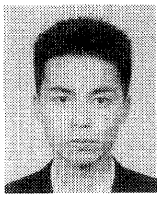
であるので，式(A.10)の $Pr[\ ]$の[ ]内の式は

$$\begin{aligned}&|\{k_1k_2+(n_1-k_1)k_3\}-n_1n_2\{p_1^*p_2^*+(1-p_1^*)p_3^*\}|\\&\quad\le\sqrt{\frac{V[k_1k_2+(n_1-k_1)k_3]}{\alpha}}\end{aligned} \tag{A.11}$$

と変形することができ，これを真のパラメータ $p_1^*p_2^*+(1-p_1^*)p_3^*$ について解くと，式(A.9)が得られる． □

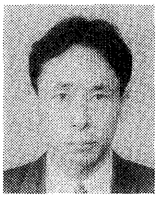
定理2の証明と同様に，補題1の式(6)を k, n の連立方程式として解き，その k, n について解かれた式を補題10の式(A.9)に代入する．こうして，定理4の連鎖の推論法が導かれる．ここで，式(A.9)の真の分散 $V[k_1k_2+(n_1-k_1)k_3]$ は実際には求められないので，推定したパラメータで書き換えた統計量で真の分散 $V[k_1k_2+(n_1-k_1)k_3]$ を近似している．

(平成5年3月31日受付)
(平成6年1月13日採録)



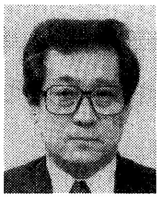
鈴木 誠（正会員）

平成3年早稲田大学理工学部工業経営学科卒業．平成5年同大学院理工学研究科修士課程修了．同年，(株)東芝入社．現在，研究開発センターシステム・ソフトウェア生産技術研究所．人工知能および知識工学の研究・開発に従事．最近は，エキスパートシステム，機械学習に興味を持つ．



松嶋 敏泰（正会員）

昭和53年早稲田大学理工学部工業経営学科卒業．昭和55年同大学院理工学研究科博士前期課程修了．同年，日本電気(株)入社．昭和61年早稲田大学大学院理工学研究科博士後期課程入学．平成3年同課程修了．平成元年横浜商科大学講師．平成4年同大学助教授．平成5年早稲田大学理工学部工業経営学科助教授，現在に至る．知識情報処理および情報理論とその応用に関する研究に従事．工学博士．電子情報通信学会，人工知能学会，情報理論とその応用学会等会員．



平澤 茂一（正会員）

昭和36年早稲田大学理工学部数学科卒業．昭和38年電気通信学科卒業．同年，三菱電機(株)入社．昭和56年早稲田大学理工学部工業経営学科教授，現在に至る．工学博士．データ伝送方式，計算機応用システムの開発，ならびに情報理論とその応用の研究に従事．昭和54年UCLA計算機科学科客員研究員．昭和60年ハンガリー科学アカデミー，伊トリエステ大学客員教授．電子情報通信学会，情報理論とその応用学会，人工知能学会等会員．