

D N A鑑定の確率計算

北 村 篤

島大法学第64巻第1・2号抜刷〔研究ノート〕

2021年3月

DNA 鑑定 of 確率計算

北 村 篤

1 はじめに

・ DNA 鑑定は、今日、刑事司法においても、犯人性認定の極めて有力な証拠 (のひとつ) となることが少なくない。そこでは、被疑者が、犯人が残した (形成に関与した) と考えられる資料 (「現場資料」ということとする。) の形成者 (の一人) として矛盾しない DNA 型をもつときに、その被疑者が犯人であることの確からしさが検討されるが、その犯人性の確からしさは、確率の問題である。その確率は、DNA 鑑定の検査結果の生物学的な知見に基づく解釈を前提とするものであるが、実務では、その確率の計算も、DNA 鑑定の専門家に依存している。公判廷で、専門家により本稿で説明する確率の計算式の導出方法が説明されることもあるが、短時間でその説明を理解することは容易ではない。しかしながら、現場資料に含まれるアレルの有無の情報だけにに基づく確率計算であれば、多少複雑ではあるが、難しいものではない。例えば、歪みのないサイコロ 2 個を同時に振ったとき、“1” のぞろ目が出る確率は $1/36$ で、“3” と “4” の目が出る確率は $1/18$ (2 個のサイコロを区別し、“3, 4” 又は “4, 3” の目が出る確率である。) であり、出目に “1” が含まれ “4”, “5”, “6” が含まれない確率は $5/36 (= (3/6)^2 - (2/6)^2)$ であることは、専門的知識の力を借りなければ判断できないようなことではないが、これと同様である。訴訟法的にいうならば、専門家による鑑定など証拠により証明しなければならないような事柄ではない。

こうした、アレルの有無の情報だけにに基づく確率計算の考え方は、例えば『DNA 鑑定とタイピング』(J.M.Butler 著、福島弘文ほか監訳、**Butler**) と略記する。) などで紹介されているが、具体的な計算方法 (計算式) は

十分には説明されていない。本稿では、その計算式とその導出方法を説明する。あわせて、その確率の値を容易に求めることができるように、現場資料や被疑者などの DNA 型を入力するだけで自動的に—本稿で説明する計算式を用いて—確率計算ができる、Microsoft の Excel のファイルを筆者が作成したサイト <https://sites.google.com/view/dnafrequency> からダウンロードできるようにした。この計算ツールの提供も本稿の大きな目的である。

- ・ DNA 鑑定は、警察で行われている常染色体上のローカスを対象とした STR 型検査法によるものを前提とし、本稿では、従来から行われてきた 15 のローカスを対象とする検査法（試薬 Identifiler を用いるもの）を想定して確率計算の方法を説明するが、その確率（出現頻度）はローカスごとに求めるので、常染色体上の 21 のローカスを対象とする検査法（試薬 GlobalFiler を用いるもの）でも、計算方法は異ならない。アレルの出現頻度は、公表されている、科学警察研究所による日本人を対象としたサンプル調査の結果（“DNA.xlsm”のシート“頻度表”）を—母集団における真の出現頻度とみなして—利用している¹。また、Excel のファイルでは、サンプル調査における出現数が 5 を下回るアレルの出現頻度²は、（必要に応じ）慎重（被疑者有利）に補正值を用いて計算できるようにしている。
- ・ なお、本稿は、確率の計算方法（数学の問題）を説明するものであり、意見にわたる部分も含め、もとより筆者の私見である。

2 数学記号の説明

(1) 確率を求めるとき、順列・組合せの数の計算が必要となるが、その計算に用いられる数学記号を説明すると、次のとおりである。

- ・ 1 から n （正の整数）までの整数の総乗を n の「階乗」というが、階乗は、 $n!$ で表され

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

であり、例えば、 $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$ である。

- ・ n 枚の異なる札（カード）の中から重複なく r 枚の札を選んで順に並

べる順列 (permutation) の数は, nPr で表され,

$${}_n P_r = n \times (n-1) \times (n-2) \times \cdots \times (n-r+1)$$

である。例えば, 1 ~ K までの 13 枚のトランプ札の中から (重複なく) 5 枚の札を選んで並べる順列の数は, ${}_{13} P_5 = 13 \times 12 \times 11 \times 10 \times 9 = 154,440$ であり, これは, 15万4,440通りの並べ方があるということである。なお, n 枚 (n 種類) の異なる札の中から重複を許して r 枚の札を選んで並べる重複順列の数は, n^r (n の r 乗) である。

- ・ n 枚の異なる札の中から重複なく r 枚の札を選ぶ組合せ (combination) の数は, ${}_n C_r$ で表され

$${}_n C_r = \frac{n!}{(n-r)! \times r!} = \frac{n \times (n-1) \times (n-2) \times \cdots \times (n-r+1)}{r!}$$

である。選ばれた r 枚の札を並べる順列の数 (r 枚の札の並べ方の数) は, $r!$ であるから,

$${}_n C_r \times r! = {}_n P_r$$

である。例えば, 1 ~ K までの 13 枚の札の中から (重複なく) 5 枚の札を選ぶ組合せの数は, ${}_{13} C_5 = {}_{13} P_5 / 5! = 154,440 / 120 = 1,287$ である。

- ・ 例えば, 3 枚の A, 2 枚の B, 1 枚ずつの D と E の合計 7 枚の札をすべて並べる順列の数は, 3 枚の A, 2 枚の B が A_1, A_2, A_3, B_1, B_2 と区別できるとしたときの順列の数を (そのように区別できるとしたときの) 3 枚の A 札を並べる順列の数と 2 枚の B 札を並べる順列の数で除して, $7! / 3! \times 2! = 420$ で求められるが, 札が並べられる位置を考えて, 3 枚の A 札の位置の組合せの数, (3 枚の A 札を配置した後の) 2 枚の B 札の位置の組合せの数, (3 枚の A 札と 2 枚の B 札を配置した後の) 1 枚の D 札の位置の組合せの数の積で, ${}_7 C_3 \times {}_4 C_2 \times {}_2 C_1 = 420$ で求めることもできる。

(2) 総和記号 Σ と総乗記号 Π の意味は, 次のとおりである。

$$\sum_{k=1}^n a_k = a_1 + a_2 + \cdots + a_n$$

$$\prod_{k=1}^n a_k = a_1 \times a_2 \times \cdots \times a_n$$

- (3) アレルの出現頻度は、 $p(\{a_i\})$ (a_i の出現頻度)、 $p(\{10\})$ (“10” の出現頻度) などと表す。アレル a_1, a_2, \dots, a_i の各出現頻度を、 p_1, p_2, \dots, p_i で略記することもある。

また、例えば、 a_1, a_2, a_3 のいずれかである確率（出現頻度）を $p(\{a_1, a_2, a_3\})$ で表す。アレルの型は、互いに排反であるので、 $p(\{a_1, a_2, a_3\}) = p(\{a_1\}) + p(\{a_2\}) + p(\{a_3\})$ である。

- (4) 指数表記をEを用いて $5.67 \text{ E-}23$ ($= 5.67 \times 10^{-23}$) の形で表記する。

3 DNA 鑑定の証拠としての意味など

- (1) 本稿での確率計算の前提

DNA 鑑定の検査では、アレルは、量的情報（検査結果を表すエレクトロフェログラムに現れるピーク高やピーク面積）の形で検出されるが、本稿で説明する確率計算は、そうした量的情報を捨象した、アレルの有無（現場資料にそのアレルが含まれるか否か）の情報だけに基づくものである。例えば、あるローカスで、被害者と犯人との2人に由来する（2人の混合資料である）と認められる現場資料のDNA型が“11, 12, 13”で、被害者は“11, 12”であるケースを例にすると、現場資料の3つのアレルのうち“12”のピークがほかのアレルと比べてかなり大きく、犯人のDNA型は“12, 13”の可能性が高いと考えられるとしても、そのことは考慮せずに、そのDNA型は“11,13”, “12,13”, “13,13”のいずれかであるということだけに基づいて確率を求めるということである³。

また、現場資料のDNA鑑定では、検出されないアレルが含まれている可能性があることもある。本稿では、混合資料は、この可能性が否定されるという前提での確率計算の方法を説明するが、この可能性が否定されな

いのであれば、上記の設例のケースでいうと、犯人は（“13”のアレルをもたなければならないが）“11”, “12”, “13” 以外のアレルをもつ可能性があることを前提に確率計算（下記5(2)のような計算など）をしなければならない。

(2) 形成者として矛盾しない DNA 型をもつという資格

（現場資料と被疑者の）DNA 型が被疑者の犯人性の証拠となるのは、被疑者が現場資料の形成者（の一人）として矛盾しない DNA 型をもつ場合であり、また、形成者として矛盾しない DNA 型をもつことを現場資料の形成者の「資格」と表現すると、その資格をもつ者は限られているので、犯人性の証拠となる。例えば、上記の設例のケースでいうと、“13”のアレルをもたない者や“11”, “12”, “13” 以外のアレルをもつ者は犯人たり得ず、犯人は“11,13”, “12,13”, “13,13” のいずれかの DNA 型をもつという形成者の資格をもつ者でなければならないが、その資格をもつ者は限定されるから、（被疑者が）その資格をもつことは犯人性の積極証拠となる。

その証拠価値は、現場資料の形成者の資格（犯人資格である。）をもつ人の割合（何人に一人の割合で存在するか）で定まる。その割合は、その DNA 型（の資格をもつ人）の出現頻度であり、また、極めて多数の人の集団（集合）の中から任意（無作為）に抽出された人がその DNA 型の資格をもつ確率と同じであって、アレルの出現頻度により求めることができる。例えば、あるローカスで DNA 型が“10,10”である人の割合は、 $p(\{10\})^2$ （アレル“10”の出現頻度の2乗）である。さらに、ローカスごとの DNA 型（の出現頻度）は相互に独立であると考えられる。すなわち、あるローカスである DNA 型をもつ人の割合は、ほかのローカスにおける DNA 型如何によって左右されず、例えば、ローカス「A」で“11,12”の DNA 型をもつ人の割合は、ローカス「B」で“20,21”である人の集合でも“22,23”である人の集合でも、変わらない。歪みのないサイコロ1個を2回振ったとき、2回目に“1”の目が出る確率は、1回目の出目に影響されないのと同じである。したがって、複数のローカスで形成者の資格を

もつ人の割合は、それぞれのローカスで資格をもつ人の割合の総乗で求められる。その割合が非常に小さいと、その資格をもつ被疑者は、僅かな割合の人だけがもつ資格（レアな属性）をもつ者として、一卵性の多胎児がない限り、犯人であると強く推認されることになる。

現場資料の形成者としての資格は、現場資料の DNA 型から導かれるが、例えば、犯人一人に由来すると認められる現場資料の DNA 型が3つのローカスで全く不明であるとき、それらのローカスで形成者の資格をもつ者は、すべての人であって、その割合（出現頻度）は1（100%）であり、DNA 型の証拠価値は、残りの12のローカスで形成者としての資格をもつ人の割合で定まる（下記5(3)参照）。この場合、12のローカスで形成者としての資格をもつ被疑者は、現場資料の DNA 型が不明であるローカスで、現場資料と異なる DNA 型をもつ可能性があるが、そのことは、確率的には（12のローカスにおける）DNA 型の証拠価値に影響しない。それは、検査対象外の数多あるローカスで DNA 型が一致しない可能性に影響されない（検査対象のローカスで形成者の資格をもつ人の割合が非常に小さいとき、その可能性を考慮しなければならない現実的な必要性はない。）のと、確率の問題としては、同じである。また、上記の設例のケースでは、現場資料の DNA 型から導かれる犯人の DNA 型は、“11,13”、“12,13”、“13,13”のいずれかであるという、ある範囲のものであり、犯人は被疑者と異なる DNA 型をもつ者である可能性がある（被疑者は“12,13”で、犯人は“11,13”又は“13,13”である可能性がある。）が、同様に、このことは、アレルの有無だけで考えると、形成者として矛盾しない DNA 型をもつという犯人資格をもつ人の割合が小さいことによる犯人性の証拠価値を減殺するものではない。それは、（現実的な喩えではないが）犯人は成人で身長が140cm～145cm（非常にレアであるとする。）であり、被疑者（成人）の身長がその範囲内にある143cmであるとき、その範囲で身長が一致するという事実は、その限度で、-真犯人の身長は140cmや145cmで1桁の数値で一致しない想定上の可能性があるとしても-、犯人性を推

認させる証拠価値をもつものと同じであると考えられる。

(3) 尤度比

現場資料の形成者として矛盾しない DNA 型をもつ者が形成者であること（犯人性）の確からしさは、上記(2)のように DNA 型の資格をもつ人の割合で考えるほか、形成者として矛盾しない DNA 型をもつ者（被疑者）が形成者（の一人）であるという仮説と、任意の 1 人の人が形成者（の一人）であるという仮説を想定し、それぞれの仮説の下で現場資料と一致する DNA 型が形成される各確率の比（**尤度比**：LR：Likelihood Ratio）⁴で考えることもできる。例えば、被害者と犯人との 2 人に由来すると認められる現場資料の DNA 型が被害者と被疑者で形成されたとして矛盾のないものであるケースを例とする。このとき、前者の確率は、アレルの有無だけで考えると、1（100%）である。後者の確率は、被害者と任意の 1 人の人との混合で現場資料と一致する DNA 型が形成される確率であるが、そのローカスごとの確率は、上記の設例のケースでは、アレルの有無だけで考えると、任意の人が“11,13”，“12,13”，“13,13”のいずれかの DNA 型をもつ確率である。そして、前者の確率が後者の確率（ローカスごとに求めた確率の総乗）より大きいとき、前者の仮説（被疑者がもう一人の形成者であるという仮説）は、後者の仮説（任意の 1 人の人がもう一人の形成者であるという仮説）よりも、尤もらしい（可能性が高い）ということになる。ただし、これは、任意の 1 人の人が形成者であるという仮説との対比であって、被疑者以外の者が形成者である（被疑者以外の多数の者の中に形成者がいる）という仮説との対比ではない。

(4) 犯人確率の計算

蛇足的であるが、犯人資格の DNA 型をもつ被疑者が犯人であることの確からしさを数値的に評価するとすれば、以下のような計算が考えられる。その計算の前提として、犯人は N 人の集団の中の一人であるとする。

- ・ 犯人資格の出現頻度 p が小さいとき、例えば、4.7兆人に 1 人の割合であるときでも、その資格をもつ者は、被疑者一人に限られるというこ

とはならない。 N 人の集団の中に被疑者以外にその資格をもつ者が存在する確率は、

$$1 - \frac{Np \times (1-p)^{N-1}}{1 - (1-p)^N} = 1 - \frac{(1-p)^{N-1}}{(1 - (1-p)^N)/Np}$$

であると考えられる。仮に、被疑者以外の $(N-1)$ 人の各人が犯人資格の DNA 型をもつ各確率を p とすると、 $(N-1)$ 人の中に犯人資格の DNA 型をもつ人が存在する確率は、

$$1 - (1-p)^{N-1}$$

となる⁵が、被疑者以外に犯人資格をもつ者が存在する確率は、これより小さい。この式は、展開・整理すると、

$$= (N-1)p - {}_{N-1}C_2 p^2 + {}_{N-1}C_3 p^3 \cdots \cdots - {}_{N-1}C_{N-1} (-p)^{N-1}$$

であり、 p が N の逆数に比して十分に小さい (Np が 1 に比して十分に小さい) とき、最初からの数項で近似計算できる。その 1 次近似値は、 $(N-1)p \cong Np$ である。

・ 犯人資格をもつ者が犯人である確率は、複数の中から選び出したボールが唯一の当たりのボールである確率と同様のものとして、次のように求めることもできると思われる。

ここで説明する確率に関する数学記号を説明する。ある事象 X が発生する確率を $P(X)$ で表す。 X と Y の 2 つの事象が同時に発生することを $X \cap Y$ で表し、その確率を $P(X \cap Y)$ で表す。また、 X が発生するという条件下で Y が発生する確率 (条件付き確率) を $P(Y/X)$ で表す。

ここでは、選別されたある者 (被疑者) が犯人であることを A とし、その者以外の $(N-1)$ 人の各人が犯人であることをそれぞれ B_1, B_2, \dots, B_{N-1} で表す。

まず、犯人に関する情報は全くないとすると、被疑者が犯人である確率は、 $P(A) = 1/N$ である。

次に、被疑者の DNA 型は既知であるとし、被疑者が形成者であるときに (単独で又は被害者などとの混合で) 形成される DNA 型を α 型で

表し、現場資料の DNA 型が α 型であることを D とする。また、犯人資格の出現頻度を p とする。そして、事象 D が生じているという情報 (現場資料の鑑定結果) が得られたとする。このとき、被疑者が犯人である確率は、ベイズの定理⁶により

$$P(A/D) = \frac{P(A)P(D/A)}{P(A)P(D/A) + \sum_{i=1}^{N-1} (P(B_i)P(D/B_i))}$$

で求められる。なお、 $\sum_{i=1}^{N-1} (P(B_i)P(D/B_i)) = P(B_1)P(D/B_1) + P(B_2)P(D/B_2) + \dots + P(B_{N-1})P(D/B_{N-1})$ である。 $P(A) = 1/N$, $P(B_i) = 1/N$ ($i=1,2,3, \dots, N-1$) である。 $P(D/A)$, $P(D/B_i)$ は、それぞれ、被疑者が犯人であるとき、それ以外の各人が犯人であるときに、現場資料の DNA 型が α 型である (その旨の鑑定結果が得られる) 確率 (尤度) である。 $P(D/A) = 1$ である。① $P(D/B_i)$ は、

$$\frac{\binom{Np}{1 - (1-p)^N} - 1}{N-1}$$

であると考え、被疑者が犯人である確率は

$$P(A/D) = \frac{1 - (1-p)^N}{Np}$$

であり、逆に被疑者が犯人ではない確率は

$$1 - \frac{1 - (1-p)^N}{Np} \quad 7$$

となる。②仮に、

$$P(D/B_i) = p$$

であるとすると

$$P(A/D) = \frac{1}{1 + (N-1)p}$$

となるが、 N 人の集団の中には (必ず) 犯人資格の DNA 型をもつ人がいるので、 $P(D/B_i)$ は p より小さく、被疑者が犯人ではない確率は

$$1 - \frac{1}{1 + (N-1)p}$$

より小さい。この式も、 p が N の逆数に比して十分に小さいとき、1次近似値は $(N-1)p \cong Np$ である。

これらの計算の値は、“犯人確率.xlsm”で一応求めることができるが、被疑者が犯人ではない確率は、目的的には、 p が N の逆数に比して十分に小さいとき、 Np である。

4 各種の確率計算

被疑者「甲」が現場資料の形成者（の一人）として矛盾しない DNA 型（犯人資格）をもつとき、その者が形成関与者であること（犯人性）の確からしさにかかわるものとして、いくつかの確率計算が考えられる。

まず、次の5項では、犯人一人に由来すると認められる現場資料（単独犯の犯人が残した精子など）のケース（**単独資料のケース**）で、

犯人資格をもつ人の出現頻度

を求める。

次に、2人以上の者で形成された（可能性がある）現場資料のケース（**混合資料のケース**）を扱うが、6項(1)では、基本的に、現場資料は2人の人で形成されたとして、（次に述べる、ある者「乙」の形成関与を前提とせずに）

犯人資格をもつ人の出現頻度 ((1)a)

被疑者「甲」の形成関与の尤もらしさ－尤度比 ((1)b)

を求める。なお、6項(1)a前段の出現頻度は、形成関与者の数が2人であることを前提としないものである。6項(2)では、ある者「乙」（被害者など）と犯人との2人に由来すると（DNA型以外の証拠によって）認められる現場資料（性犯罪の単独犯の犯人が触れた被害者の衣服で覆われている部位から採取された現場資料など）のケースで、「乙」の形成関与（とそのDNA型）を前提として、

犯人資格をもつ人の出現頻度 ((2)a)

被疑者「甲」の形成関与の尤もらしさー尤度比 ((2)b)

を求める。また、6項(3)では、2人の犯人に由来する認められる現場資料(2人の犯人の精子が混合したものなど)のケースで、

犯人資格をもつ2人ペアの出現頻度

を求める。

さらに、7項では、**現場資料の形成関与者の数の推定**を考える。主として、2人で形成可能な現場資料(すべてのローカスでアレルの種類数が4以下であるもの)が3人以上の人で形成される可能性の程度を考えるが、1人で形成可能な現場資料(すべてのローカスでアレルの種類数が2以下であるもの)が2人以上の人で形成される可能性の程度も考える。7項(1)では、 n ($n=2, 3, \dots$) 人混合で**すべてのローカスでアレルの種類数が4(又は2)以下となる確率**を求め、7項(2)では、ローカスごとの**アレルの種類数の合計数による推定**を考えるが、-現場資料のDNA型が与えられているときにアレルの種類数だけに基づいて推定を行う意味はないので、7項(3)で、 n ($n=2, 3, \dots$) 人混合で**現場資料と一致するDNA型が形成される確率**を求める。さらに、7項(4)では、現場資料が**3人の人で形成された**として、ある者「乙」の形成関与を前提とし、**被疑者「甲」の形成関与の尤もらしさー尤度比**を求める。

なお、これらの確率計算は、筆者のサイトからダウンロードできる、数式などを設定している Excel ファイル (“DNA.xlsx”の赤線で囲ったセル範囲)にDNA型を入力するだけで、その結果を得られるが、計算方法の説明とともに、そのファイル名・シート名を掲記する。また、計算量が膨大になる計算は、Java によるプログラムを作成して行った(その実行方法は“Java ファイルの実行方法 .pdf”参照)が、そのプログラムもソースコードとともにダウンロードすることができる。

5 単独資料のケース - “DNA.xlsx”のシート“犯人単独” -⁸

ここでは、犯人一人で形成されたと認められる現場資料(犯人の精子など)のDNA型が被疑者のDNA型と矛盾しないケースを考える。

- (1) 典型的には、現場資料と被疑者の DNA 型が〔第1図〕のように一致しているケースである。

第1図

ローカス					出現頻度
D8S1179	11	0.1089	13	0.2248	0.0490
D21S11	28	0.0422	30	0.3370	0.0284
D7S820	10	0.2189			0.0479
CSF1PO	9	0.0467	11	0.2078	0.0194
D3S1358	14	0.0289	16	0.3063	0.0177
TH01	6	0.2233	7	0.2667	0.1191
D13S317	10	0.1148	13	0.0511	0.0117
D16S539	10	0.1967	12	0.1785	0.0702
D2S1338	19	0.2089	24	0.1078	0.0450
D19S433	13	0.2874	14	0.0881	0.0506
vWA	18	0.2256			0.0509
TPOX	11	0.3630	12	0.0356	0.0258
D18S51	12	0.0478	14	0.2222	0.0212
D5S818	10	0.2007			0.0403
FGA	20	0.0889	22	0.2015	0.0358
					2.07E-22
					4.83E+21

まず、D8S1179 のローカスをみると、その DNA 型は、“11,13”であり、“11”と“13”のヘテロ接合であるが、これは、父親・母親に由来の各アレルが①“11”,“13”であるか、②“13”,“11”であるということの意味している。父親由来のアレルと母親由来のアレルをこの順に表記する順列を考えると、①“11,13”又は②“13,11”である。①である人の出現頻度（任意の人が①の DNA 型をもつ確率）も、②である人の出現頻度も、 $p(\{11\}) \times p(\{13\})=0.0245$ である。したがって、D8S1179 のローカスで“11,13”の DNA 型をもつ人の出現頻度は、 $2 \times p(\{11\}) \times p(\{13\})=0.0490$ である。これは、その DNA 型をもつという資格（犯人資格）をもつ人は、20.42 (=1/0.0490) 人に 1 人の割合で存在するということである。2つ目の D21S11 のローカスについても同様である。一般化すると、 a_1, a_2 のヘテ

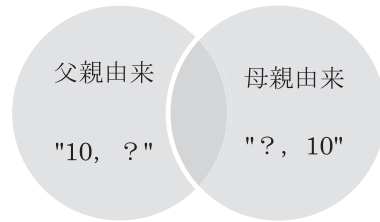
ロ接合の DNA 型をもつ人の出現頻度は、 $2 \times p(\{a_1\}) \times p(\{a_2\})$ である。

3つ目の D7S820 のローカスの DNA 型は、“10” のホモ接合である。これは、父親・母親に由来のアレルがいずれも“10”である（順列では“10,10”）ということであり、その DNA 型をもつ人は、出現頻度は $p(\{10\}) \times p(\{10\}) = 0.0479$ で、20.87 (=1/0.0479) 人に 1 人の割合で存在する。一般化すると、 a_1 のホモ接合の DNA 型をもつ人の出現頻度は、 $p(\{a_1\})^2$ である。

こうして、ローカスごとに出現頻度を求めることができるが、15 のローカスで〔第 1 図〕の DNA 型をもつという犯人資格をもつ人は、① D8S1179 で“11,13”の DNA 型をもつという資格をもち、かつ、② D21S11 で“28,30”の DNA 型をもつという資格をもち、かつ、……、かつ、⑮ FGA で“20,22”の DNA 型をもつという資格をもつ人であるから、その出現頻度は、①～⑮の出現頻度の総乗で得られ、 $2.07 \text{ E-}22$ (2.07×10 のマイナス 22 乗) であり、約 48 垓人 (1 垓は 1 京の 1 万倍) に 1 人の割合でしか存在しないということになる。

- (2) 上記(1)では、例えば、現場資料の D7S820 のローカスは、“10”のホモ接合であるとして確率計算を行ったが、あるローカスの DNA 型が部分的に不明であるとき、例えば、アレル“10”が含まれているということしかいえず、“10”とそのほかのアレルとのヘテロ接合である可能性もあるとき（いわば“10,?”の DNA 型であるとき）は、その DNA 型をもつという資格（アレル“10”をもつという資格）をもつ人の出現頻度は、次のとおりである。

“10”のアレルが含まれるというの、①父親由来のアレルが“10”であるときと、②母親由来のアレルが“10”であるときがあるが、“10,10”の人は①にも②にも該当するので、D7S820のローカスで“10”をもつ人の出現頻度



は、 $2 \times p(\{10\}) - p(\{10\})^2 = 0.3899$ (2.56人に1人の割合)である。一般化すると、アレル a_1 をひとつはもつ人の出現頻度は、 $2 \times p(\{a_1\}) - p(\{a_1\})^2$ である。

- (3) 歪みのないサイコロの出目の確率は、1～6のいずれの目でも(1/6であって)等しいが、アレルの出現頻度はそうではない。例えば、D8S1179のローカスで“11,13”のDNA型をもつ人の出現頻度は、 $2 \times 0.1089 \times 0.2248 = 0.0490$ であるが、このローカスで出現頻度が大きいのは、“13”, “14”のアレルであり、“13,14”である人の出現頻度は、 $2 \times p(\{13\}) \times p(\{14\}) = 0.2248 \times 0.2048 = 0.0921$ である。ローカスごとに、最も出現頻度が高いDNA型は、[第2図]のとおりであり、15のローカスでこれらのDNA型をもつ人(いわば最もポピュラーな人である。)の出現頻度は約4.7兆人に1人の割合

第2図

ローカス		
D8S1179	13	14
D21S11	29	30
D7S820	11	12
CSF1PO	10	12
D3S1358	15	16
TH01	7	9
D13S317	8	11
D16S539	9	10
D2S1338	18	19
D19S433	13	14
vWA	17	18
TPOX	8	11
D18S51	13	14
D5S818	11	12
FGA	22	23

(“DNA.xlsx”のシート“犯人単独(最高頻度)”)である。したがって、15のローカスであるDNA型をもつ人((2)の部分的不明もないとする。)の出現頻度は、約4.7兆人に1人の割合以下であるので、15のローカスのDNA型を調べることで、人の識別は高度に可能であるといえる。なお、21の

ローカスを対象とする検査法（GlobalFiler）では、最も高い DNA 型の出現頻度は、約565京分の1となる。

他方、15 のローカスのすべてで DNA 型が一致していることは、出現頻度（確率）的には、重要ではない。例えば、〔第3図〕のように、現場資料の DNA 型が CSF1PO, D2S1338, D18S51 の3つのローカスで不明であっても、12 のローカスで現場資料と一致（*の付された2つのローカスでは(2)の部分的不明もあるとしている。）する DNA 型をもつ人の出現頻度は、 4.21×10^{-17} （1京人以上に1人の割合）であり、最もポピュラーな人のケースで15の

第3図

ローカス			出現頻度
D8S1179	11		0.0119
D21S11	28	30	0.0284
D7S820	10	*	0.3899
CSF1PO			1.0000
D3S1358	14	16	0.0177
TH01	6	*	0.3967
D13S317	10	13	0.0117
D16S539	10	12	0.0702
D2S1338			1.0000
D19S433	13	15	0.0291
vWA	18		0.0509
TPOX	11	12	0.0258
D18S51			1.0000
D5S818	10		0.0403
FGA	20	22	0.0358
			4.21E-17

ローカスの DNA 型が一致するという偶然の重なりが生じる確率より相当に小さく、犯人性の積極証拠としての価値はより大きい。

6 混合資料のケース

この項では、現場資料が2人（以上）の者で形成された（可能性のある）混合資料であるケースで、形成者（の一人）として矛盾しない DNA 型をもつ者「甲」（被疑者）が形成者であること（犯人性）の確からしさを考える。

(1) ある DNA 型をもつ者（被害者など）の形成関与を前提としないとき

ここでは、〔第4図〕のように、「甲」（被疑者）が現場資料の形成者として矛盾しない DNA 型（網掛けを付した DNA 型）をもつケースで、ある DNA 型をもつ者「乙」（被害者など）のもう一人の形成者としての関与の事実を前提とせず、「甲」が形成者として矛盾しない DNA 型をもつこと（矛盾して排除されることがないこと）だけに基づいて、「甲」が形成者であること（犯人性）の確からしさを考える。

もう一人の形成者の DNA 型が不明であるケース（例えば、現場資料が一人の犯人と誰であるか全く不明である者との混合で形成されたと考えられるケースや、現場資料が2人の犯人の混合で形成されたと認められるが、形成者として矛盾しない DNA 型をもつ被疑者「甲」一人しか犯人として浮上していないケースなど）では、ここで述べる方法で、「甲」の形成関与の確からしさを考えることになる。また、現場資料の DNA 型が「甲」と「乙」

第4図

ローカス	DNA型			
D8S1179	11	12	13	14
D21S11	29	30		
D7S820	10	12	13	
CSF1PO	9	11	12	
D3S1358	14	15	18	
TH01	6	7	8	
D13S317	11	13		
D16S539	9	11	12	
D2S1338	17	18	20	23
D19S433	13	14	16	
vWA	17	18		
TPOX	11	12	13	
D18S51	13	15	16	17
D5S818	10	12		
FGA	18	21	22	24

で形成されたとして矛盾はないが、「乙」が形成者の一人であると（現場資料の採取状況などからは）認定できないケースでも、ここで述べる方法で、「乙」の形成関与を前提とせずにその確からしさを考えることができる。さらに、下記 a 前段の計算は、現場資料の形成関与者の数を前提としないものであるので、現場資料が2人の者で形成されたと前提できないケースで「甲」の形成関与の確からしさを考えるのにも役立つ。

a 出現頻度 - “DNA.xlsm”のシート“混合DNA（非排除）”⁹

現場資料の形成者の一人である資格は、それぞれのローカスで、現場資料の DNA 型に含まれるアレルのみをもつことである。例えば、〔第4図〕の D8S1179 のローカスであれば、形成者の資格は、“11”、“12”、“13”、“14”のアレルのみをもつことであり、その資格をもつ人の出現頻度は、 $p(\{11,12,13,14\})^2$ であり、D21S11 のローカスでは $p(\{29,30\})^2$ である。一般化すると、現場資料の DNA 型に含まれるアレルの出現頻度の総和 $\sum p_i$ の2乗 $(\sum p)^2$ である。こうして、15 のローカスごとに、形成者の資格をもつ人の出現頻度を求めて、それらの総乗を求めると、

同図の現場資料の形成者の資格をもつ人の出現頻度は、 3.29×10^{-9} で、約3億人に1人の割合となる。ここでの確率計算は、「甲」(被疑者)が形成者として矛盾しないDNA型をもつことだけに基づくものであり、その意味で、「甲」の犯人性を控え目に(被疑者有利に)見積もることになるものであるが、約3億人に1人の割合というのは、僅かな割合であり、「甲」(被疑者)の犯人性の証拠として相応の価値をもつとも考えられる。

ローカス	2人混合	非前提
D8S1179	0.3177	0.4371
D21S11	0.3407	0.3407
D7S820	0.2386	0.2386
CSF1PO	0.4529	0.4529
D3S1358	0.2382	0.2382
TH01	0.3095	0.3095
D13S317	0.0741	0.0741
D16S539	0.5172	0.5172
D2S1338	0.1913	0.2587
D19S433	0.4124	0.4124
vWA	0.2583	0.2583
TPOX	0.1604	0.1604
D18S51	0.2397	0.3301
D5S818	0.1900	0.1900
FGA	0.1780	0.2608
	8.76×10^{-10}	3.29×10^{-9}

上記の計算は、現場資料の形成者関与者の数を前提としないとき(2人で形成されたととは限らないとしたとき)のものであり、現場資料が2人の者で形成されたとすると、計算は異なる。すなわち、2人の者で形成されたとすると、4種類のアレルが検出されているローカスでは、形成者のDNA型は、いずれも、それらのアレルから成るヘテロ接合でなければならず、例えば、[第4図]のD8S1179のローカスを例にすると、形成者であるためには、“11”、“12”、“13”、“14”のアレルのみをもち、かつ、ヘテロ接合である(ホモ接合ではない)ことが必要であり、その出現頻度は、 $p(\{11,12,13,14\})^2 - p(\{11\})^2 - p(\{12\})^2 - p(\{13\})^2 - p(\{14\})^2$ である。一般化すると、 $(\sum p_i)^2 - \sum (p_i^2)$ である。なお、現場資料のアレルの種類数が3以下であるローカスでは、上記のとおり、 $(\sum p_i)^2$ である。これにより、2人の者で形成されたとして、同図の現場資料の形成者の資格をもつ人の出現頻度を求めると、 8.76×10^{-10} であり、約11億人に1人の割合となる。

b 尤度比 – “DNA.xlsm”のシート“混合DNA (2人)” –¹⁰

- ・ 現場資料の形成者として矛盾しない DNA 型をもつ「甲」(被疑者)が形成者であること(犯人性)の確からしさは、現場資料が2人の者で形成されたと前提したとき、①「甲」と「任意の1人の人」との混合で形成されたという仮説の下で、現場資料と一致する DNA 型が形成される確率(「 $A\alpha$ 確率」ということとする。)と、②「任意の2人の人」の混合で形成されたという仮説の下で、現場資料と一致する DNA 型が形成される確率(「 $\alpha\alpha$ 確率」ということとする。)の比(尤度比)で考えることもできる。ここでは、形成関与者のもつアレルは現場資料の DNA 鑑定ですべて検出されているとして、この尤度比を求める。
- ・ 先に、 $A\alpha$ 確率と $\alpha\alpha$ 確率を求める計算式を記すと、現場資料の DNA 型に含まれるアレルの種類の数又は「甲」のもつアレルを除いた「非重複」のアレルの種類の数に応じ、次のとおりである。

$A\alpha$ 確率 – シート“混合DNA (2人)”の $X\alpha$ 確率及び $Y\alpha$ 確率 –
[非重複のアレル]

$$a_1, a_2 \text{ のとき } 2 \times \prod p_i = 2 p_1 p_2 \quad (\text{①式})$$

$$a_1 \text{ のとき } (\sum p_i)^2 - (\sum p_i - p_i)^2 = 2 \times p_i \times \sum p_i - p_i^2 \quad (\text{②式})$$

$$\text{ないとき } (\sum p_i)^2 \quad (\text{③式})$$

$\alpha\alpha$ 確率 – シート“混合DNA (2人)”の $\alpha\alpha$ 確率 –
[現場資料のアレル]

$$a_1, a_2, a_3, a_4 \text{ のとき } 24 \times \prod p_i = 24 \times p_1 \times p_2 \times p_3 \times p_4 \quad (\text{④式})$$

$$a_1, a_2, a_3 \text{ のとき } 12 \times \prod p_i \times \sum p_i \\ = 12 \times (p_1 \times p_2 \times p_3) \times (p_1 + p_2 + p_3) \quad (\text{⑤式})$$

$$a_1, a_2 \text{ のとき } (\sum p_i)^4 - \sum p_i^4 = 4p_1^3 p_2 + 6p_1^2 p_2^2 + 4p_1 p_2^3 \quad (\text{⑥式})$$

$$a_1 \text{ のとき } p_i^4 \quad (\text{⑦式})$$

- ・ $A\alpha$ 確率の計算式(①~③式)の導出は、下記(2)aで説明する。
 $\alpha\alpha$ 確率は、ローカスごとに、観念的に、2人の人を(同時にではなく)順に抽出し、父親由来と母親由来のアレルを区別できるとして、そ

の 2 人のアレルを「1 人目の父親由来, 1 人目の母親由来, 2 人目の父親由来, 2 人目の母親由来」の順に並べる順列 (例えば, 2 人の DNA 型が父親由来と母親由来のアレルを区別して “11,12”, “13,11” であるときは, “11,12,13,11” の順列) を想定することで, 次のとおり, 現場資料の DNA 型に含まれるアレルの種類数に応じ, 上記の④~⑦式で求めることができる。

(現場資料のアレルの種類数が「4」であるとき)

4 種類のアレル a_1, a_2, a_3, a_4 を 1 つずつ並べる順列の数は, ${}_4P_4=24$ であり, ひとつひとつの順列の出現頻度は, いずれも, $p_1 \times p_2 \times p_3 \times p_4$ であるから, $\alpha \alpha$ 確率は

$$24 \times \prod_{i=1}^4 p_i = 24 \times p_1 p_2 p_3 p_4 \quad (\text{④式})$$

である。

(現場資料のアレルの種類数が「3」であるとき)

3 種類のアレル a_1, a_2, a_3 のうち, 1 種類のアレルだけ 2 つ含まれるが, それが a_1, a_2, a_3 のいずれであるときも, それぞれ, 順列の数は, $4!/2!=12$ であり, また, ひとつひとつの順列の出現頻度は, 2 つ含まれるアレルを a_i とすると, $p_1 \times p_2 \times p_3 \times p_i$ である。したがって, $\alpha \alpha$ 確率は

$$12 \times \prod_{i=1}^3 p_i \times \sum_{i=1}^3 p_i = 12 \times (p_1 p_2 p_3) \times (p_1 + p_2 + p_3) \quad (\text{⑤式})$$

である。

(現場資料のアレルの種類数が「2」であるとき)

4 つのアレルはすべて a_1 又は a_2 であり, かつ, 4 つとも a_1 又は a_2 である順列は除かれる。前者の出現頻度は, $(p_1+p_2)^4$ であり, 後者の出現頻度は, $p_1^4+p_2^4$ であるから, $\alpha \alpha$ 確率は

$$\left(\sum_{i=1}^2 p_i\right)^4 - \sum_{i=1}^2 p_i^4 = (p_1 + p_2)^4 - p_1^4 - p_2^4$$

$$= 4p_1^3p_2 + 6p_1^2p_2^2 + 4p_1p_2^3 \quad (\text{⑥式})$$

である。

(現場資料のアレルの種類数が「1」であるとき)

4つのアレルはすべて a_1 であり、 $\alpha\alpha$ 確率は

$$p_1^4 \quad (\text{⑦式})$$

である。

上記の計算式により、15 のローカスごとに $A\alpha$ 確率を求めて、それらの総乗で、「甲」と任意の1人の人との混合で現場資料と一致する DNA 型が形成される確率が得られるが、現場資料と「甲」の DNA 型が〔第4図〕のとおりであるとき、その確率は、2.65 E-18 である。同様に、任意の2人の人の混合で現場資料と一致する DNA 型が形成される確率は、ローカスごとの $\alpha\alpha$ 確率の総乗で

ローカス	$\alpha\alpha$ 確率	$A\alpha$ 確率
D8S1179	0.0148	0.0551
D21S11	0.0995	0.2798
D7S820	0.0105	0.0328
CSF1PO	0.0328	0.3882
D3S1358	0.0042	0.0274
TH01	0.0264	0.1191
D13S317	0.0031	0.0741
D16S539	0.1019	0.2249
D2S1338	0.0058	0.0466
D19S433	0.0040	0.0036
vWA	0.0577	0.2074
TPOX	0.0001	0.0272
D18S51	0.0082	0.0422
D5S818	0.0314	0.1497
FGA	0.0021	0.0410
	1.55E-30	2.65E-18

あり、1.55 E-30 である。 $A\alpha$ 確率は $\alpha\alpha$ 確率の約1.7兆倍（尤度比）であるが、これは、同図の現場資料の DNA 型は、(網掛けの DNA 型をもつ)「甲」と任意の1人の人で形成される確率は任意の2人の人で形成される確率に比して（非常に）大きいということであり、「甲」（被疑者）の形成関与の尤もらしさを意味している。

なお、15 のローカスにおける $A\alpha$ 確率と $\alpha\alpha$ 確率の比（尤度比）は、ローカスごとの尤度比の総乗で求めることもできるが、ローカスごとの

尤度比は、次の式¹¹で得られる。

現場資料	被疑者	尤度比	計算式
a_1, a_2, a_3, a_4	a_1, a_2	$\frac{1}{12p_1p_2}$	① / ④
a_1, a_2, a_3	a_1, a_2	$\frac{2p_1 + 2p_2 + p_3}{12p_1p_2(p_1 + p_2 + p_3)}$	② / ⑤
a_1, a_2, a_3	a_1	$\frac{1}{6p_1(p_1 + p_2 + p_3)}$	① / ⑤
a_1, a_2	a_1, a_2	$\frac{(p_1 + p_2)^2}{2p_1p_2(2p_1^2 + 3p_1p_2 + 2p_2^2)}$	③ / ⑥
a_1, a_2	a_1	$\frac{2p_1 + p_2}{2p_1(2p_1^2 + 3p_1p_2 + 2p_2^2)}$	② / ⑥
a_1	a_1	$\frac{1}{2p_1^2}$	③ / ⑦

(※) 例えば、2行目でいうと、現場資料の DNA 型が3種類のアレル a_1, a_2, a_3 から成るものであるとき、そのうちの2種類のアレル a_1, a_2 から成る DNA 型をもつ者の尤度比は、②式の⑤式に対する比として、添え字を適当に読み替えて

$$\frac{2p_3(p_1 + p_2 + p_3) - p_3^2}{12(p_1p_2p_3)(p_1 + p_2 + p_3)} = \frac{2p_1 + 2p_2 + p_3}{12p_1p_2(p_1 + p_2 + p_3)}$$

で求められるということである。

(2) 被害者と犯人との2人に由来すると認められるとき

– “DNA.xlsm”のシート“混合DNA (p2)” –

現場資料がある者「乙」(被害者など)と犯人との2人で形成されたと、DNA 型以外の情報(証拠)によって、認定できるケースがある。例えば、現場資料が性犯罪の単独犯の犯人が触れた被害者の衣服で覆われている部位から採取されたものであることなどから、被害者と犯人との2人だ

けで形成されたと認められる ((1)の確率計算で被害者の形成関与の事実の認定を支えることもできると考えられる。) ようなケースである。

ここでは, [第5図] のように, 現場資料の DNA 型が, 丸囲みの DNA 型をもつ「乙」(被害者など) と網掛けを付した DNA 型をもつ「甲」(被疑者) との2人で形成されたとして矛盾のないものであるケースで, 「乙」と犯人との2人で形成されたことを前提とすることができるときに, これを前提として, 「甲」が形成者であること(犯人性)の確からしさを考える。

第5図

ローカス	DNA型			
D8S1179	11	(12)	(13)	14
D21S11	29	(30)		
D7S820	10	(12)	(13)	
CSF1PO	9	(11)	(12)	
D3S1358	(14)	15	18	
TH01	(6)	(7)	8	
D13S317	(11)	(13)		
D16S539	(9)	11	(12)	
D2S1338	17	(18)	20	(23)
D19S433	13	(14)	(16)	
vWA	(17)	18		
TPOX	11	(12)	(13)	
D18S51	13	(15)	(16)	17
D5S818	10	(12)		
FGA	18	(21)	22	(24)

a 出現頻度

もう一人の形成者(「乙」ではない形成者)である資格(犯人資格)は, 「乙」との混合で現場資料と一致する DNA 型を形成できる DNA 型をもつことである。ローカスごとに, その資格をもつ人の出現頻度は, 形成関与者のもつアレルは現場資料の DNA 鑑定ですべて検出されているとすると, 現場資料の DNA 型に含まれるアレルから「乙」のもつアレルを除いた「非重複アレル」の種類の数に応じ, 次のとおりである。なお, 現場資料の DNA 型に含まれるアレルの出現頻度の総和を $\sum p_i$ で表す。

(非重複アレルの種類数が「2」であるとき)

例えば, [第5図] の D8S1179 や D3S1358 のローカスでは, 非重複アレルが2つあるが, この場合, もう一人の形成者(犯人)は, その2つアレルをもつことが必要である。D8S1179 のローカスでいうと, “11,14” の DNA 型をもつことである。したがって, もう一人の形成者の資格(犯人資格)をもつ人の出現頻度は, $2 \times p(\{11\}) \times$

$p(\{14\})=0.0446$ (22.4人に1人の割合) である。非重複のアレルを a_1, a_2 とすると,

$$2 \times \prod_{i=1}^2 p_i = 2p_1p_2 \quad (\text{①式})$$

である。

(非重複アレルの種類数が「1」であるとき)

例えば、同図の D21S11 や D7S820 のローカスでは、非重複アレルが1つあるが、この場合、もう一人の形成者(犯人)は、現場資料の DNA 型に含まれるアレルのみをもち、かつ、非重複アレルを(少なくとも1つ)もつことが必要である。D7S820 のローカスでいうと、“10”、“12”、“13”のアレルのみをもち、かつ、“10”をもつことであり、すなわち、“10,10”、“10,12”、“10,13”のいずれかであることである。したがって、もう一人の形成者の資格(犯人資格)をもつ人の出現頻度は、それらの DNA 型をもつ人の出現頻度の総和で求められ、 $p(\{10\})^2+2 \times p(\{10\}) \times p(\{12\})+2 \times p(\{10\}) \times p(\{13\})=0.0479+0.1028+0.0152=0.1660$ (6.03人に1人の割合) である(シート“混合DNA (p2) 列挙”)。このように、もう一人の形成者の資格の DNA 型を列挙した上で、その資格をもつ人の出現頻度を求めることもできるが、“10”、“12”、“13”のアレルのみをもち、かつ、“10”のアレルをもつ人の集合は、“10”、“12”、“13”のアレルのみをもつ人(2つのアレルがともに“10”、“12”、“13”のいずれかである人)の集合から“10”のアレルをもたない人(2つのアレルがともに“12”、“13”のいずれかである人)の集合を引いた差集合であるので、もう一人の形成者の資格をもつ人の出現頻度は、2つとも“10”、“12”、“13”のいずれかのアレルをもつ人の出現頻度から2つとも“12”、“13”のいずれかのアレルをもつ人の出現頻度を減ずることで求められ、 $p(\{10,12,13\})^2-p(\{12,13\})^2$ である。一般化すると、現場資料の DNA 型に含まれるアレルの出現頻度の和の2乗から、重複するアレルの出現頻度の和(現場資料の

DNA 型に含まれるアレルの出現頻度の和から非重複アレルの出現頻度を減じて得られる。)の2乗を減ずることで、求められる。また、「非重複アレル」と「現場資料のDNA型に含まれるアレルのいずれか」をもつ人の出現頻度から(ダブルカウントされる)「非重複アレルのホモ接合」である人の出現頻度を減ずることで求められる。すなわち、非重複アレルを a_1 とすると

$$\left(\sum p_i\right)^2 - \left(\sum p_i - p_1\right)^2 = 2 \times p_1 \times \sum p_i - p_1^2 \quad (2 \text{式})$$

である。

(非重複アレルがないとき)

例えば、同図の D13S317 のローカスでは、現場資料と「乙」のDNA型は一致しているが、この場合、もう一人の形成者(犯人)は、現場資料のDNA型に含まれるアレルのみをもつことが必要である。このローカスでいうと、「11」、「13」のアレルのみをもつことであり、列挙すれば、「11,11」、「11,13」、「13,13」のいずれかであることである(シート「混合DNA (p2) 列挙」)。したがって、もう一人の形成者の資格をもつ人の出現頻度は、 $p(\{11,13\})^2$ である。一般化すると、混合資料のDNA型に含まれるアレルの出現頻度の和の2乗、すなわち

$$\left(\sum p_i\right)^2 \quad (3 \text{式})$$

である。

こうして、15のローカスごとに、非重複アレルの種類数に応じて、もう一人の形成者の資格をもつ人の出現頻度を求めて、それらの総乗を求めると、[第5図]のケースでは、「乙」(被害者など)との混合で現場資料と

ローカス	出現頻度
D8S1179	0.0446
D21S11	0.2271
D7S820	0.1660
CSF1PO	0.0606
D3S1358	0.0502
TH01	0.0694
D13S317	0.0741
D16S539	0.2341
D2S1338	0.0206
D19S433	0.2866
vWA	0.1784
TPOX	0.1589
D18S51	0.0325
D5S818	0.1347
FGA	0.0087
	3.89E-17

2.57E+16

一致する DNA 型を形成できるというもう一人の形成者の資格 (犯人資格) をもつ人は, 出現頻度は 3.89×10^{-17} であり, 約 2.6 京人に 1 人の割合でしか存在しないということになる。

b 尤度比

ある者「乙」(被害者など)との混合で現場資料と一致する DNA 型を形成できる DNA 型をもつ者「甲」(被疑者)がもう一人の形成者であること(犯人性)の確からしさは, ①「甲」と「乙」との混合で形成されたという仮説と, ②任意の 1 人の人と「乙」との混合で形成されたという仮説の下で, それぞれ, 現場資料と一致する DNA 型が形成される確率の比(尤度比)で考えることもできる。

形成関与者のもつアレルは現場資料の DNA 鑑定ですべて検出されているとすると, ①の仮説の下での確率は, 1 (=100%) であり, ②の仮説の下での確率は, 上記 a で求めた確率(出現頻度)と同じである。したがって, [第 5 図] のケースでは, ①の確率と②の確率の比(尤度比)は, 約 2.6 京(上記 a で求めた確率の逆数)となる。

これは, 現場資料が「甲」(被疑者)と「乙」(被害者など)との混合で形成されたという仮説は(任意の 1 人の人と「乙」との混合で形成されたという仮説より)はるかに尤もらしいということであるが, この比は, ①の仮説の下での確率が 1 である(「甲」と「乙」との混合で必ず現場資料と一致する DNA 型が形成されると考える)とき, 必ず, 上記 a で求めた確率の逆数であり, もう一人の形成者の資格(犯人資格)をもつ人はその逆数の人数(約 2.6 京人)に 1 人の割合で存在するにすぎないと説明する方が証拠価値を理解しやすいように思われる。

(3) 2 人の犯人に由来すると認められるとき

—“DNA.xlsm”のシート“混合 DNA (2 人)”の $\alpha \alpha$ 確率—

ここでは, 2 人の犯人だけで形成されたと認められる現場資料(犯人 2 人の精液が混合したものなど)の DNA 型が, [第 5 図] のように, X (同図の丸囲みの DNA 型をもつとする。)と Y (同図の網掛けの DNA 型

をもつとする。)の2人で形成されたとして矛盾のないものであるケースで、XとYの2人ペアが形成者であること(2人ペアの犯人性)の確からしさを考える。

現場資料と一致するDNA型を2人で形成できるという資格(犯人資格)をもつ2人ペアの出現頻度は、(極めて多数の2人ペアから抽出された)任意の1組の2人ペアで現場資料と一致するDNA型が形成される確率であり、上記(1)bの α 確率である。[第5図]のケースでは、その出現頻度は、 1.55×10^{-30} であり、約65種(1種は1京の(1万)³倍)組に1組の割合で存在するにすぎない。この出現頻度の逆数は、 N 人の集団のうちの2人から成るペアの総数(組合せの数 ${}^N C_2$ であり、 $N=70$ 億で約2,450京、 $N=1$ 億で約5,000兆である。)に比して、はるかに大きい。

なお、ここで想定しているケースで、Yが形成者の一人であること(Yの犯人性)がDNA型以外の証拠(Yの自白や防犯カメラの映像など)によって認められるとき、Xがもう一人の形成者であること(Xの犯人性)の確からしさは、上記(2)の確率計算で考えることができる。

ローカス	α 確率
D8S1179	0.0148
D21S11	0.0995
D7S820	0.0105
CSF1PO	0.0328
D3S1358	0.0042
TH01	0.0264
D13S317	0.0031
D16S539	0.1019
D2S1338	0.0058
D19S433	0.0040
vWA	0.0577
TPOX	0.0001
D18S51	0.0082
D5S818	0.0314
FGA	0.0021

1.55×10^{-30}
6.46E+29

7 混合資料の形成関与者の数

すべてのローカスでアレルの種類数が4以下であるとき、その現場資料は、2人の人で形成可能であるが、3人以上の人で形成された可能性もある。ここでは、すべてのローカスでアレルの種類数が4以下である現場資料の形成関与者の数の推定(3人以上の人で形成される確率)を、形成関与者のもつアレルは現場資料のDNA型鑑定ですべて検出されているという前提で、考える。

(1) アレルの種類数(4以下であること)による推定

a シミュレーション的方法

– “混合実験.xlsx”, “混合実験 2p 10万回.xlsx”, “混合実験 3p 10万回.xlsx” –

任意の 3 人の人で形成される混合資料がすべてのローカス (15 のローカス) でアレルの種類数が 4 以下となる確率は, MS Excel を用いてシミュレーション的な方法で求めることもできる。具体的には, ①ローカスごとに任意の 3 人の人の 6 つのアレルを父親由来・母親由来の順に並べる数列 (例えば, 3 人の DNA 型が父親由来と母親由来のアレルを区別して “10,11”, “12,12”, “11,13” であれば, “10,11,12,12,11,13” の数列) を想定し, ② Excel のワークシートの一行で, ① 15 のローカス

3 人分 : 出現頻度に応じてランダム発生
15 loci × 3 persons × 2

D9S1170											D21S11					D7S820
16	13	15	10	11	13	5	31	30	32	30	30	30	3	11	11	
13	14	14	14	15	14	3	28	32	30	30	30	32	3	13	14	
11	15	12	11	13	13	2	28	30	31	31	30	28	3	10	12	
14	13	12	13	14	11	4	32	29	32	32	33	29	3	9	10	
15	12	15	11	13	14	5	1	30	29	29	30	32	4	11	11	
12	14	16	15	13	11	6	30	29	29	31	29	33	4	10	11	

種類の数

N 回

ごとに, その数列の要素をアレルの出現頻度に従いランダム発生させて, その要素の種類の数 (“16,13,15,10,11,13” の数列であれば, “16”, “13”, “15”, “10”, “11” の 「5」種類) を求め, ②アレルの種類数が 4 以下となるローカスの数をカウントする, ③ワークシートの一行で行う②の計算を N 行 (回) 繰り返し, アレルの種類数が 4 以下となるローカスの数が 15 である行 (15 のすべてのローカスでアレルの種類数が 4 以下となる行) の数をカウントして, その N 行中の割合を求めるという方法であり, N を大きくすることで, 3 人の人で形成される混合資料がすべてのローカスでアレルの種類数が 4 以下となる確率のおおよその値が得られる。

b 数値計算

上記 a の方法で求める値は実験値のようなものである（計算するごとに値は変動する）ので、数値計算で、 n 人の人で形成される混合資料のローカスごとのアレルの種類数が m ($m=1, 2, \dots, 2n$) となる確率を求めて、任意の 3 人の人で形成される混合資料がすべてのローカス（15 のローカス）でアレルの種類数が 4 以下となる確率を求める。

その計算方法の考え方は、次のとおりである。3 人混合で説明する。a に記載したように、3 人の 6 つのアレルを並べる順列を想定する。あるローカスのアレルの種類数が m ($m=1, 2, \dots, 6$) となる確率は、その要素（並べられたアレル）の種類数が m となる順列の出現頻度の総和であるが、これは、次の方法で求めることができる。例として種類数が「3」となる確率を求めるとする。あるローカスで、アレルは、説明のために簡略化して、 a_1, a_2, a_3, a_4 の 4 種類であるとする、種類数が 3 になる順列の要素（アレル）の組合せは $(a_1, a_2, a_3), (a_1, a_2, a_4), (a_1, a_3, a_4), (a_2, a_3, a_4)$ のいずれかであり、また、3 種類の各要素が順列に現れる数の組合せは $(1,1,4), (1,2,3), (1,3,2), (1,4,1), (2,1,3), (2,2,2), (2,3,1), (3,1,2), (3,2,1), (4,1,1)$ のいずれかである。例えば、 $(a_1, a_2, a_3) \& (1,3,2)$ は、1 つの a_1 と 3 つの a_2 と 2 つの a_3 から成る順列ということであるが、それらの順列の数は $60 (= {}_6C_1 \times {}_5C_3)$ であり、ひとつひとつの順列の出現頻度はいずれも $p_1 \times p_2^3 \times p_3^2$ であるから、それらの順列の出現頻度の合計は、 $60 \times p_1 \times p_2^3 \times p_3^2$ で求められる。これにより、順列の要素の組合せと各要素が順列に現れる数の組合せのセットごとに、順列の出現頻度の合計を求め、すべてのセットのそれらの合計の総和で、種類数が「3」となる確率を求める。こうして、ローカスごとに、アレルの種類数が m ($m=1, 2, \dots, 6$) となる各確率を求めて、4 以下になる確率（1～4 になる確率の合計）を求め、ローカスごとのその確率（4 以下になる確率）の総乗で、すべてのローカスでアレルの種類数が 4 以下となる確率を得ることができる。PrAlleleNumEachLocus は、この方法で、任意の n ($n=2, 3, \dots$) 人の人の混合のときに、ローカスごとに、アレルの種類

数が $1 \sim 2n$ 個になる各確率を計算するために作成した Java のプログラムである（“混合人数 (VB) .xlsx”のシート“種類数”でも、2人混合と3人混合の場合に限って VBA でこの数値計算を行っているが、その計算にはかなりの時間を要する。）。これによる計算結果から、すべてのローカス（15のローカス）でアレルの種類数が4以下になる確率を求めると、3人混合の場合、4.18...%（4～6人混合の場合、それぞれ、0.0217...%, 0.0000518...%, 0.0000000945...%）である（“混合人数.xlsx”のシート“種類数”）。

なお、任意の n ($n=2, 3, \dots$) 人の人で形成される混合資料がすべてのローカスでアレルの種類数が2以下になる確率は、上記の計算結果から求めると、2人混合の場合、0.000000152% (1.52 E-09), 3人混合の場合、6.14 E-19 である（“混合人数.xlsx”のシート“種類数”）。

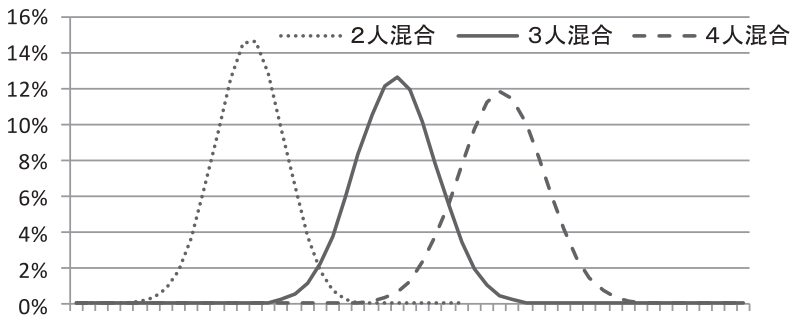
(2) アレルの種類数の合計数による推定

–“混合人数.xlsx”のシート“合計数”–

任意の3人の人で形成される混合資料は、上記(1)のとおり、4パーセント程度の確率で、すべてのローカスでアレルの種類数が4以下になるにすぎず、その意味では、すべてのローカスでアレルの種類数が4以下であるとき、その現場資料は、3人以上の人で形成された可能性は低い。しかしながら、(1)で求めたのは、混合資料の具体的な DNA 型（“11,12”や“13,14,15”や“16,17,18”などであること）や種類数（4,3,2,1のいずれであるか）を問わず、単に文字どおり（すべてのローカスで）4以下となる確率である。すべてのローカスでアレルの種類数が4以下であっても、極端な例として、すべてのローカスでアレルの種類数が4であるときは、直感的に、3人の人で形成された可能性がより高いと考えられる。

そこで、次に、任意の n 人の人で形成される混合資料のローカスごとのアレルの種類数の合計数が m となる確率（確率分布）を考える。例えば、3人の人で形成される混合資料の15のローカスごとのアレルの種類数の合計数は、 $15(=1 \times 15) \sim 90(=3 \times 2 \times 15)$ であるが、その合計数が15、

16, 17, ……89, 90 となる各確率を求めるのである。計算方法の考え方は、次のとおりである。混合資料のローカスごとのアレルの種類数を並べる順列（例えば、D8S1179 が2種類、D21S11 が3種類、……FGA が5種類であるときは、“2, 3, ……5”）を想定すると、その総数は $(2n)^{15}$ であるが、これらの個数 $(2n)^{15}$ の順列をアレルの種類数の合計数（15～15×2n）で合計数15の集合、合計数16の集合、……合計数15×2nの集合に分類し、それぞれの集合ごとに、混合資料のローカスごとのアレルの種類数（集合の要素である）順列の要素となる確率（例えば、“2, 3, ……”の順列であれば、アレルの種類数が、D8S1179 で2種類、D21S11 で3種類、……となる確率）－これは、上記(1)で求めた、 n 人混合でローカスごとのアレルの種類数が1～2nになる確率の総乗で求められる－の総和を求めると、種類数の合計数が15～15×2nとなる各確率を得ることができる。ただし、この方法では計算量が膨大となるので、15のローカスを5つずつのグループ（D8S1179～D3S1358, TH01～D19S433, vWA～FGA）に分けて、5つのローカスごとの種類数の合計数が5～5×2nとなる各確率を求めた上、これらの3グループを結合する形で、同様の方法で、種類数の合計数が15～15×2nとなる各確率を求める。PrAlleleTotalNum は、この方法で、任意の n ($n=2, 3, 4, …$) 人の人の混合のときのアレルの種類数の合計数の確率分布を求めるために作成した Java のプログラムである（“混合人数 (VB) .xslm”のシート“合計数”



では、2人混合と3人混合の場合に限って VBA でこの計算を行っている。)。これにより、2人混合、3人混合、4人混合のときのローカスごとのアレルの種類数の合計数の確率分布を求めると、左図のとおりであり、 m が49以上であるとき、3人混合で合計数が m になる確率は、2人混合で合計数が m になる確率より大きい。

(3) 具体的な DNA 型を前提とした推定

現場資料の形成関与者の数を、すべてのローカスでアレルの種類数が4以下であるという属性やローカスごとのアレルの種類数の合計数という属性（をもつ DNA 型が形成される確率）で推定するとすると、上記のとおりであるが、“11, 12, 13, 14” などという具体的な DNA 型が与えられているときに、“11, 12, 13, 14” であることを4種類以下であるなどと）抽象化した属性で形成関与者の数を推定するのは、あまり意味がないと思われる。

a そこで、現場資料の具体的な DNA 型を前提として、任意の n ($n = 2, 3, \dots$) 人の人の混合でその DNA 型が形成される確率を求め、これで、形成関与者の数を推定することを考える。ローカスごとに、 n 人の人の混合で、ある具体的な DNA 型が形成される確率は、 n 人の人の個数 $2n$ のアレルを並べる順列を想定すると、現場資料の DNA 型に対応する順列（現場資料に含まれるアレルの順列）の出現頻度の総和であるが、これは、上記(1)bに記載したとおり、順列に現れるアレルの種類ごとの数の組合せごとに、順列の数とひとつひとつの順列の出現頻度の積を求め、それらの積の総和で求められる。PrMatchEachLocus は、この計算方法で、任意の n ($n=2, 3, \dots$) 人の人の混合で、ある具体的な DNA 型が形成される確率を計算するために作成した Java のプログラムである。これを、現場資料の DNA 型を記録した MixAlleleList.csv や形成関与者の数（混合人数）を引数として実行すると、（ローカスごとに）現場資料と一致する DNA 型が形成される確率が求められる。現場資料の DNA 型が [第6図] のとおりであるとき、(15 のローカスで)

その DNA 型が形成される確率は、2人混合では 8.95×10^{-31} 、3人混合では 1.31×10^{-34} （4～6人混合では、それぞれ、 4.61×10^{-41} 、 2.23×10^{-48} 、 4.27×10^{-56} ）であり、2人混合のときの確率は3人混合のときの確率の約7,000倍となる。

なお、同様に、すべてのローカスでアレルの種類数が2以下である現場資料の形成関与者の数の推定（2人以上の人で形成される確率）を考えることもできる。現場資料の DNA 型が〔第1図〕（5

第6図

ローカス	DNA型			
D8S1179	11	12	13	14
D21S11	29	30		
D7S820	10	12	13	
CSF1PO	9	11	12	
D3S1358	14	15	18	
TH01	6	7	8	
D13S317	11	13		
D16S539	9	11	12	
D2S1338	17	18	20	23
D19S433	13	14	16	
vWA	17	18		
TPOX	11	12	13	
D18S51	13	15	16	17
D5S818	10	12		
FGA	18	21	22	24

(1)に掲げたもの)のとおりであるとき、任意の n ($n = 2, 3, \dots$) 人の人の混合でその DNA 型が形成される確率は、上記の PrMatchEachLocus で求めると、2人混合では 4.13×10^{-35} （3, 4人混合では、それぞれ、 6.86×10^{-50} 、 3.52×10^{-65} ）である。他方、任意の1人の人でこの DNA 型が形成される確率（この DNA 型をもつ人の出現頻度）は 2.07×10^{-22} （上記5(1)）であるから、その確率は、任意の2人の人の混合で形成される確率の約5兆倍となる。

b - “DNA.xlsm”のシート“混合 DNA（人数）”-

上記 a では、任意の n ($n=2, 3, \dots$) 人の人の混合で現場資料の具体的な DNA 型が形成される確率を求める方法の考え方を説明したが、2人混合のとき、その確率は、6(1)bの $\alpha \alpha$ 確率であり、また、3人混合のとき、そのローカスごとの確率は、3人の人の6つのアレルを並べる順列を想定し、現場資料の DNA 型に含まれるアレルの種類数に応じて、次の計算式で求められる。なお、現場資料の DNA 型に含まれるアレルの出現頻度の総和を $\sum p_i$ 、総乗を $\prod p_i$ などと表す。

(現場資料のアレルの種類数が「1」であるとき)

6つのアレルはすべて a_1 であるから, p_1^6 である。

(現場資料のアレルの種類数が「2」であるとき)

2種類のアレル a_1, a_2 から成る順列は, a_1 又は a_2 から成る順列のうち, a_1 のみから成る順列と a_2 のみから成る順列を除いたものであるから,

$$\left(\sum p_i\right)^6 - \sum p_i^6 = (p_1 + p_2)^6 - p_1^6 - p_2^6$$

である。

(現場資料のアレルの種類数が「3」であるとき)

3種類のアレル a_1, a_2, a_3 から成る順列は, ① a_1, a_2 又は a_3 から成る順列のうち, ②① a_1 又は a_2 から成る順列, ③① a_1 又は a_3 から成る順列, ④① a_2 又は a_3 から成る順列 (それぞれ, a_3, a_2, a_1 を含まない順列である。)を除いたものであるが, 1種類のアレルのみから成る順列は, ②の①③④に重複して含まれるので, 3種類のアレル a_1, a_2, a_3 から成る順列の出現頻度は, 「①の順列の出現頻度」から「②③④の順列の出現頻度の和から (ダブルカウントされる) 1種類のアレルのみから成る順列の出現頻度を減じたもの」を減ずることで求められる。すなわち

$$\left(\sum p_i\right)^6 - \left\{\left(\sum p_i - p_1\right)^6 + \left(\sum p_i - p_2\right)^6 + \left(\sum p_i - p_3\right)^6\right\} + \sum p_i^6$$

である。

(現場資料のアレルの種類数が「4」であるとき)

4種類のアレル a_1, a_2, a_3, a_4 のうち, 2つ以上含まれるアレルに着目すると, ①2種類のアレルが2つずつ含まれるか, ②1種類のアレルが3つ含まれるかのいずれかである。①の2つ含まれるアレルを a_i, a_j とすると, i, j の組合せは (1,2), (1,3), (1,4), (2,3), (2,4), (3,4) のいずれかであるが, いずれであるときも, それぞれ, 順列の数は $6!/(2! \times$

2)=180 であり、また、ひとつひとつの順列の出現頻度は $p_1 p_2 p_3 p_4 \times p_i p_j$ であるから、2種類のアレルが2つずつ含まれる順列の出現頻度は $180 \times p_1 p_2 p_3 p_4 \times (p_1 p_2 + p_1 p_3 + p_1 p_4 + p_2 p_3 + p_2 p_4 + p_3 p_4)$ である。②の順列の出現頻度は、3つ含まれるアレルが a_1, a_2, a_3, a_4 のいずれであるときも、それぞれ、順列の数は $6!/3! = 120$ であり、また、ひとつひとつの順列の出現頻度は、3つ含まれるアレルを a_k とすると、 $p_1 p_2 p_3 p_4 \times p_k^2$ であるから、 $120 \times p_1 p_2 p_3 p_4 \times (p_1^2 + p_2^2 + p_3^2 + p_4^2)$ である。したがって、4種類のアレル a_1, a_2, a_3, a_4 から成る順列の出現頻度は

$$30 \times \prod p_i \times \left\{ 6 \times \left(\left(\sum p_i \right)^2 - \sum p_i^2 \right) \times 1/2 + 4 \times \sum p_i^2 \right\}$$

$$= 30 \times \prod p_i \times \left\{ 3 \times \left(\sum p_i \right)^2 + \sum p_i^2 \right\}$$

である。

これらの計算式により、任意の3人の人で現場資料の具体的なDNA型が形成される確率を求めると、そのDNA型が〔第6図〕のとおりであるとき、(15のローカスでの)その確率は、一当然のことながら、上記aのPrMatchEachLocusによる計算結果と同じであるが、 1.31×10^{-34} であり、2人混合のときの確率(6(1)bの $\alpha \alpha$ 確率)の約7,000倍となる。

ローカス	2人混合	3人混合
D8S1179	0.0148	0.0264
D21S11	0.0995	0.0379
D7S820	0.0105	0.0045
CSF1PO	0.0328	0.0276
D3S1358	0.0042	0.0021
TH01	0.0264	0.0143
D13S317	0.0031	0.0003
D16S539	0.1019	0.0903
D2S1338	0.0058	0.0061
D19S433	0.0040	0.0031
vWA	0.0577	0.0166
TPOX	0.0001	0.0000
D18S51	0.0082	0.0111
D5S818	0.0314	0.0066
FGA	0.0021	0.0023
	8.95E-31	1.31E-34

(4) 3人混合であるとしたときの尤度比

—“DNA.xlsm”のシート“混合DNA (p3)”—

ここでは、現場資料のDNA型がある者「甲」(被疑者)とある者「乙」(被害者など)との2人で形成されたとして矛盾のないものであるケース

で、現場資料が3人の人で形成されたとして、「乙」の形成関与を前提に、「甲」が形成関与者であること（犯人性）の確からしさを考える。

「甲」の犯人性の確からしさは、6(1)aに記載したとおり、現場資料のDNA型に含まれるアレルのみをもつという形成関与者の資格をもつ人の割合（出現頻度）で考えるほか、「甲」と「乙」と任意の1人の人との混合で形成されたという仮説の下で、現場資料と一致するDNA型が形成される確率（「 $AV\alpha$ 確率」ということとする。）と、「乙」と任意の2人の人との混合で形成されたという仮説の下で、現場資料と一致するDNA型が形成される確率（「 $V\alpha_1\alpha_2$ 確率」ということとする。）の比（尤度比）で考えることができる。

ローカスごとに、 $AV\alpha$ 確率と $V\alpha_1\alpha_2$ 確率の計算方法は、形成関与者のもつアレルは現場資料のDNA鑑定ですべて検出されているとすると、次のとおりである。なお、現場資料のDNA型に含まれるアレルの出現頻度の総和を $\sum p_i$ で表す。

- i $AV\alpha$ 確率は、「甲」と「乙」の2人との混合で現場資料と一致するDNA型を形成できる人「 α 」の出現頻度と同じであるが、「甲」と「乙」とで既に現場資料のDNA型の形成に足りないものはない（現場資料のDNA型に「甲」も「乙」ももたないアレルは含まれていない。）ので、 α の資格は、現場資料のDNA型に含まれるアレルのみをもつことであり、そのローカスごとの確率（出現頻度）は、 $(\sum p_i)^2$ である。
- ii $V\alpha_1\alpha_2$ 確率は、「乙」との混合で現場資料と一致するDNA型を形成できる2人ペアの出現頻度と同じであるが、そのローカスごとの確率（出現頻度）は、現場資料のDNA型に含まれるアレルから「乙」のもつアレルを除いた「非重複アレル」の種類の数に応じ、次のとおりである。

（非重複アレルがないとき）

2人ペアの資格は、現場資料のDNA型に含まれるアレルのみをもつことであるので、 $(\sum p_i)^4$ である。

(非重複アレルの種類数が「1」であるとき)

2人ペアの資格は、現場資料のDNA型に含まれるアレルのみをもち、かつ、2人のいずれか(又は双方)が非重複アレル a_1 をもつことであるので、 $(\sum p_i)^4 - (\sum p_i - p_1)^4$ である。

(非重複アレルの種類数が「2」であるとき)

2人ペアの資格は、現場資料のDNA型に含まれるアレルのみをもち、かつ、2人で、2つの非重複アレル a_1, a_2 をもつことである。この資格をもつ2人ペアの集合は、現場資料のDNA型に含まれるアレルのみをもつ2人ペアの集合から2つの非重複アレルの1つ又は2つをもたない2人ペアの集合を引いた差集合であり、2つの非重複アレルの1つ又は2つをもたない2人ペアの集合は、アレル a_1 をもたない2人ペアの集合とアレル a_2 をもたない2人ペアの集合の和集合である。したがって、この資格をもつ2人ペアの出現頻度は、

$$\left(\sum p_i \right)^4 - \left\{ \left(\sum p_i - p_1 \right)^4 + \left(\sum p_i - p_2 \right)^4 - \left(\sum p_i - p_1 - p_2 \right)^4 \right\}$$

である。

こうして、形成関与者のアレルはすべて検出されているという前提の下であるが、「甲」(被疑者)と「乙」と任意の1人との3人で形成されたという仮説と、「乙」と任意の2人の人との3人で形成されたという仮説の下で、現場資料と一致するDNA型が形成される各確率(AV α 確率とV $\alpha_1\alpha_2$ 確率)を求めると、[第5図]のケース(「甲」と「乙」はそれぞれ網掛けと丸囲みのDNA型をも

ローカス	AV α 確率	V $\alpha_1\alpha_2$ 確率
D8S1179	0.4371	0.0693
D21S11	0.3407	0.1032
D7S820	0.2386	0.0517
CSF1PO	0.4529	0.0513
D3S1358	0.2382	0.0241
TH01	0.3095	0.0381
D13S317	0.0741	0.0055
D16S539	0.5172	0.1873
D2S1338	0.2587	0.0207
D19S433	0.4124	0.1542
vWA	0.2583	0.0603
TPOX	0.1604	0.0257
D18S51	0.3301	0.0375
D5S818	0.1900	0.0330
FGA	0.2608	0.0085
	3.29E-09	9.28E-22

つとする。)では, 15 のローカスで, $AV\alpha$ 確率と $V\alpha_1\alpha_2$ 確率は, それぞれ, 3.29×10^{-9} , 9.28×10^{-22} である。 $AV\alpha$ 確率は, $V\alpha_1\alpha_2$ 確率の約3.5兆倍(尤度比)であり, 「甲」(被疑者)が(「乙」とともに)現場資料の形成に関与したことの尤もらしさを意味している。

8 おわりに

例えば

ローカス D8S1179 で, 現場資料の DNA 型は“11,12,13”で, 被害者は“11,12”であるから, 犯人の DNA 型は“11,13”, “12,13”, “13,13”のいずれかである。“11”の出現頻度は 0.1089 であり, “13”の出現頻度は 0.2248 であるから, “11,13”の出現頻度は $2 \times 0.1089 \times 0.2248 = 0.0490$ である。“12”と“13”の出現頻度は 0.1226 と 0.2248 であるから, “12,13”の出現頻度は 0.0551 であり, …… “13,13”は……0.0505 である。したがって, 犯人として矛盾しない DNA 型をもつ者の出現頻度は, これらを合計して, 0.1546 である。

というのは, 理解も容易であるが, 15 のローカスでこうした計算を繰り返すのは手間がかかる。本稿は, こうした確率(出現頻度)の値を一般化して求めるための数式 $(\sum p_i)^2 - (\sum p_i - p_1)^2$ の導出方法を説明したものである。それは, アレルの有無だけに基づく確率計算であり, あらゆるケースで万能的に利用できるものではないが, それらの計算式によって作成した Excel のファイルは, 手間のかかる計算作業を省くのに役立つはずである。

注

- 1 科学警察研究所のサンプル調査による出現頻度は, 小数点以下4桁の数値の形で公表されているが, 調査対象者の数(1,350人)からこの調査で計数されたデータを推定することができる(例えば, D8S1179 のアレル“10”の出現頻度は 0.1326 であるが, この調査で計数された“10”の数は $358 (\approx 0.1326 \times 1350 \times 2)$ であるはずである。)ので, 出現頻度として, (0.1326 ではなく) $358/2700$ の値(“頻度表.xlsx”)を用いることも考えられる。これによる出現頻度(表)を取って用いる意味はあまりないが, Excel のファイルでは, この出現頻度を用いた確率計算を行うこともできる。また, 本稿7項の(1)b

と(2)の計算結果は、この出現頻度で求めた値である。

2 Butler 410~411頁参照

3 アレルの有無だけにに基づく確率計算は、本文の設例のケースでいうと、犯人の DNA 型は“12,13”の可能性が高いことを考慮しないので、その DNA 型をもつ被疑者の犯人性を控え目に(被疑者有利に)見積もることになる。ただし、被疑者の DNA 型が“11,13”又は“13,13”であるときは、被疑者が犯人であるのに“12”のピークが相対的に高い理由を検討しなければならない。

4 Butler 442~444, 453頁参照

5 Butler 444~445頁参照

6 事象 Y_1, Y_2, \dots, Y_n は、互いに排反であり、これらで全事象を尽くしているとする。このとき

$$P(Y_i/X) = \frac{P(X \cap Y_i)}{P(X)} = \frac{P(Y_i)P(X/Y_i)}{\sum_{k=1}^n (P(Y_k)P(X/Y_k))}$$

が成立する。

7 この式は、展開・整理すると

$$\begin{aligned} & \frac{(N-1)}{2!}p^1 - \frac{(N-1)(N-2)}{3!}p^2 + \frac{(N-1)(N-2)(N-3)}{4!}p^3 - \dots \\ & \dots - \frac{(N-1)!}{N!}(-p)^{N-1} \end{aligned}$$

であるが、その値は、 p が N の逆数に比して十分に小さいとき、この式の最初からの数項で近似計算できる。

8 Butler 430~431, 433~434頁参照

9 Butler 451~453頁には、排除率 (PE : probability of exclusion) について、「ローカス i での排除率は、現場資料の DNA 型に含まれるアレルの出現頻度の和を p とし、 $q=1-p$ とすると、 $PE_i=2pq+q^2$ であり、複数のローカスからの総合排除率は、

$$1 - \prod (1 - PE_i) \quad (\text{※式})$$

で求められる。」旨説明されている。これは、「任意に抽出された人は、現場資料の DNA 型に含まれないアレルをもつとき、形成者ではないとして排除されるが、ローカスごとの排除率(確率)は、 $PE=2pq+q^2$ (現場資料の DNA 型に含まれないアレルを1つだけもつ人と2つもつ人の出現頻度の和)であり、排除されない確率は、 $1-PE_i$ である。15のローカスを考えたときの総合排除率は、※式で求められる。」という意味であろう。

総合排除率は、任意に抽出された人が1つ又は複数のローカスで現場資料の DNA 型に含まれないアレルをもつことで形成者から排除される確率であると考えられるが、 $1-PE_i=1-\{2p_i(1-p_i)+(1-p_i)^2\}=p_i^2$ (p_i はローカス i で混合資料の DNA 型に含まれるアレルの出現頻度の和) であるから、

$$1 - \prod (1 - PE_i) = 1 - \prod p_i^2$$

であり、すなわち、総合排除率は、1 から、本文 6(1)a で求めた、現場資料の形成者の資格をもつ人の出現頻度 (本文の表記では $(\sum p_i)^2$) を減じたものである。現場資料の DNA 型が本文第 4 図のとおりであるとき、その出現頻度は 3.2873 E-09 であるが、総合排除率は、 $1 - (3.2873 \text{ E-}09) = 99.99999967\ldots\%$ となる。

10 Butler 453 頁参照

11 これらの式は、Butler 454 頁にも、米国・国立研究評議会の報告 (NRC II) からの引用として掲載されているが、本稿では、その導出方法を説明した。