

# 3

## 潜在結果変数， 割付けメカニズム，統計的推測

この章では、Rubin 因果モデルの 3 つの要素である潜在結果変数、割付けメカニズム、確率分布について述べる。このモデルの特徴は、潜在結果変数の比較に基づいて、因果効果を定義することである。因果効果を識別するためには、割付けメカニズムについて 2 つの仮定が必要になる。1 つ目は非交絡性、2 つ目はオーバーラップである。また、この章の表記法は、一致性・非干渉性（あわせて Stable Unit Treatment Value Assumption と呼ばれる）を、暗黙の裡に仮定している。

一般に、観察研究では非交絡性が満たされる保証はない。また、ランダム化臨床試験では、ランダム化という操作によって非交絡性が成り立つとされているが、単純ランダム化とアウトカム適応的ランダム化では、統計的推測の違いが生じることに注意すべきである。

### キーワード

アウトカム適応的ランダム化、一致性、オーバーラップ、解析単位レベルの因果効果、干渉、効果修飾因子、効果の均一性、コントロールにおける非交絡性、弱オーバーラップ、条件付因果リスク差、Stable Unit Treatment Value Assumption (SUTVA)、正值性、潜在結果変数、治療群における母集団因果リスク差、治療群における有限集団因果リスク差、強い無視可能性、比較可能性、非干渉性、非交絡、プロペンシティスコア、併合可能性 (collapsibility)、Bayes 流の推測、母集団因果リスク差、無視可能性、有限集団因果リスク差、尤度に基づく推測、割付けメカニズム

### 事例

ハーバード ECMO 試験、ミシガン ECMO 試験

### 3.1 潜在結果変数と因果リスク差

医学において、原因と結果の関係について言及するとき、しばしば反事実の形をとる。ECMO 臨床試験 (Bartlett, et al. 1985, O'Rourke, et al. 1989, UK Collaborative ECMO Trial Group 1996) は、どれも ECMO によって死亡率を減らせるかを調べるのが目的だった。これを研究仮説にすると

仮説：遷延性肺高血圧症の新生児に ECMO を用いて酸素を供給していたら、命が助かったか

となるだろう。一部の患者では実際には ECMO は用いられなかったのだから、これは事実に反する命題についての推測である。

Rubin 因果モデルでは、この問題を定式化するために、観測されないアウトカム (潜在結果変数, potential outcomes) を導入する (Neyman 1923, Rubin 1974)。ECMO ではなく従来療法を用いたある個人について、ECMO を用いたときのアウトカムは、実際には観測することのできない潜在的な (事実に反する) 変数である。これを導入することで、反事実に基づく因果効果を定義することができる<sup>\*1)</sup>。

対象者  $i$  について観察されるアウトカムを、0 または 1 の値をとる 2 値変数  $Y_i$  で表す。また、関心のある試験治療とコントロール治療に対応する潜在結果変数をそれぞれ  $Y_i^0$  と  $Y_i^1$  と定義する。ここでは便宜上、試験治療とコントロール治療という呼び方をするが、疫学研究などでは曝露・非曝露といった方が適切なときもある。試験治療がアウトカムに与える因果効果は、両者に差があるかどうか、つまり

$$\tau_i = Y_i^1 - Y_i^0$$

によって定義することができる。 $\tau_i$  を解析単位レベルの因果効果という。これは、同一個人がコントロール治療を受けたときと試験治療を受けたときの差を表している。現実の医学研究では、ある個人について、両方の潜在結果変数つ

<sup>\*1)</sup> 潜在結果変数の起源は Neyman (1923) とされており、初期の実験計画法のテキストにも記載がみられる (Fisher 1935, Kempthorne 1952, Cox 1958)。その後、これらのアイデアを潜在結果変数、割付けメカニズム、確率分布という 3 つの概念で整理し、観察研究を含めて定式化したものを Rubin 因果モデルと呼んでいる。この章で述べる一貫性、非干渉性、効果の均一性は、Rubin (1974, 1978) によって導入された概念である。

まり  $Y_i^1$  と  $Y_i^0$  を同時に観測することはできない。そこで,  $N$  人からなる有限集団を考え, 統計的に推測を行う。集団全体の潜在結果変数は

$$\begin{aligned} Y_i^0 &= Y_i^1 = 0 \\ Y_i^0 &= 0, Y_i^1 = 1 \\ Y_i^0 &= 1, Y_i^1 = 0 \\ Y_i^0 &= Y_i^1 = 1 \end{aligned}$$

の 4 パターンに分類できる。これらのパターンの人数を, それぞれ  $n_{00}$ ,  $n_{01}$ ,  $n_{10}$ ,  $n_{11}$  で表す。対象者全員に試験治療を用いたときとコントロール治療を用いたときの比較は, 有限集団因果リスク差 (finite-population causal risk difference)

$$\bar{\tau} = \frac{\sum_{i=1}^N Y_i^1}{N} - \frac{\sum_{i=1}^N Y_i^0}{N} = \frac{n_{01}}{N} - \frac{n_{10}}{N}$$

により行うこととする。通常のリスク差と異なり, 分母が対象者全員である点に注意してほしい。

さらに区別が必要なのは, 有限集団と無限母集団の違いである。 $N$  人全員の潜在結果変数の組  $\{Y_i^0, Y_i^1; i = 1, \dots, N\}$  を無限母集団からのランダムサンプルとみなすことがある。そのときの  $\tau_i$  の期待値は,  $\Pr(Y_i^0 = 1) = \pi^0$  と  $\Pr(Y_i^1 = 1) = \pi^1$  の差になる。このように定義された

$$\tau = E(Y_i^1 - Y_i^0) = \pi^1 - \pi^0$$

を母集団因果リスク差 (super-population causal risk difference) または (連続データでは) 母集団平均因果効果 (super-population average causal effect) という。この章では一例として因果リスク差に注目するが, 因果リスク比は  $\pi^1/\pi^0$ , 因果オッズ比は  $\{\pi^1/(1-\pi^1)\}/\{\pi^0/(1-\pi^0)\}$  と定義することができる。

さらに, 実際に試験治療群に組み込まれた  $N_1$  人に注目し, その集団に試験治療を用いたときとコントロール治療を用いたときの因果リスク差を定義することができる。それはそれぞれ, 治療群における有限集団因果リスク差 (finite-population causal risk difference for the treated)

$$\bar{\tau}^T = \frac{\sum_{i: A_i=1} Y_i^1}{N_1} - \frac{\sum_{i: A_i=1} Y_i^0}{N_1}$$

と治療群における母集団因果リスク差 (super-population causal risk difference for the treated)

$$\tau^T = E(Y_i^1 - Y_i^0 | A_i = 1)$$

表 3-1 本書で扱う因果パラメータ

種類	定義	識別	本書の該当箇所
<b>解析単位レベル</b>			
解析単位レベルの因果効果	$\tau_i = Y_i^1 - Y_i^0$	難	3 章
均一性の下での因果効果	$\tau^H = Y_i^1 - Y_i^0$ for all $i$	易	13 章
時間依存性治療効果	$Y_i^{ab} - Y_i^{cd}$	難	9 章
制御された直接効果	$Y_i^{1m} - Y_i^{0m}$	難	10 章
自然な直接効果	$Y_i^{1,M_i^0} - Y_i^{0,M_i^0}$	難	10 章
自然な間接効果	$Y_i^{1,M_i^1} - Y_i^{1,M_i^0}$	難	10 章
<b>ターゲット集団：全体</b>			
有限集団平均因果効果 (因果リスク差)	$\bar{\tau} = \frac{\sum_{i=1}^N Y_i^1}{N} - \frac{\sum_{i=1}^N Y_i^0}{N}$	易	3 章
母集団平均因果効果 (因果リスク差)	$\tau = E(Y_i^1 - Y_i^0) = \pi^1 - \pi^0$	易	3 章
母集団因果リスク比*	$\exp(\psi_0) = \frac{\Pr(Y_i^1=1)}{\Pr(Y_i^0=1)} = \frac{\pi^1}{\pi^0}$	易	3, 13 章
母集団因果オッズ比	$\frac{\pi^1(1-\pi^0)}{\pi^0(1-\pi^1)}$	易	3 章
介入を伴う直接効果	$E(Y_i^{a,G_{a^*}} - Y_i^{a^*,G_{a^*}})$	難	10 章
介入を伴う間接効果	$E(Y_i^{a,G_{a^*}} - Y_i^{a,G_{a^*}})$	難	10 章
<b>ターゲット集団 <math>\{i : A_i = 1\}</math></b>			
治療群における有限集団 平均因果効果 (因果リスク差)	$\bar{\tau}^T = \frac{\sum_{i: A_i=1} Y_i^1}{N_1} - \frac{\sum_{i: A_i=1} Y_i^0}{N_1}$	易	3 章
治療群における母集団 平均因果効果 (因果リスク差)	$\tau^T = E(Y_i^1 - Y_i^0   A_i = 1)$	易	3 章
治療群における母集団 因果リスク比*	$\exp(\psi_0) = \frac{\Pr(Y_i^1=1   A_i=1)}{\Pr(Y_i^0=1   A_i=1)}$	易	3, 13 章
<b>ターゲット集団 <math>\{i : L_i = l\}</math></b>			
条件付有限集団 平均因果効果 (因果リスク差)	$\bar{\tau}(l) = \frac{\sum_{i: L_i=l} Y_i^1}{N_l} - \frac{\sum_{i: L_i=l} Y_i^0}{N_l}$	易	3 章
条件付母集団 平均因果効果 (因果リスク差)	$\tau(l) = E(Y_i^1 - Y_i^0   L_i = l)$	易	3 章
条件付母集団因果リスク比	$\varphi(l) = \frac{\Pr(Y_i^1=1   L_i=l)}{\Pr(Y_i^0=1   L_i=l)}$	易	3 章
<b>ターゲット集団 <math>\{i : A_i^0 = 0, A_i^1 = 1\}</math></b>			
服用遵守者における平均因果効果 (因果リスク差)	$\tau^C = E(Y_i^1 - Y_i^0   A_i^0 = 0, A_i^1 = 1)$	難	13 章
<b>周辺構造モデル</b>			
構造平均モデル	$\psi_0, \psi_1, \psi_2, \psi_3$	—	8, 9, 13 章

識別の難易度：非交絡性・オーバーラップが満たされれば識別可能なものは易、不可能なものは難とした。

\* $\psi_0$  は以下の対数線型構造平均モデルのパラメータである。

$$\frac{\Pr(Y_i^a=1 | A_i=1, Z_i)}{\Pr(Y_i^0=1 | A_i=1, Z_i)} = \exp[(\psi_0 + \psi_1 Z_i) a]$$

と表される。

表 3-1 は本書で扱う因果効果の一覧である。必要に応じて参照してほしい。

#### コントロールの重要性

ここで述べたように、Rubin 因果モデルにおける因果効果は、コントロール（比較対照群）の選び方に依存する。

環境疫学や放射線疫学を通じて因果関係を調べることがある。これらの研究では、特定の地域で生じた有害な曝露を研究対象とすることが多く、研究を実施するにあたってコントロールの選択基準をどのように設定するかが問題になる。理想的には、地域住民のうち曝露を受けておらず、しかも曝露集団と特徴が近い集団（内部対照）を特定したい。しかしながら、住民ひとりひとりの曝露状況は研究とは無関係に決まるものなので、理想的なコントロールを自由に選択できるわけではない。別の地域との比較（外部対照）に頼らざるを得ないこともある。内部対照・外部対照の選択はよくある問題の一例である。コントロールの選び方に関する批判は、疫学にはつきものである。

## 3.2

### ランダム化の下での因果リスク差の推定

対象者  $i$  に試験治療を用いたかどうかを  $A_i$  ( $A_i = 1$  なら試験治療群,  $A_i = 0$  ならコントロール治療), アウトカムを  $Y_i$  ( $Y_i = 0$  ならイベントなし,  $Y_i = 1$  ならイベントあり) で表す。また、治療前に測定された共変量 (covariate) を  $L_i$  で表す。 $L_i$  を省略すれば、データは表 3-2 のような  $2 \times 2$  表に要約できる。 $S_1$ ,  $S_0$ ,  $S$  はそれぞれ試験治療群, コントロール群, 全体のイベント数である。

ここで、試験治療とコントロール治療がランダムに割付けられていると仮定する。このとき、いわゆるリスク差は

$$\hat{\tau} = \frac{S_1}{N_1} - \frac{S_0}{N_0}$$

で与えられ、これを有限集団または母集団因果リスク差の推定量として用いることができる。

これまで述べたように、潜在結果変数は 4 パターンに分類できる\*2)。各パ

\*2) 潜在結果変数のパターンに基づく層別を、主要層別という (11 ~ 13 章)。ただし、特定の対象者  $i$  が、潜在結果変数の 4 パターンのどれに該当するかはわからない。それだけではなく、表

ターンの試験治療群の人数を、それぞれ  $x_{00}$ ,  $x_{10}$ ,  $x_{01}$ ,  $x_{11}$  で表すと、表 3-3 のようにまとめることができる。この表は、 $N$  組の  $Y_i^0$  と  $Y_i^1$  から構成される有限集団のパターンを表している。つまり、各パターンの人数  $n_{00}$ ,  $n_{01}$ ,  $n_{10}$ ,  $n_{11}$  がわかれば（もちろん  $n_{00} + n_{01} + n_{10} + n_{11} = N$  という制約がある）、この有限集団で因果効果があるかが判断できるから、これが推測の対象（未知パラメータ）である。一方で治療をランダムに割付けるとき、 $x_{00}$ ,  $x_{10}$ ,  $x_{01}$ ,  $x_{11}$  は確率変数とみなすことができる。

繰り返しランダム割付けを行ってデータを発生させた仮想的な状況を考えてみよう。ある個人がどちらの治療に割付けられるかによって、 $x_{00}$ ,  $x_{10}$ ,  $x_{01}$ ,  $x_{11}$  の値がばらつく。観測データ（表 3-2）でいえば、割付け結果によって、 $S_1 = x_{01} + x_{11}$  や  $S_0 = n_{10} - x_{10} + n_{11} - x_{11}$  が確率的に変動する。ランダム化に基づく推測では、これらの確率的変動が、因果リスク差を推定するときの誤差の源になる。

表 3-2 観測データの記法

	試験治療 ( $A_i = 1$ )	コントロール ( $A_i = 0$ )	合計
イベントなし ( $Y_i = 0$ )	$N_1 - S_1$	$N_0 - S_0$	
イベントあり ( $Y_i = 1$ )	$S_1$	$S_0$	$S$
合計	$N_1$	$N_0$	$N$

表 3-3 潜在結果変数の記法

	試験治療 ( $A_i = 1$ )	コントロール ( $A_i = 0$ )	合計
常にイベントなし ( $Y_i^0 = Y_i^1 = 0$ )	$x_{00}$	$n_{00} - x_{00}$	$n_{00}$
試験治療のみイベントあり ( $Y_i^0 = 0, Y_i^1 = 1$ )	$x_{01}$	$n_{01} - x_{01}$	$n_{01}$
コントロール治療のみ イベントあり ( $Y_i^0 = 1, Y_i^1 = 0$ )	$x_{10}$	$n_{10} - x_{10}$	$n_{10}$
常にイベントあり ( $Y_i^0 = Y_i^1 = 1$ )	$x_{11}$	$n_{11} - x_{11}$	$n_{11}$
合計	$N_1$	$N_0$	$N$

3-3 のセル度数と観測されるイベント数とは、 $S_1 = x_{01} + x_{11}$ ,  $S_0 = n_{10} - x_{10} + n_{11} - x_{11}$  という関係があるものの、セル度数をすべて知ることはできない。

### 3.3 一貫性, 非干渉性, 効果の均一性

割合の差を計算することで因果リスク差を推定できると述べたが, そこにはいくつかの仮定が隠れている. そこで, 対象者  $i$  と  $j$  の 2 人がいて, それぞれ ECMO と従来療法により治療されたという単純化した状況で, ここまでの議論を振り返ってみよう. 両者の比較  $Y_i - Y_j$  が因果効果としてなんらかの意味を持つための条件はなんだろうか. 結論からいうと, (3.7 節の非交絡性・オーバーラップに加えて)

- 一貫性 (consistency)
- 非干渉性 (no interference between unit)
- 比較可能性 (comparability)
- 効果の均一性 (between-unit homogeneity of effects)

が必要になる (Rubin 1978).

一貫性とは, 観察されるデータと潜在結果変数が,  $Y_i = (1 - A_i)Y_i^0 + A_iY_i^1$  という関係でリンクしていることを意味する. わかりやすくいうと, 治療  $A_i = 0$  を受けると  $Y_i^0$  が観察され,  $A_i = 1$  だと  $Y_i^1$  が観察されるという反事実の世界と現実世界を結ぶ仮定である. このとき,  $Y_i - Y_j = Y_i^1 - Y_j^0$  のように, 対象者  $i$  と  $j$  の 2 人を比べることで, 潜在結果変数の比較を行うことができる. それでは, どのようなときに一貫性が成り立たないのだろうか. それはたとえば,  $K$  人の術者がいて, ECMO または従来療法の成績が異なるときである. このとき, 仮に対象者  $i$  が術者  $k$  ( $k = 1, \dots, K$ ) により治療  $A_i = a$  を受けたときの潜在結果変数を定義するには,  $Y_i^{ak}$  のように術者と治療の組み合わせを表現する必要がある (VanderWeele and Hernán 2013). ここから明らかなように, 同じ治療  $A_i = a$  であっても別のバージョン ( $k = 1, \dots, K$ ) が存在するとき (これを multiple versions of treatment という),  $Y_i = (1 - A_i)Y_i^0 + A_iY_i^1$  という単純な関係にはならない.

次に非干渉性とは, 他の解析単位の治療が, 別の解析単位に影響しないという仮定である. 逆に影響があることを干渉 (interference) と呼んでいる (Cox 1958). 仮に対象者  $i$  が治療  $A_i = a$  を, 対象者  $j$  が治療  $A_j = b$  を受けたときの, 対象者  $i$  のアウトカムを  $Y_i^{ab}$  と定義する. これが, 対象者  $j$  の治療が, 対象者  $i$  に影響を干渉するときのモデルである. このモデルが意味を持つのは, たとえば

対象者  $i$  と  $j$  の ECMO を行う術者が同じで、同じ治療を続けて行くと技術が向上するような状況である。なぜなら、別の治療を行ったとき ( $A_i = a \neq A_j = b$ ) のアウトカム  $Y_i^{ab}$  よりも、 $Y_i^{aa}$  の方がよいかもしれないからである。別の例として、同じ地域内に住む対象者間の感染や、1 人の対象者の治療部位（たとえば右目と左目）が挙げられる。このように、干渉は一定の範囲内に起こることが多い。そのとき、治療  $A_i$  だけによる効果と、干渉のある範囲内に共通に起こる効果を、区別する必要がある。非干渉性はこの問題を避けるための仮定といえる。

一致性と非干渉性は、合わせて Stable Unit Treatment Value Assumption (SUTVA) と呼ばれることもあるし、別の仮定が論じられることもある（黒木・小林 2011, VanderWeele and Hernán 2013）。いずれにしても、これらの前提をおけば、潜在結果変数を、 $\{Y_i^0, Y_i^1, Y_j^0, Y_j^1\}$  と通常の記法で表すことができる。そして、一致性により  $Y_i^0$  と  $Y_i^1$  のうちどちらかは観測されることになり、もう一方は欠測データとなる。

一致性と非干渉性が成り立っていたとしても、 $Y_i - Y_j$  という比較は、個人の因果効果 ( $\tau_i = Y_i^1 - Y_i^0$  または  $\tau_j = Y_j^1 - Y_j^0$ ) とは一致しない。逆に一致するための条件を考えると、 $Y_i^0 = Y_j^0$  と  $\tau_i = \tau_j$  が成り立っていればよいことがわかる。前者は、対象者  $i$  と  $j$  の疾患の状態に差がないことを意味する。これは 6 章で交絡と関連付けて論じる比較可能性という概念に相当する。後者は、解析単位間で因果効果が等しいことであり、これを効果の均一性と呼ぶ<sup>\*3)</sup>。たとえば投与すればかならず出血を止められる止血剤を考えてみよう。止血の有無に与える効果はどの個人においても均一であろう。しかし、実際には効果に個人差がない治療の方がまれである。統計的にみても、効果の均一性は、 $Y_i^0$  と  $Y_i^1$  の相関が 1 ということを意味するから強い仮定である。

### 3.4 条件付因果効果と効果の修飾

#### ■ 3.4.1 条件付因果リスク差

状況によっては、効果が均一とは考えにくく、しかも共変量によって効果の大きさや方向が変わるようなこともあり得る。このような共変量を効果修飾因

<sup>\*3)</sup> 本書では、効果の均一性の下での因果効果を、 $\tau^H = \tau_i = \tau_j$  と表している。



子 (effect modifier) という。潜在結果変数を用いて表現するなら、 $\tau_i = Y_i^1 - Y_i^0$  と相関する因子を効果修飾因子といってもよい。効果の修飾の有無は、共変量  $L_i$  に基づくサブグループを設定し、サブグループ間で因果効果が不均一かどうかによって検討することが多い。この場合、因果効果として、共変量  $L_i$  によって条件付けた因果リスク差

$$\tau(l) = E(Y_i^1 - Y_i^0 | L_i = l)$$

$$\bar{\tau}(l) = \frac{\sum_{i:L_i=l} Y_i^1}{N_l} - \frac{\sum_{i:L_i=l} Y_i^0}{N_l}$$

を用いることができる (有限集団条件付因果リスク差の分母  $N_l$  は共変量が  $L_i = l$  であるような人数である)。そして効果の修飾は、共変量  $L_i = 0$  と  $L_i = 1$  の2つのサブグループがあるときを例にすると、各サブグループの因果リスク差の差

$$\tau(1) - \tau(0)$$

$$\bar{\tau}(1) - \bar{\tau}(0)$$

がゼロかどうかにより定義することができる。また、母集団因果リスク差と条件付因果リスク差には

$$\tau = E(Y_i^1 - Y_i^0) = \sum_l \tau(l) \Pr(L_i = l)$$

$$\bar{\tau} = \frac{1}{N} \sum_{i=1}^N \bar{\tau}(L_i)$$

という関係がある。つまり、母集団因果リスク差は、条件付因果リスク差を対象者  $N$  人について平均したものである。

#### ■ 3.4.2 条件付因果リスク比

アウトカム  $Y_i$  が2値のとき、差の指標ではなく、母集団因果リスク比  $\pi^1/\pi^0$  や母集団因果オッズ比  $\{\pi^1/(1-\pi^1)\}/\{\pi^0/(1-\pi^0)\}$  を用いることもある。しかし比の指標を扱うときは、併合可能性 (collapsibility) について注意しなければならない。ここでいう併合可能な指標とは、効果の修飾がないとき、集団全体と一部の集団で、数値的に一致するような因果効果のことである。

すでに述べたように条件付因果リスク差と母集団因果リスク差には

$$\tau = \sum_l \tau(l) \Pr(L_i = l)$$

という関係がある。条件付因果リスク比を

$$\varphi(l) = \frac{\Pr(Y_i^1 = 1 | L_i = l)}{\Pr(Y_i^0 = 1 | L_i = l)}$$

と定義して、母集団因果リスク比との関係が、リスク差のケースと同じになるか調べてみよう。母集団因果リスク比は

$$\begin{aligned} \frac{\Pr(Y_i^1 = 1)}{\Pr(Y_i^0 = 1)} &= \frac{\sum_l \Pr(Y_i^1 = 1 | L_i = l) \Pr(L_i = l)}{\sum_l \Pr(Y_i^0 = 1 | L_i = l) \Pr(L_i = l)} \\ &= \frac{\sum_l \varphi(l) \Pr(Y_i^0 = 1 | L_i = l) \Pr(L_i = l)}{\sum_l \Pr(Y_i^0 = 1 | L_i = l) \Pr(L_i = l)} \end{aligned}$$

と表される。ここで、効果の修飾がないと仮定しよう。これは条件付因果リスク比が定数ということだから、 $\varphi(l) = \varphi$ と書くことができる。このとき上の式から

$$\frac{\Pr(Y_i^1 = 1)}{\Pr(Y_i^0 = 1)} = \varphi$$

という関係が導かれる。これは、母集団因果リスク比（左辺）と条件付因果リスク比（右辺）が等しい、つまり母集団因果リスク比が併合可能であることを意味する。

母集団因果オッズ比など他の因果効果では、併合可能性はかならずしも成り立たない。第1巻14章で述べたように、一般化線型モデルに基づいて効果の指標を定義したときも、同様の結果が得られる。

### ■ 3.4.3 数値例

表3-4は、有限集団因果効果の場合分けによって、効果の修飾のパターンを整理したものである。アウトカム・治療・共変量がすべて2値のとき、個人レベルの潜在結果変数  $\{Y_i^0, Y_i^1\}$  のパターンは4通りあるから、サブグループが2つのとき、組み合わせは全部で16通りになる。効果の修飾がみられたとき、それをどのように解釈すればよいだろうか。もっとも注意すべきなのは、因果効果の方向が逆転しているかどうかである。効果の方向がサブグループ間で逆転しているということは、治療が利益だったり不利益だったりすることを意味す

るからである（これを疫学では質的交互作用という）。治療の有効性を判断する状況では、効果の方向が同じで、定量的な違いに過ぎない方が都合がよい（量的交互作用）。質的交互作用がみられる状況は、表では2パターンがある。

表には、 $L_i = 0$  または  $L_i = 1$  のサブグループだけで治療の効果が生じているパターンが8通りある。このとき、 $L_i = 0$  または  $L_i = 1$  の個人で治療効果が期待できるから、共変量  $L_i$  は効果予測マーカーということができる。それに加えて、 $L_i = 0$  と  $L_i = 1$  で因果効果の方向が逆になっているパターンが2通りある。このような状況では、個人によって治療が利益になることもあれば害をもたらすこともあるから、治療すべきかどうかより慎重に考える必要がある。

表 3-4 効果の修飾のパターン

共変量の水準別の潜在結果変数の値				$L_i = 1$ のサブグループの $Y_i^1 - Y_i^0$ と $L_i = 0$ のサブグループの $Y_i^1 - Y_i^0$ の差	
$L_i = 1$ $Y_i^1$	$L_i = 1$ $Y_i^0$	$L_i = 0$ $Y_i^1$	$L_i = 0$ $Y_i^0$		
1	1	1	1	0	効果なし
1	1	1	0	-1	$L_i = 0$ は効果予測マーカー
1	1	0	1	1	$L_i = 0$ は効果予測マーカー
1	1	0	0	0	効果の修飾なし
1	0	1	1	1	$L_i = 1$ は効果予測マーカー
1	0	1	0	0	効果の修飾なし
1	0	0	1	2	効果の方向が逆転
1	0	0	0	1	$L_i = 1$ は効果予測マーカー
0	1	1	1	-1	$L_i = 1$ は効果予測マーカー
0	1	1	0	-2	効果の方向が逆転
0	1	0	1	0	効果の修飾なし
0	1	0	0	-1	$L_i = 1$ は効果予測マーカー
0	0	1	1	0	効果の修飾なし
0	0	1	0	-1	$L_i = 0$ は効果予測マーカー
0	0	0	1	1	$L_i = 0$ は効果予測マーカー
0	0	0	0	0	効果なし

ランダム化臨床試験では、治療  $A_i$  と共変量  $L_i$  の交互作用項（interaction）を含む一般化線型モデルを当てはめることで、効果の不均一性を調べることができる。

$$g[E(Y_i|A_i, L_i)] = \beta_0 + \beta_1 A_i + \beta_2 L_i + \beta_3 A_i L_i$$

$$= \text{INTERCEPT} + \text{TREATMENT} + \text{COVARIATE} + \text{INTERACTION}$$

ここで  $\beta_3 A_i L_i$  が交互作用項である。ここで注意してほしいのは、第1巻14章で述べたように、一般化線型モデルにおいて交互作用が生じるかどうかは、

リンク関数の選択（たとえばリスク差とリスク比のどちらを効果の指標に選ぶか）に左右される点である。つまり、ある共変量が効果修飾因子かどうかは、当てはめているモデルに依存することがあり得る。

#### ■ 3.4.4 事例：広島・長崎の寿命調査

表 3-5 は、広島・長崎の寿命調査の対象者 80145 人全体における放射線被ばくが乳癌発生に与える効果と、共変量（性別）によって層別した後の効果を示したものである（Brenner, et al. 2018）。リスク差をみると、層別前は 0.005 Gy 以上の被ばくによって乳癌発生リスクが 0.74% 増加しており、女性に限った解析では 1.17% のリスクの増加がみられる。一方で、男性ではリスク差は 0.02% と、女性より放射線被ばくの影響が明らかに小さい。この現象は、男性における乳癌発生率が女性よりずっと低いことに起因することは明らかである。これは性別による効果の修飾の一例である。また、この傾向はリスク差では顕著に認められるが、リスク比とオッズ比ではむしろ男性の方が女性より数字が大きいことにも注意しよう。

表 3-5 広島・長崎の寿命調査における層別前後のリスク差・リスク比・オッズ比

	全体		女性		男性	
	≥ 0.005 Gy	< 0.005 Gy	≥ 0.005 Gy	< 0.005 Gy	≥ 0.005 Gy	< 0.005 Gy
乳癌発生なし	44695	34372	26825	20255	18297	14117
乳癌発生あり	756	322	750	320	6	2
合計	45451	34694	27575	20575	18303	14119
乳癌発生リスク	1.66%	0.93%	2.72%	1.56%	0.03%	0.01%
リスク差	0.74%		1.17%		0.02%	
リスク比	1.79		1.75		2.31	
オッズ比	1.81		1.77		2.31	

このように、共変量による層別前と層別後で、曝露とアウトカムの関連の程度が大きく変化する現象を Simpson（シンプソン）のパラドックスという<sup>\*4)</sup>。

<sup>\*4)</sup> 複数の 2×2 表から得られた効果の指標が併合可能かどうかという問題のこと。第 1 巻 14 章では、Simpson のパラドックスには、一般化線型モデルの 3 つの要素が関係していることを説明した。1 つ目の要素は、層別する共変量が治療およびアウトカムと相関するかどうかである。2 つ目は、治療と共変量の交互作用項があるかどうかである。そして 3 つ目は、効果の指標やリンク関数の特性である。

### 3.5 割付けメカニズム

対象者全体が, どの治療を受けるかを決める規則を, 割付けメカニズム (assignment mechanism) という. 臨床試験で用いられる割付けメカニズムでもっとも単純なのは, 個人ごとに一定の確率でランダムに割付けるものである. このとき, 割付け確率は潜在結果変数  $\{Y_i^0, Y_i^1\}$  と共変量  $L_i$  に依存しない. この単純ランダム化は

$$\Pr(A_1, A_2, \dots, A_N | Y_1^0, Y_1^1, L_1, Y_2^0, Y_2^1, L_2, \dots, Y_N^0, Y_N^1, L_N) = \prod_{i=1}^N \Pr(A_i)$$

と表すことができる. このとき, 期待値の上では, 割付けたグループ間で共変量に偏りは生じない.

Imbens and Rubin (2015) によるテキストは, いくつかの割付けメカニズムを取り上げているが, それを無視可能 (ignorable)・非交絡 (unconfounded) という概念を用いて分類している. 無視可能とは, 割付け確率が観測データだけに依存することと定義され, 数式では

$$\begin{aligned} \Pr(A_1, A_2, \dots, A_N | Y_1^0, Y_1^1, L_1, Y_2^0, Y_2^1, L_2, \dots, Y_N^0, Y_N^1, L_N) \\ = \Pr(A_1, A_2, \dots, A_N | Y_1, L_1, Y_2, L_2, \dots, Y_N, L_N) \end{aligned}$$

のように表される (これは欠測メカニズムにおける無視可能と同義である).

非交絡とは, 割付け確率が共変量だけに依存することを指す. この用語は (Rosenbaum and Rubin 1983 でいう) 強い無視可能 (strongly ignorable) とほぼ同じ意味で用いられている. 強い無視可能割付けメカニズムは, 割付けが個人ごとに独立になされるとき

$$\Pr(A_1, A_2, \dots, A_N | Y_1^0, Y_1^1, L_1, Y_2^0, Y_2^1, L_2, \dots, Y_N^0, Y_N^1, L_N) = \prod_{i=1}^N \Pr(A_i | L_i)$$

と表すことができる.  $e(l) = \Pr(A_i = 1 | L_i = l)$  をプロペンシテイスコア (propensity score) という (Rosenbaum and Rubin 1983).

### 3.6 アウトカム適応的ランダム化

#### ■ 3.6.1 事例: ミシガン ECMO 試験

交絡割付けメカニズムを採用したランダム化臨床試験の例を挙げよう. 交絡

割付けメカニズムの典型例は、アウトカム適応的ランダム化である。これは過去に治療を受けた対象者のアウトカムに基づいて、割付けのアルゴリズムを変更する試験デザインのことである。1982年のミシガン ECMO 試験 (Bartlett, et al. 1985) では、割付け確率を決めるためにアウトカム適応的ランダム化の一種 (ランダム化プレイザウィナールール) が用いられた (Wei and Durham 1978, Wei, et al. 1990)。このルールは白と黒のボールが入っている壺に例えられる。はじめに壺には白と黒のボールが1個ずつ入っている。壺からボールを1つ取り出して、白だったら試験治療群に割付け、黒だったらコントロール群に割付け、ボールを壺に戻す。試験治療群でよい結果が出たか、コントロール群で悪い結果が出た場合には、壺に白いボールを入れる。そうすると、壺には白いボール2個と黒いボール1個が入っているから、白いボールが出て試験治療群に割付けられる確率は1/2から2/3になる。コントロール群でよい結果が出たか、試験治療群で悪い結果が出た場合には、逆に壺に黒いボールを入れる。このように、対象者のアウトカムが評価される度に、壺の中の白と黒のボールの割合が更新され、次の対象者が成績のよい治療を受ける確率を高めるのが、このランダム化プレイザウィナールールである。

ミシガン ECMO 試験において、2番目以降の対象者の割付け確率は

$$\Pr(A_i = 1 | A_1, Y_1, \dots, A_{i-1}, Y_{i-1}) = \frac{1 + \sum_{j=1}^{i-1} [(1 - A_j)Y_j + A_j(1 - Y_j)]}{i + 1}$$

となる (死亡すなわち  $Y_i = 1$  が悪い結果のため)。これは、割付け確率が観測データだけで決まるため、無視可能割付けメカニズムに分類される。しかし、強い無視可能 (非交絡) ではない。割付け確率が共変量だけで決まらず、アウ

表 3-6 ミシガン ECMO 試験のデータ

登録順序 $i$	$\Pr(A_i = 1   Y_j, A_j, j = 1, \dots, i-1)$	$A_i$	$Y_i^1$	$Y_i^0$
1	1/2	1	0	欠測
2	2/3	0	欠測	1
3	3/4	1	0	欠測
4	4/5	1	0	欠測
5	5/6	1	0	欠測
6	6/7	1	0	欠測
7	7/8	1	0	欠測
8	8/9	1	0	欠測
9	9/10	1	0	欠測
10	10/11	1	0	欠測
11	11/12	1	0	欠測
12	12/13	1	0	欠測

トカム  $Y_i$  に依存しているからである。

表 3-6 にミシガン ECMO 試験で実際に用いられた割付け確率を示す。すでに述べたように、この試験では実現した割付けパターン以外でどのような割付け確率が生じるかはわからない。

### ■ 3.6.2 事例：ハーバード ECMO 試験

その後ハーバード大学で行われた試験 (O'Rourke, et al. 1989) では、個人ではなく集団単位で割付け確率を変える 2 ステージデザインが採用された (Ware 1989)。これは、ランダム化を行う第 1 ステージと、それまでの成績のよい治療を強制的に割付ける第 2 ステージから構成される。第 1 ステージでは、ブロックサイズ 4 の置換ブロック法により、4 人ごとに均等に試験治療群とコントロール群を割付ける。そして、一方の治療で 4 件の死亡が観測されたとき、第 1 ステージを終了する。第 2 ステージでは、4 件の死亡が観察されるか優越性を結論できるまで、成績のよい方の治療を用いて試験を継続する。このデザインもまたアウトカム適応的ランダム化の一種である。

ハーバード ECMO 試験のデザインは、頻度論の立場によるものである。主たる解析には、第 1・第 2 ステージを併合した正確検定が採用された。デザインの根拠を得るため、過去の遷延性肺高血圧症のある新生児についてチャートレビューが行われ、従来療法の死亡割合は 13 人中 11 人という事前情報が得られた。これに基づき、死亡確率 20% 対 80% という対立仮説を用いて、片側有意水準 5% の下で検出力 74% が得られるように、デザイン（たとえば死亡 4 件という第 2 ステージの開始基準）が決められた。

このようなアウトカム適応的ランダム化では、割付けメカニズムはアウトカムの値に依存するから、無視可能であるが、強い無視可能には分類されない。この試験の結果の解釈については 5 章で述べる。

### ■ 3.6.3 事例：英国 ECMO 試験

ミシガン ECMO 試験・ハーバード ECMO 試験の結果は ECMO の有効性を示していたが、コントロール群の人数が少なすぎるなどの批判があり、医師間でコンセンサスには至らなかった。その後、英国で行われた大規模ランダム化臨床試験 (UK Collaborative ECMO Trial Group 1996) では、原疾患、重症度、

施設を用いた最小化法<sup>\*5)</sup>が採用された。

### 3.7 因果効果の識別に必要な2つの仮定

因果効果を識別するためには、割付けメカニズムについて2つの仮定が必要になる。第一の仮定は、強い無視可能性（非交絡性）と呼ばれるもので、潜在結果変数 $\{Y_i^0, Y_i^1\}$ と治療 $A_i$ は、共変量 $L_i$ により条件付けた下で、独立であることと定義される。ここで、 $A \perp\!\!\!\perp B|C$ は、「確率変数 $A$ と $B$ は $C$ で条件付けたとき、独立であること」を意味する記号である。

$$\{Y_i^0, Y_i^1\} \perp\!\!\!\perp A_i | L_i$$

第二の仮定は、確率的に割付けが行われるという条件である。

$$0 < \Pr(A_i = 1 | L_i) < 1$$

このことをオーバーラップ（overlap）や正値性（positivity）という。

どうして、非交絡性とオーバーラップの下で、平均因果効果が識別できるのだろうか。その説明は、たとえば以下の関係により与えられる。

$$E(Y_i^a | L_i = l) = E(Y_i | A_i = a, L_i = l)$$

この右辺は、それぞれの群で観測されたデータに回帰モデルを当てはめることで推定できる。回帰モデルの平均構造を関数 $\mu^a(l)$ で特定したとすると

$$E(Y_i | A_i = 0, L_i = l) = \mu^0(l)$$

$$E(Y_i | A_i = 1, L_i = l) = \mu^1(l)$$

と表すことができる。条件付因果効果 $\tau(l)$ は関数 $\mu^a(l)$ によって決まるから、 $\tau(l)$ の $l$ の分布についての平均を考えれば、 $\tau$ が識別できていることがわかる。

$$\tau(l) = \mu^1(L_i) - \mu^0(L_i)$$

これを共変量の経験分布について平均をとることで、母集団平均因果効果を推定できる。

$$\hat{\tau} = \frac{1}{N} \sum_{i=1}^N [\hat{\mu}^1(l_i) - \hat{\mu}^0(l_i)]$$

<sup>\*5)</sup> 最小化法とは、共変量の周辺分布が群間で偏らないように工夫されたランダム化の方法の一種である。



ここで、仮にオーバーラップの仮定が、 $L_i = l$  という共変量の値を持つ集団で満たされなかったら何が起きるだろうか。そのサブグループでは、関数  $\mu^0(l)$  または関数  $\mu^1(l)$  のどちらかを推定できない。つまり平均因果効果が識別できない、ということになってしまう。

非交絡性・オーバーラップよりも、弱い条件が用いられることもある。それはたとえば、治療群における有限集団因果リスク差や治療群における母集団因果リスク差を推定するときである。このための必要条件は、コントロールにおける非交絡性

$$Y_i^0 \perp\!\!\!\perp A_i | L_i$$

と弱オーバーラップ

$$\Pr(A_i = 1 | L_i) < 1$$

である。

## 3.8 統計的推測

### ■ 3.8.1 母集団因果リスク差のモデルに基づく頻度論の推測

本書では統計的推測について、主に最尤法と M 推定を扱っている。しかし統計学全体でいえば、頻度論と Bayesian の 2 つの流派がある。また、推測の対象となるパラメータも、有限集団と無限母集団では意味が違ってくる。ランダム化臨床試験の解析は、ランダム化に基づく推測と尤度（またはモデル）に基づく推測に大別される。このように統計的推測にはバリエーションがあるにもかかわらず、 $2 \times 2$  表のような単純な場合には、似たような推定量に帰着することが多い。

このことは潜在結果変数モデルを考えるとときにもいえる。有限集団因果リスク差と母集団因果リスク差のどちらも、リスク差  $\tau$  によって推定することができる。ただし、 $\tau$  の誤差分布は、無限母集団と有限集団で異なることを指摘しておく。これは、前者はモデル（ランダムサンプリング）に基づいて推測を行ったときの誤差を表しているのに対して、後者はランダム化に基づくバラツキを反映しているためである。有限集団を考えたとき、特定の状況以外では、分散

$$\text{Var}(\hat{\tau}) = \frac{\hat{\pi}^1(1 - \hat{\pi}^1)}{N_1} + \frac{\hat{\pi}^0(1 - \hat{\pi}^0)}{N_0}$$

には大標本であっても過大評価の方向にバイアスがある．この点は4章で詳しく述べる．

ここでは，母集団因果リスク差の推定において，潜在結果変数にパラメトリックモデルを仮定すると， $2 \times 2$  表の積2項分布モデルに帰着することを示そう． $N$  人の対象者は，無限母集団からのランダムサンプルと仮定する．そして， $Y_i^0$  と  $Y_i^1$  の母集団における確率分布に，以下のような独立な Bernoulli 分布を仮定する．

$$\Pr(Y_i^0 = y, Y_i^1 = v; \pi^0, \pi^1) = \binom{1}{y} (\pi^0)^y (1 - \pi^0)^{1-y} \binom{1}{v} (\pi^1)^v (1 - \pi^1)^{1-v}$$

$Y_i^0$  と  $Y_i^1$  は独立と仮定するのは， $Y_i^0$  と  $Y_i^1$  の相関はデータから識別不能だからである．このとき尤度関数は単純ランダム化の下で

$$\begin{aligned} L(\pi^0, \pi^1) &= \prod_{i=1}^N \Pr(Y_i | A_i; \pi^0, \pi^1) \Pr(A_i) \\ &\propto \prod_{i:A_i=0} \Pr(Y_i | A_i = 0; \pi^0) \prod_{i:A_i=1} \Pr(Y_i | A_i = 1; \pi^1) \\ &= \prod_{i:A_i=0} \Pr(Y_i^0; \pi^0) \prod_{i:A_i=1} \Pr(Y_i^1; \pi^1) \\ &\propto (\pi^0)^{S_0} (1 - \pi^0)^{N_0 - S_0} (\pi^1)^{S_1} (1 - \pi^1)^{N_1 - S_1} \end{aligned}$$

ここで第2式から第3式は，潜在結果変数と治療の独立性を利用して導かれる\*6)．これは積2項分布モデルと同じ尤度関数であるから，最尤法を適用すれば，リスク差  $\tau = \pi^1 - \pi^0$  についての推測を行うことができる．

母集団因果リスク差に関する帰無仮説は

$$H_0: \tau = 0$$

と表される．この命題が正しいかどうかは，

$$\chi^2 = \frac{\widehat{\tau}^2}{\widehat{\pi}^1 (1 - \widehat{\pi}^1) / N_1 + \widehat{\pi}^0 (1 - \widehat{\pi}^0) / N_0}$$

\*6) ここでは  $Y_i^0$  と  $Y_i^1$  の独立性を仮定して尤度を構成したが，これは自明ではないから，補足が必要であろう．データは  $Y_i^0$  と  $Y_i^1$  の相関に関する情報を持たないため，尤度関数は真の分布によらず上の式になる．両者の相関関係について，単調性  $Y_i^0 < Y_i^1$  という仮定をおくことで， $Y_i^0$  と  $Y_i^1$  の両方に基づく層別（11章で扱う主要層別）を行うことがある（Frangakis and Rubin 2002）．また，Imbens and Rubin（2015）の8.6節には，Bayes流の推測において，事前分布を通じて相関に関する情報を与えると，事後分布の分散が大きくなる数値例が示されている．

が、積 2 項分布モデルにおける帰無仮説の下で、漸近的に自由度 1 の  $\chi^2$  分布に従うことを利用して、検定することができる。この統計量は、分割表の解析における、いわゆる独立性の  $\chi^2$  検定と同じものである。

### ■ 3.8.2 母集団因果リスク差の Bayes 流の推測

$Y_i^0$  と  $Y_i^1$  にパラメトリックモデルを仮定して導かれた  $\pi^0$  と  $\pi^1$  の尤度関数は、Bayes（ベイズ）推測でも用いられる。本書の立場は頻度論だが参考のため触れておこう。単純ランダム化の下で、 $\pi^0$  と  $\pi^1$  について無情報事前分布を用いるとき、上の 95%信頼区間の公式は、 $\tau$  の事後分布の 95%信用区間（credible interval）と一致する。このときの 95%という確率は、頻度確率（たとえば割付けを繰り返したときの実現頻度の相対値）とは異なり、主観確率として解釈される。

Bayes 推測の特徴は、潜在結果変数のうち観測できない部分  $Y_i^{\text{mis}} = A_i Y_i^0 + (1 - A_i) Y_i^1$  を欠測データとみなし、その事後予測分布を経由して、因果効果の事後分布を求める点である（Rubin 1978）。Imbens and Rubin（2015）に従って、無視可能割付けメカニズムの下で  $Y_i^{\text{mis}}$  の事後予測分布を導いてみよう。割付けメカニズムを含む確率分布を考えるため、 $Y^0$  と  $Y^1$  に加え  $A$  の同時分布を考える。 $Y_i^{\text{mis}}$  の事後予測分布は、 $\Pr(Y^0, Y^1, A|L)$  のうち、観測できる部分で条件付けたものである。したがって、無視可能性の仮定の下で以下の結果を得ることができる。

$$\begin{aligned} \Pr(Y^{\text{mis}}|Y, A, L) &= \frac{\Pr(A|Y^0, Y^1, L) \Pr(Y^0, Y^1|L)}{E^{Y^{\text{mis}}} \{\Pr(A|Y^0, Y^1, L) \Pr(Y^0, Y^1|L)\}} \\ &= \frac{\Pr(A|Y, L) \Pr(Y^0, Y^1|L)}{E^{Y^{\text{mis}}} \{\Pr(A|Y, L) \Pr(Y^0, Y^1|L)\}} \\ &= \frac{\Pr(Y^0, Y^1|L)}{E^{Y^{\text{mis}}} \{\Pr(Y^0, Y^1|L)\}} \end{aligned}$$

ここで、 $E^{Y^{\text{mis}}}$  は、 $Y^{\text{mis}}$  についての期待値を表す。第 1 式から第 2 式は無視可能性から、第 2 式から第 3 式は  $\Pr(A|Y, L)$  に  $Y^{\text{mis}}$  が含まれないことから導かれる。これらの式にはパラメータ  $\pi$  が明示的に含まれていないが、 $\Pr(Y^0, Y^1|L, \pi)$  を  $\pi$  の事前分布  $f(\pi|L)$  にそって積分することで計算される。すなわち、 $\Pr(Y^0, Y^1|L, \pi)$  と  $f(\pi|L)$  を特定し、観測データから事後予測分布を求めれば、因果効果に欠測データ  $Y^{\text{mis}}$  が含まれていたとしても、その確率分布を計算することができる。

第3式の分母に含まれている  $Y^0$  と  $Y^1$  の条件付同時分布は、一般に

$$\Pr(Y^0, Y^1 | L) = \int \prod_{i=1}^N \Pr(Y_i^0 | L_i, \pi) \prod_{i=1}^N \Pr(Y_i^1 | L_i, \pi) f(\pi | L) d\pi$$

と書ける。これに独立な Bernoulli 分布を仮定し、 $f(\pi | L)$  に無情報事前分布を仮定すれば、やはり  $2 \times 2$  表の積2項分布モデルの尤度に帰着する。

この議論で重要なのは割付けメカニズムが無視可能かどうかである (Rubin 1978)。一般に、因果効果の事後分布は、割付けメカニズムが無視可能なとき、観測データ  $\{Y, A, L\}$ ,  $Y^0$  と  $Y^1$  の条件付同時分布  $\Pr(Y^0, Y^1 | L, \pi)$  とすべてのパラメータ  $\pi$  の事前分布  $f(\pi | L)$  だけで決まる。それに対して、割付けメカニズムが無視不可能だと、因果効果の事後分布に割付けメカニズムの部分が含まれるため、具体的な確率分布を特定する必要が生じる。そしてそのとき事後分布は割付けメカニズムの誤特定に鋭敏であり、誤った推測を導きやすい。

### ■ 3.8.3 ランダム化プレイザウィナールールの下での推測

ここからはやや高度な話題になるが、上の議論はランダムサンプリング・単純ランダム化を仮定している。それに対してアウトカム適応的ランダム化の下では、尤度関数が積2項分布モデルと同一にならない。このとき妥当な推測を行うにはどうすればよいか。

第一に考えられるアプローチは、最尤法である。Wei, et al. (1990) は、ランダム化プレイザウィナールールの下で、リスク差  $\tau$  がどのような挙動を示すか調べた。治療割付けの列  $A$  はランダム化プレイザウィナールールに従って決まるとする。さらに、 $\Pr(Y_i | A_i)$  に2項分布を仮定する。このとき  $\tau$  は、やはり大標本では正規分布に従い、分散は  $\text{Var}(\tau) = \pi^1 (1 - \pi^1) / N_1 + \pi^0 (1 - \pi^0) / N_0$  で与えられる。ただしこれはサンプルサイズがかなり大きいときの結果である。サンプルサイズが50よりも小さいときは、通常の95% Wald 信頼区間を用いるべきではない。Wei, et al. (1990) には、十分統計量  $S$  と  $N_1$  で条件付けた下での正確な信頼区間と、無条件の正確な信頼区間が示されており、こちらが推奨されている。

別の有力な選択肢として、独立性の  $\chi^2$  検定の代わりに、ランダム化に基づく並び替え検定を用いるアプローチが考えられる。一般に並び替え検定は、特定の確率分布に依存しないノンパラメトリックな手法のため、頑健性が高いとされる。しかしランダム化プレイザウィナールールでは、表3-5でみたように、

実現した割付けパターン以外で、どのような割付け確率が生じるかはわからない。これは（4章で述べる Fisher 帰無仮説の下でなければ）、並び替え検定の確率計算ができないことを意味する。一般的にいて、強い無視可能割付けメカニズムでなければ、並び替え検定を構成することは難しい。

第三のアプローチは前述の Bayes 流の推測である。ランダム化プレイザウイナールールの下ではこのアプローチも有力とされている。

#### 因果モデルとはなにか 2

因果と一括りに論じてはいるが、原因と結果の具体的な対象は分野によって異なるわけで、因果モデルに多様性が生じることはむしろ自然である。疫学発祥の因果モデルとして Rubin 因果モデルと十分原因構成要素モデルが挙げられるが、計量経済学に目を移すと、構造方程式モデルや Granger（グレンジャー）因果というまったく別のモデルがある。因果モデルのオーバービューは、Pearl (2009), Greenland and Brumback (2002), VanderWeele (2015) に与えられている。これらの抽象的な枠組みは、因果の概念を積極的に維持しているといえ、逆説的にいえばそこになんらかの必然性があるはずである。これを論じるには、それぞれの分野で議論の対象となっている因果関係が、どのような特徴を持つか考えなければならない。

放射線疫学の典型例として、広島・長崎の寿命調査がある。この研究は、放射線被ばくとがんなどの疾患の因果関係を明らかにすることが目的である。このとき、人によっては、放射線被ばくとがんの間に因果関係があることは科学的事実であって、議論の前提だと考えるかもしれない。高線量の放射線被ばくによって、DNA 損傷や染色体異常が生じ、細胞レベルの異常や発癌につながることは科学的にわかっているからである。しかしこの意見には明確な反論がある。特定の個人に生じたがんの原因を考えると、どのくらいの線量なら因果関係があるといえるのだろうか？ 放射線以外のがんの原因があり得るのではないだろうか？ つまり物理学・生物学と疫学では、同じ現象を扱っていたとしても、因果の意味するところが違うのである。実はこの場合の因果には少なくとも 3 つの含意がある。

- 特定の個人に生じたがんの原因は、放射線被ばくによるものか（単称因果）
- 放射線被ばくによってがんが発生するという現象を、科学的に説明できるか（メカニズムの解明）
- 寿命調査データから、放射線被ばくとがんの間に因果関係があると判断できるか（統計的推測）

原因と結果の関係について統計学的に推測するためには、データを確率分布で表現することになるが、それは2つに大別することができる。1つ目は、観察されるデータの確率分布であり、その具体例が一般化線型モデルといえる。もう1つの定式化がRubin因果モデルであり、このモデルでは潜在結果変数(potential outcomes)を含めた確率分布を考える(Neyman 1923, Rubin 1974)。仮に対象者  $i$  が放射線に曝露したときの潜在結果変数を  $Y_i^1$ 、仮に曝露しなかったときの潜在結果変数を  $Y_i^0$  で表す。因果効果の有無は、両者に差があるかどうか、つまり  $\tau_i = Y_i^1 - Y_i^0$  によって定義することができる。

もうちょっと議論を進めると、潜在結果変数を導入する必然性がわかってくる。一般化線型モデルの発想は、真の確率分布にもっとも近いモデルを当てはめることにある。ところが、以下の2つのモデルが同時に真実だったとしたらどうだろうか。

$$g[E(Y_i|A_i, L_i)] = \beta_0 + \beta_1 A_i + \beta_2 L_i$$

$$g[E(Y_i|A_i)] = \beta_0^* + \beta_1^* A_i$$

前者は、治療  $A_i$ 、共変量  $L_i$  で条件付けたアウトカム  $Y_i$  の期待値である。一方で、後者は、治療  $A_i$  だけで条件付けたアウトカム  $Y_i$  の期待値である。治療  $A_i$  の回帰係数  $\beta_0$  と  $\beta_0^*$  は、明らかに治療  $A_i$  のある種の効果を表しているが、どちらもパラメータ真値といえる。もしパラメータの併合可能性がかならずしも成り立たないとしたら（つまり  $\beta_1 \neq \beta_0^*$  だとしたら）どうだろう。因果効果が特定の値に定まらないことになって具合が悪い。そのため、因果推論の推定目標を潜在結果変数によって定義することから、本書はスタートしたのである。

### 演習問題

#### 〈効果の修飾〉

〔問1〕 以下の式は、効果の修飾に関する制約のパターンを表している。もっとも制約が強いものをひとつ選べ。

(A)  $\tau = E(Y_i^1 - Y_i^0)$

(B)  $E(Y_i^1 - Y_i^0 | L_i = l) = E(Y_i^1 - Y_i^0 | L_i = m) \quad \text{for all } l \neq m$

(C)  $Y_i^1 - Y_i^0 = Y_j^1 - Y_j^0 \quad \text{for all } i \neq j$

## 〈割付けメカニズム〉

問2 臨床試験では, さまざまなランダム化の方法が用いられ, 因果効果に関する統計的推測はその影響を受ける. 通常の統計手法で因果効果を推定すると問題が生じるのは, 次のうちどの方法を用いたときだろうか.

- (A) 単純ランダム化 (割付け確率は定数)
- (B) 共変量に基づくブロックランダム化 (割付け確率は共変量に依存)
- (C) アウトカム適応的ランダム化 (割付け確率はアウトカムに依存)
- (D) 最小化法 (割付け確率はこれまで登録された対象者の割付け結果に依存)

問3 治療をランダムに割付けたはずなのに, ベースライン共変量の一部に治療群間で有意な差がみられることがある. この現象に当てはまるのは次のうちどれか.

- (A)  $\alpha$  エラー
- (B)  $\beta$  エラー
- (C) Simpson のパラドックス
- (D) 交絡