

混合試料の構成人数 Nuisance パラメタ

法数学勉強会

2019/04/20

京都大学(医)統計遺伝学分野

山田 亮

ネタが切れました…

法医学講座 玉木研からお題をいただきました

2016年
8月10日

ELSEVIER

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Correspondence

About the number of contributors to a forensic sample*Dear Editor,*

According to the ISFG recommendations [1], “There is a common misconception that the numbers of contributors under H_p and H_d should be the same. There is no requirement for this” (a footnote further elaborates: “Other things being equal, the aim of the defense is to maximize the probability of the evidence under H_d . Similarly, the prosecution aim is to maximize the probability of the evidence under H_p , consistent with their theory of the case. The number of contributors is a secondary consideration ...”).

population structure (F_{st} or θ). The v cannot be estimated only from the data must be supplied using external (previous) information. In this regard, prosecution and defense have different interests, since low values of F_{st} and θ are favorable to the prosecution and by the person of interest and the evidence support to the prosecution hypothesis is favorable to the defense. Consultants contend about the values that should be used, and they often even finish in disagreement.

The above ISFG recommendation is perfectly acceptable to calculate a ratio computed with different sets of parameters. Such a likelihood ratio is uninterpretable in any way to present the results in case of



Contents lists available at ScienceDirect

Correspondence

About the number of contributors to a sample

Dear Editor,

According to the ISFG, a common misconception that the number of contributors to a sample (Hp and Hd) should be the same (a footnote further elaborates: "Other things being equal, the aim of the defense is to maximize the probability of the evidence under Hd. Similarly, the prosecution aim is to maximize the probability of the evidence under Hp, consistent with their theory of the case. The number of contributors is a secondary consideration ...").

ISFG推奨ってというのがあって
検察側の仮説
Vs.
弁護側の仮説

それらの尤度は
「同条件」
にした上で
尤度比を計算するべきでは？

The v
the da
l (prev
on an
st and
the ev
hthesis
sultants
t should

even minor in disagreement.

The above ISFG recommendation is perfectly acceptable to calculate a ratio computed with different sets of parameters. Such a likelihood ratio is uninterpretable in any way to present the results in case of

DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures

P. Gill^{a,*}, C.H. Brenner^b, J.S. Buckleton^c, A. Carracedo^d, M. Krawczak^e, W.R. Mayr^f,
N. Morling^g, M. Prinz^h, P.M. Schneiderⁱ, B.S. Weir^j

^a Forensic Science Service, Trident Court, 2960 Solihull Parkway, Birmingham, UK

^b Forensic Science Group, School of Public Health, University of California, Berkeley, CA 510-339-1911, USA

^c ESR, Private Bag 92021, Auckland, New Zealand

^d Institute of Legal Medicine, Faculty of Medicine, University of Santiago de Compostela, 15705 Santiago de Compostela, Spain

^e Institute of Medical Informatics and Statistics, Kiel, Germany

^f Division of Blood Group Serology, Medical University of Vienna, Austria

^g Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark

^h Office of the Chief Medical Examiner, Department of Forensic Biology, 520 First Avenue, New York, NY 10016, USA

ⁱ Institute of Legal Medicine University Clinic of Cologne, Melatengürtel, 60-62 D 50823 Köln, Germany

^j University of Washington, Department of Biostatistics, Box 357232, Seattle, WA 98195, USA

Received 4 April 2006; accepted 10 April 2006

Available online 5 June 2006

Abstract

The DNA commission of the International Society of Forensic Genetics (ISFG) was convened at the 21st congress of the International Society for Forensic Genetics held between 13 and 17 September in the Azores, Portugal. The purpose of the group was to agree on guidelines to encourage best practice that can be universally applied to assist with mixture interpretation. In addition the commission was tasked to provide guidance on low copy number (LCN) reporting. Our discussions have highlighted a significant need for continuing education and research into this area. We have attempted to present a consensus from experts but to be practical we do not claim to have conveyed a clear vision in every respect in this difficult subject. For this reason, we propose to allow a period of time for feedback and reflection by the scientific community. Then the DNA commission will meet again to consider further recommendations.

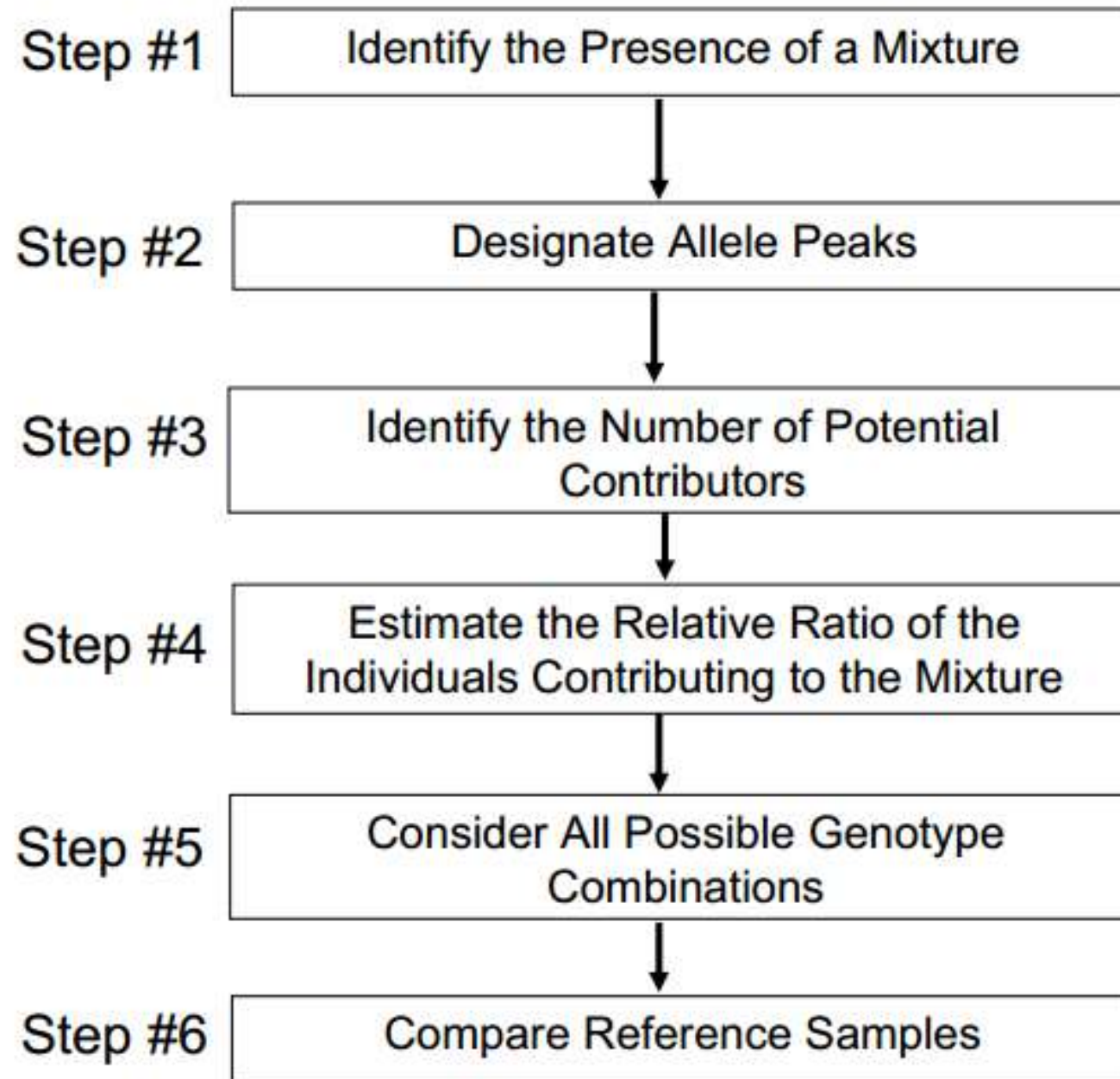
© 2006 Elsevier Ireland Ltd. All rights reserved.

Keywords: STR typing; Biostatistical analysis; Likelihood ratio; Probability of exclusion; Mixtures; ISFG DNA commission

2006年

Steps in the interpretation of mixtures

(Clayton *et al.*
Forensic Sci. Int.
1998; 91:55-70)



https://www.isfg.org/files/SFG2007_Statistics_Gill_ISFG_Recommendations.pdf

Peter Gill (Norway)のコメント

2016年
9月30日

オリジナルコレスポンドンスが1
ページなのに対して、
このコメントは4ページ。
『部分を取り出して問題点を指摘
せず、全体の文脈から適切な判断
を』

Charles Brenner (USA) のコメント

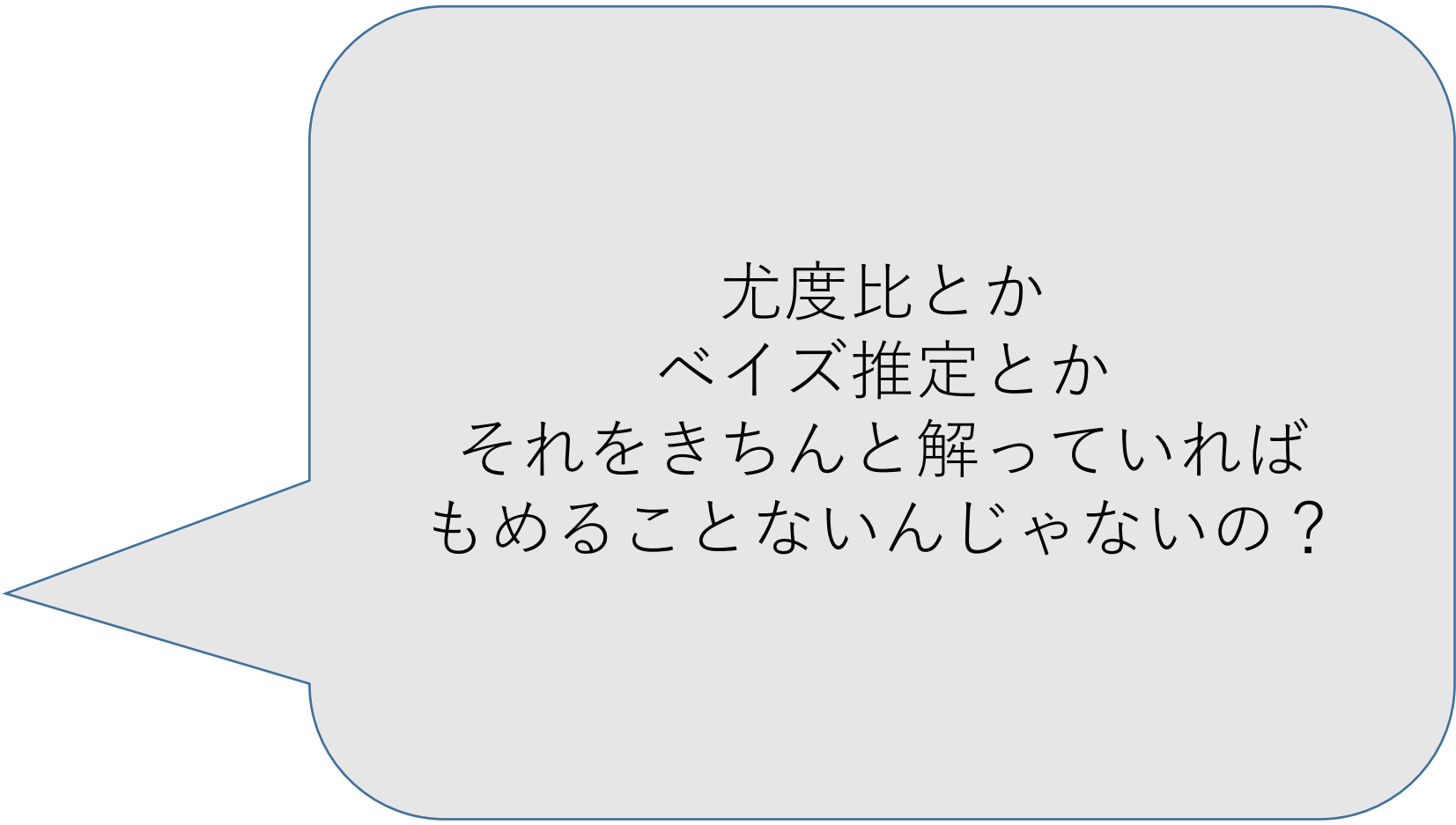
2016年
10月18日

混合試料の構成人数のように不確
かなものを『情報として使う』の
は難しく
それに基づく『条件付確率』は無
理なんだからしかたないじゃない
...

とやや、無理押しなコメント

Ian Evett (UKE)のコメント

2017年
1月7日



尤度比とか
ベイズ推定とか
それをきちんと解っていれば
もめることないんじゃないの？

少し、気合を入れて考えてみよう

Forensic Science International: Genetics 37 (2018) 116–125



Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen



Research paper

Contributors are a nuisance (parameter) for DNA mixture evidence evaluation

K. Slooten^{a,*}, A. Caliebe^b

^a VU University Amsterdam, Netherlands Forensic Institute (NFI), The Netherlands

^b Institut für Medizinische Informatik und Statistik, Christian-Albrechts-Universität zu Kiel, Germany



ARTICLE INFO

Keywords:

DNA mixtures
Number of contributors
Weight of evidence
Nuisance parameters

ABSTRACT

Recently, a debate has arisen around the number of contributors postulated in hypotheses for the purpose of weight of evidence calculations on DNA mixture profiles. Specifically the issue at stake is whether or not one should have the same number of contributors under both hypotheses for which a likelihood ratio is calculated. In this paper we aim to clarify this issue. We take the general approach of considering the number of contributors as a nuisance parameter. Two central assumptions then determine the form of the overall likelihood ratio: whether the prior distributions of the nuisance parameter are equivalent given both hypotheses and whether they depend on the hypotheses. Examples are given for both scenarios where we have either independence or strong dependence between the prior distributions of the number of contributors and the hypotheses. Moreover, examples for different kinship scenarios are presented. In conclusion, the overall likelihood ratio does not only depend on likelihood ratios for fixed values of the nuisance parameter but may also vary considerably with different prior distributions.

Nuisance parameter

- 厄介もの、邪魔くさい
- 「知りたいことそのもの」ではないけれど
- 「確率モデルに入り込む」パラメタ
- 今回の文脈では、「混合試料の構成人数」
 - じゃまくさい、1人でも2人でも3人でも、どうでもええやん！

Nuisance parameter

- 厄介もの、邪魔くさい
- 「知りたいことそのもの」ではないけれど
- 「確率モデルに入り込む」パラメタ
- 今回の文脈では、「混合試料の構成人数」
 - じゃまくさい、1人でも2人でも3人でも、どうでもええやん！
 - . . . と、いうわけに行くか、という話

京大 統計遺伝学分野的 結論

Nuisance Parameter(s)

混合試料構成人数

を

無視したり・勝手に固定したり

してよいわけがない

極端な例をいくつか考えてみる

あるローカス

- アレルが3つ (A,B,C) あった
 - 弁護側
 - 『(もろもろを勘案して)寄与者は1人である』
 - 『寄与者1人の下で、被疑者1名が3つのアレルを残すことは不可能である』
 - 『尤度は0である』
 - 検察側
 - 『(もろもろを勘案して)寄与者は2人である』
 - 『寄与者のうちの1人が被疑者であるとしてい矛盾しない』
 - 『尤度は十分に大きい』
- まったくかみ合わない～

3 アレル検出

寄与人数 = 1 人という主張がナンセンス

- 確かに . . .

3 アレル検出

寄与人数 = 1 人という主張がナンセンス

- 確かに . . .
- では「寄与人数 = 2 人」という主張はどれほどナンセンスから遠いのか ???

3 アレル検出

寄与人数 = 1 人という主張がナンセンス

- 確かに・・・
- どうして「ナンセンス」なのか
- 「寄与人数 = 1 人」という『事前確率』は 0 ではなかった
- しかるに、「データ」が指し示す尤度を考慮したら
- 「寄与人数 = 1 人」という仮説の『事後確率』が 0 になったことを「ナンセンス」と称している

3 アレル検出

寄与人数 = 1 人という主張がナンセンス

- 確かに . . .
- どうして「ナンセンス」なのか
- 「寄与人数 = 1 人」という『事前確率』は 0 ではなかった
- しかるに、「データ」が指し示す尤度を考慮したら
- 「寄与人数 = 1 人」という仮説の『事後確率』が 0 になったことを「ナンセンス」と称している
- 「寄与人数 = 1 人」という仮説を維持しつつ、3 アレルを許す仮説 (たとえば、実験上の理由で第三アレルのシグナルが検出される) はナンセンスではなくなる

小総括

- 仮説(モデル)が単純に過ぎる(実験エラー等を許さない)と、事後確率が0になる場合が増えてしまう
- モデルをある程度複雑にする～現実近づけると、色々な仮説に0でない事後確率が得られる
- 2つの仮説の比較なら、2つの尤度を比べればよい
- 2つの尤度の比べ方には「尤度比」という手がある
- 3つ以上の仮説が尤度を持ってしまったら『尤度比』でよいのか？

寄与人数を推定して決めうちにする事

- 寄与人数が不明であるとき、1人か、2人か、3人か・・・が決まらない
- とはいえ、最大尤度をもたらすのは、 n 人である、という計算はできる
- このとき、 n を寄与人数の最尤推定値という
- Nuisance パラメタ(寄与人数)の最尤推定値を一つ取り出して、それを固定し、条件付き尤度を計算するとはどういうことか？

最尤推定

- Nuisance parameter(寄与人数)も含めて、諸々のパラメタのすべてについて、尤度を最大にする値を推定することはできる
- (複数パラメタの)最尤推定
- 最尤推定結果は、「もっともありそうな仮説」の「そのもっともらしさ」
- 検察側 vs. 弁護側とで、最尤推定をそれぞれ行って、その尤度を比較するのは
- 最大尤度の比較。最大尤度比
- この比を決断の根拠にしてよいのか ???

Nuisance parameterが離散的な場合

- 寄与人数が1のときの、検察仮説尤度/弁護仮説尤度 = LR1
- 2 = LR2
- 3 = LR3
- ...
- すべての i について、 $LR_i > Q$ なら
- 絶対に
 - 検察仮説尤度/弁護仮説尤度 $> Q$ が言える

Nuisance parameterが「離散的な場合

- 寄与人数が1のときの、検察仮説尤度/弁護仮説尤度 = LR1
- 2 = LR2
- 3 = LR3
- ...
- すべての i について、 $LR_i > Q$ とは限らないときは？
- $i = 1, 2, 3, \dots$ について、重み付き平均をとるしかない
- その重みは i の値の『事前分布』

Nuisance parameterが「離散的な場合

- 寄与人数が1のときの、検察仮説尤度/弁護仮説尤度 = LR1
- 2 = LR2
- 3 = LR3
- ...
- すべての i について、 $LR_i > Q$ とは限らないときは？
- $i = 1, 2, 3, \dots$ について、重み付き平均をとるしかない
- その重みは i の値の『事前分布』・・・それってわかるの？

i の事前分布がわからないとき...

- 困ってしまう...
- じゃあ、やっぱり、 i も最尤推定してしまう???

パラメタをいじるときには、要注意！

- モデルが複雑になると、尤度は上がります
 - オッカムのかみそり
 - 赤池の情報量基準
 - バイアス・バリアンスのトレードオフ
-
- 統計学・データサイエンスのいたるところで問題になるテーマ

尤度・確率、確率密度

- 最尤推定は
 - 「尤度関数の値を最大にするパラメタ値を推定する」
- 尤度関数と、確率密度関数は同じ
- 「確率密度関数」の値は、「確率」を教えてくれない
- 「幅があって」初めて「確率」が0ではなくなる
- 「尤度～事後確率」も「幅があって」こそ
- 「幅のある」「0でない」尤度の計算は、「最尤推定」だけでは無理

小総括

- 尤度を比べたい
- 2つの画然とした仮説を比較するなら、尤度比を使う
- Nuisance parameterがあるなら、Nuisance parameterの値を決めうちにするのは、難がある
- 特にNuisance parameterが連続値をとるときは、要注意
- 離散的なNuisance parameterでも、複数のNuisance parameter値について合算するのが無難
- 合算するには、重みが必要・・・事前分布～ある程度の恣意性が入る？