

· 基础理论与方法 ·

交互作用的统计推断

耿直 胡永华

假若因果推断关心的是仅仅一个风险因素对响应的因果作用的话,那么通过随机化试验或者通过控制混杂因素来消除混杂偏倚,就可以得到平均因果作用^[1]。如果因果推断包含两个或多个风险因素的话,除了每个因素的平均因果作用之外,还存在着多个风险因素之间相互的交互作用。如果研究因果意义上的交互作用,那么,同样需要进行随机化试验,或者控制混杂因素消除混杂偏倚。交互作用分析可以看作是多个风险因素的因果推断。

在流行病学研究中,“交互作用”是关于多风险因素分析的常用术语,但是,它被用于表示各种不同的概念。Rothman, Greenland^[2]描述了 3 种交互作用:统计交互作用,生物交互作用和公共卫生交互作用。各种交互作用的概念可以大致分为两大类:一类是关于多风险因素的统计模型和参数的定量概念,称为统计交互作用;另一类是关于多风险因素在生物机理上的定性概念,称为生物交互作用。统计交互是指暴露单个风险因素作用的总和或乘积是否等于同时暴露两个或多个风险因素的作用。生物交互是指对单独一个个体,当个体暴露每个单个风险因素都不会患病,仅当同时暴露两个或多个风险因素时才患病。在实际研究中,人们常常将两者混淆,在分析中采用统计交互方法,而将结果解释为生物交互现象。

在流行病学研究中普遍采用加法交互模型和乘法交互模型。但是,这些交互模型都是研究统计意义上的交互作用,而不是生物意义上的交互作用概念。另外,统计交互作用本身的概念也非常含糊,当使用统计意义上的“交互作用”一词时,必须同时说明所使用的尺度和所使用的关联测度,否则将无法评价其交互作用。

Miettinen^[3]讨论了生物交互作用的概念,Darroch, Borkent^[4]和 Darroch^[5]探讨了生物交互作用的最大熵估计方法。Rothman, Greenland^[2]定义了各种生物交互作用。关于生物交互作用的分析比因果推断更复杂。进行交互作用分析时,除了消除混杂偏倚之外,还需要其他的假定,这些假定不能根据观测得到的数据进行假设检验,需要试验和观察以外的专家知识。

本文主要讨论关于统计交互的概念,生物交互的概念,解释流行病学研究中常用的模型不能有效地确定是否存在

生物交互作用,并提出生物交互作用的分析方法。

一、统计交互作用

一些常见的统计交互作用的分析方法通过对无混杂偏倚的单因素因果作用和多因素因果作用进行比较,或建立模型来判断交互作用是否存在。首先介绍采用加法交互模型和乘法交互模型来确定是否存在交互作用。设 A 和 B 分别为两个二值的风险因素,取值 0 和 1,分别表示未暴露与暴露;设 Y 表示二值响应变量,取值 0 和 1,分别表示未患病和患病,令 $P_{ij} = P(Y=1|A=i, B=j)$ 表示暴露状态为 $A=i$ 且 $B=j$ 的子总体中的患病率(表 1)。

表1 暴露两个风险因素的患病率

	$B=0$	$B=1$
$A=0$	P_{00}	P_{01}
$A=1$	P_{10}	P_{11}

加法交互作用和乘法交互作用是根据下面 3 个等式是否成立来定义的:

$$P_{11} - P_{00} = (P_{10} - P_{00}) + (P_{01} - P_{00}) \quad (1)$$

$$\frac{P_{11}}{P_{00}} = \frac{P_{10}}{P_{00}} \cdot \frac{P_{01}}{P_{00}} \quad (2)$$

$$\frac{(1 - P_{11})}{(1 - P_{00})} = \frac{(1 - P_{10})}{(1 - P_{00})} \cdot \frac{(1 - P_{01})}{(1 - P_{00})} \quad (3)$$

如果等式(1)成立,称无加法交互作用,如果等式(2)成立,称无乘法交互作用,有时也根据等式(3)成立来确定无乘法交互作用,乘法交互模型可以看作对数尺度上的加法交互模型。当 A 和 B 都有作用时,即 $P_{10} \neq P_{00}$ 且 $P_{01} \neq P_{00}$ 时,等式(1)和等式(2)不可能同时成立,也就是说:无加法交互作用的话,一定有乘法交互作用,而无乘法交互作用的话,一定有加法交互作用。等式(2)还可以改写成:

$$\frac{P_{11}}{P_{01}} = \frac{P_{10}}{P_{00}}$$

该式可以解释为: $B=1$ 层上 A 对 Y 的相对风险等于 $B=0$ 层上 A 对 Y 的相对风险,也称为相对风险是均一的。有时加法交互模型和乘法交互模型采用其他关联测度(如:优势比)那么,用不同关联测度定义的交互作用的概念也不相同。Rothman, Greenland^[2]强调在使用“交互作用”这个术语时,应该同时指出所使用的尺度和关联测度。

下面介绍根据统计模型中的交互参数进行统计推断来确定是否存在交互作用的方法。例如:令 X_1 和 X_2 是两个

基金项目:国家自然科学基金资助项目(39930160),国家杰出青年基金资助项目(19825103)

作者单位:100081 北京大学数学科学学院概率统计系(耿直);北京大学公共卫生学院流行病与卫生统计学系(胡永华)

风险因素, Y 是响应变量, 建立回归模型:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \quad (4)$$

当参数 $\beta_3 \neq 0$ 时, 表示存在 X_1 和 X_2 对 Y 的交互作用, β_3 描述模型意义上 X_1 和 X_2 的乘积对预测 Y 的作用。

最后, 简单介绍通过假设检验来确定是否存在交互作用的方法, 对各个单个风险因素的暴露作用进行假设检验都不显著, 而对两个风险因素的同时作用进行检验显著, 那么, 确定有交互作用存在。这种假设检验方法相当于加法交互模型。

二、生物交互作用

我们强调生物交互作用的重要性, 指出它与统计交互作用的差别, 通常的统计交互作用的分析方法不能简单地用于推断生物交互作用。首先介绍生物交互作用的概念, 然后, 解释统计交互作用的分析方法不能简单得出生物交互作用的结论, 最后, 介绍生物交互作用的统计推断方法。

1. 生物交互作用的概念: 令 A 和 B 是两个二值暴露因素, 取值 0 和 1, 分别表示未暴露与暴露。令 Y_{ij} 是一个虚拟的二值响应变量, 表示个体在暴露状态为 $A=i$ 且 $B=j$ 时的响应, 取值为 0 和 1, 分别表示未患病和患病。注意: 对一个个体通常 $Y_{00}, Y_{01}, Y_{10}, Y_{11}$ 中只有一个是可观察到的。当暴露状态为 $A=0$ 且 $B=0$ 时, 仅 Y_{00} 可以观察到, 其他是观察不到的。例如: 当一个个体, 既不抽烟也不喝酒时, 仅能观察到该个体在不抽烟且不喝酒状态下的是否患病, 而不可能观察到该个体在抽烟且喝酒状态下是否患病的响应。所以, 我们称 Y_{ij} 为虚拟的响应。令 $\pi_{ij} = P(Y_{ij} = 1)$ 表示总体中所有个体都暴露 $A=i$ 且 $B=j$ 状态下, 患病个体的比率。对总体中每个个体, 其可能的暴露状态有 4 种, 在每种状态下分别有相应的响应结果: $Y_{00}, Y_{01}, Y_{10}, Y_{11}$ 。这里实际上假定了一个体的响应由个体本身及暴露状态确定。进一步假定各暴露组 ($A=i, B=j$) 的响应 Y 的分布等于虚拟响应 Y_{ij} 的分布, 即 $\pi_{ij} = P_{ij}$ 。最后一个假定相当于随机化试验, 或者是控制了充分的混杂因素, 使得不同暴露组是可交换的。例如, 控制了足够的混杂因素后, 假若整个总体中每个个体都不吸烟且不喝酒的话, 患病的比率与非吸烟且非喝酒组的患病比率一致。可以将总体中所有个体分为 $2^4 = 16$ 种类型的子总体 (表 2)。

我们对表 2 中某些子总体做一些说明。例如 $(Y_{00}, Y_{10}, Y_{01}, Y_{11}) = (0, 0, 1, 0)$ 表示: 只有当该个体单独暴露 B 时才患病, 单独暴露 A , 或同时暴露 A 和 B 时不患病。对这样的人群, 我们称 A 和 B 之间存在对抗作用, 更确切地说, A 因素对 B 因素有克制作用。 $(0, 0, 0, 1)$ 表示: 只有当个体同时暴露 A 和 B 时才患病, 单独暴露 A 或者单独暴露 B 时都不患病。对这样的人群, 我们称 A 和 B 两因素之间存在生物意义上的协同作用, 也称 A 和 B 对响应有交互作用。从表 2 中可以看到多种多样的交互作用, 我们统称这样的交互作用为生物意义上的交互作用。我们用每个子总体在整个总体中所占的比率来度量各种交互作用的大小。例如, 属于

子总体 $(0, 0, 0, 1)$ 的个体的数量在整个总体所占的比率为 A 和 B 对响应的生物协同作用。数据分析的目的是对这

表 2 根据虚拟响应将总体分为 16 类子总体

类型	Y_{00}	Y_{10}	Y_{01}	Y_{11}	子总体性质
1	0	0	0	0	A 和 B 都无作用的个体 (总不患病)
2	0	0	0	1	A 和 B 共同作用导致患病 (协同致病作用)
3	0	0	1	0	A 克制 B 的致病作用
4	0	0	1	1	A 无作用, B 为致病因素
5	0	1	0	0	B 克制 A 的致病作用
6	0	1	0	1	A 为致病因素, B 无作用
7	0	1	1	0	A 和 B 皆为致病因素, 但相互克制致病作用
8	0	1	1	1	A 和 B 皆为致病因素
9	1	0	0	0	A 和 B 皆为预防因素
10	1	0	0	1	A 和 B 单独地为预防因素, 但存有相互克制作用
11	1	0	1	0	A 为预防因素, B 无作用
12	1	0	1	1	A 为预防因素, B 克制 A 的作用
13	1	1	0	0	B 为预防因素, A 无作用
14	1	1	0	1	B 为预防因素, A 克制 B 的作用
15	1	1	1	0	A 和 B 共同作用预防患病 (协同预防作用)
16	1	1	1	1	A 和 B 都无作用 (总是患病)

些生物交互作用进行统计推断。比如, 估计 A 和 B 的生物协同作用的大小。根据通常的观察研究所获得的数据, 对表 2 中给出的各类子总体所占比率进行统计推断是一个非常困难的问题, 在一些特殊的情况下, 才可以简单地进行统计推断。第一种情况, 假若总体的每个个体可以分别暴露所有 4 种状态, 并假定暴露的响应结果不受暴露状态的先后顺序影响, 那么, 通过样本中每个个体分别暴露 4 种状态的结果, 可以容易地对表 2 中给出的各类子总体所占比率进行统计推断。第二种情况, 假若总体是由具有相同虚拟响应的 4 人组组成, 比如说同胞同卵四兄弟组。分配 4 人组中每一个人暴露不同的状态, 可以得到所有 4 种虚拟的响应结果, 那么, 很容易进行生物交互作用的统计推断。Rothman, Greenland^[2] 将 1 类, 4 类, 6 类, 11 类, 13 类和 16 类, 这 6 类归为无任何生物交互作用的子总体, 而将其余 10 类归为存在某种生物交互作用的子总体。他们得出了下面的结论: 当等式 (2) 不成立时, 一定存在着某种生物交互作用; 但是, 存在某种生物交互作用时, 可能等式 (2) 仍成立; 当下面不等式成立时:

$$P_{11} - P_{00} > (P_{10} - P_{00}) + (P_{01} - P_{00})$$

第 2 类子总体一定非空, 也就是说一定存在协同致病作用; 当下面不等式成立时:

$$P_{11} - P_{00} < (P_{10} - P_{00}) + (P_{01} - P_{00})$$

第 15 类子总体一定非空, 也就是说一定存在协同预防作用。

下面介绍 A 和 B 都是危险因素情况下的生物交互作用的统计推断方法。所谓危险因素意味着: 假若个体不暴露该因素就患病的话, 暴露该因素一定也患病; 不存在某个个体, 暴露时不患病, 没暴露时反而患病。用虚拟响应变量表示

为: $Y_{00} \leq Y_{10} \leq Y_{11}$, 并且 $Y_{00} \leq Y_{01} \leq Y_{11}$ 。在某些流行病学研究中, 可以假定两个因素都是危险因素。这时表 2 中一些类的子总体就不可能出现, 6 类子总体化减为表 3 中的 6 类子总体。表 3 中的 $\mu = E(Y_{01} Y_{10}) = P(Y_{01} Y_{10} = 1) = P(Y_{01} = 1, Y_{10} = 1)$ 表示总体中 Y_{01} 和 Y_{10} 都为 1 的个体在总体中所占的比率。假若没有混杂偏倚, 比如, 随机化试验或控制了充分的混杂因素的情况下, 我们有 $\pi_{ij} = P_{ij}$, 即总体中所有个体都暴露 $A = i$ 且 $B = j$ 时患病个体的概率 $P(Y_{ij} = 1)$ 等于暴露组 $A = i$ 且 $B = j$ 中患病个体的概率 $P(Y_{ij} = 1 | A = i, B = j)$ 这时 π_{ij} 的估计可以用 P_{ij} 的估计来代替。但是, μ 是不可估计的, 即使在随机化试验中, 我们只能观测到同一个体在一种暴露状态下的响应结果, 而得不到 μ 的估计。因此, 表 3 的概率公式中包含 μ 的子总体在总体中所占比率也是不可估计的, 即 $P(A), P(B), P(S), P(P)$ 都是不可估计的, 这里 $P(A)$ 表示 A 类子总体在总体中所占的比率, 其他类似定义。S 类子总体由 A 和 B 对响应的有协同作用的个体组成, 与 Rothman Greenland^[2] 不同, 我们这里不将 P 类子总体归为具有交互作用的子总体, 因为 P 类表示 A 和 B 都是致病因素, 既没有相互协同作用, 也没有相互对抗作用。生物交互作用分析的主要目标是估计协同作用子总体在总体中所占的比率, 即估计 $P(S)$ 。例如, 研究基因与环境的交互作用, 其目的是估计仅当具有某个基因, 同时暴露某种危险因素才患病的个体在总体中占多少比率。在肝癌病因研究中, 肝炎与黄曲霉素是否具有导致肝癌的交互作用? 其交互作用的大小是多少? 从致病机理上, 应该研究是否存在某些个体, 他们仅当患过肝炎, 且又暴露过黄曲霉素时, 才发展为肝癌, 而他们仅患肝炎, 或仅暴露黄曲霉素都不会导致肝癌。这类个体在总体中所占的比率作为交互作用的度量, 如果这个生物交互作用非常大的话, 意味着对肝炎患者避免暴露黄曲霉素是一种重要的预防肝癌的手段。如果通过某种方法可以确定某一个肝炎患者是否属于协同作用那一类子总体的话, 就可以有效地预防肝癌。

表3 两个危险因素可能形成的 6 类子总体

子总体 ($Y_{00}, Y_{10}, Y_{01}, Y_{11}$)	示性函数	概率	
Φ	(0 0 0 0)	$1 - Y_{11}$	$1 - \pi_{11}$
A	(0 1 0 1)	$Y_{10} - Y_{01} Y_{10}$	$\pi_{10} - \mu$
B	(0 0 1 1)	$Y_{01} - Y_{01} Y_{10}$	$\pi_{01} - \mu$
S	(0 0 0 1)	$Y_{11} - Y_{10} - Y_{01} + Y_{01} Y_{10}$	$\pi_{11} - \pi_{10} - \pi_{01} + \mu$
P	(0 1 1 1)	$Y_{01} Y_{10} - Y_{00}$	$\mu - \pi_{00}$
Ω	(1 1 1 1)	Y_{00}	π_{00}

2. 生物交互和统计交互的推断差异: 在流行病学研究中常采用加法交互模型确定交互作用是否存在。我们举例说明这种模型意义上的交互与生物意义上的交互的差别。假定两种危险因素 A 和 B 的作用为: 如果所有的人两种因素都不暴露的话, 则都不患病; 总体中 50% 的人只要暴露于 A 和 B 中一种因素则患病, 即 $P(P) = 0.5$; 另外 50% 的人必须同时暴露于 A 和 B 两种因素才患病, 即 $P(S) = 0.5$; 其他 4

类子总体为空。这表明 A 和 B 两种因素对 50% 的人具有生物意义上的协同交互作用。假设对两种危险因素进行随机化试验, 当人数足够多时, 可近似得到 $A = 0$ 且 $B = 0$ 的暴露组的患病率为 0, $A = 1$ 且 $B = 0$ 的暴露组的患病率为 50%, $A = 0$ 且 $B = 1$ 的暴露组的患病率也为 50%, $A = 1$ 且 $B = 1$ 的暴露组的患病率为 100%。根据加法模型可验证 $P_{11} - P_{00} = (P_{10} - P_{00}) + (P_{01} - P_{00})$, 因此, 不存在加法交互作用。由此可见, 无加法交互作用不能确定无生物交互作用。

我们再给出另一个例子, 说明不存在生物意义上的交互作用时, 可能会表现出统计意义上的加法交互作用。假定总体中所有的人两种因素都不暴露的话, 则都不患病; 总体中, 40% 的人暴露任何因素都不会患病, 20% 的人只要暴露因素 A 就患病, 即 $P(A) = 0.2$; 另外 20% 的人只要暴露因素 B 就患病, 即 $P(B) = 0.2$, 剩下 20% 的人暴露因素 A 和 B 中任一种因素则患病, 即 $P(P) = 0.2$ 。这时没有生物意义上的协同交互作用, 即 $P(S) = 0$, 也就是说不存在必须同时暴露两种因素才患病的人, 也不存在暴露一种因素则患病, 但同时暴露另一因素则不患病的人。假若能采用 4 组随机化试验, 那么, 将得到 $A = 0$ 且 $B = 0$ 的暴露组的患病率为 0, $A = 1$ 且 $B = 0$ 的暴露组的患病率为 40%, $A = 0$ 且 $B = 1$ 的暴露组的患病率也为 40%, $A = 1$ 且 $B = 1$ 的暴露组的患病率为 60%。根据加法模型: $P_{11} - P_{00} = 0.6 < (P_{10} - P_{00}) + (P_{01} - P_{00}) = 0.8$, 因此, 存在统计意义上的对抗交互作用, 但是, 两种因素 A 和 B 并不具有生物意义上的交互作用。

Miettinen^[3] 证明了下面不等式成立时:

$$P_{11} - P_{00} > (P_{10} - P_{00}) + (P_{01} - P_{00}) \quad (5)$$

即存在加法交互作用时, 一定存在生物协同交互作用, 即 $P(S) > 0$; 但是, 存在生物协同交互作用时, 不等式 (5) 不一定成立, 同样可以证明, 当下面不等式成立时:

$$P_{11} - P_{00} < (P_{10} - P_{00}) + (P_{01} - P_{00})$$

一定有 $P(P) > 0$ 。

类似地可以举例说明乘法交互模型也不能确定生物意义上的交互作用。与加法交互模型不同, 即使存在乘法交互作用, 即下面不等式成立时:

$$\frac{P_{11}}{P_{00}} > \frac{P_{10}}{P_{00}} \cdot \frac{P_{01}}{P_{00}}$$

也不能判断一定存在生物协同交互作用。线性模型中的交互参数实际上表示加法交互作用, 而对数线性模型中的交互参数实际上表示乘法交互作用, 可以举例说明统计模型中的交互参数也不能确定生物意义上的交互作用。

现在, 我们讨论假设检验的方法不能用于确定生物协同交互作用。假定对两种因素 A 和 B 进行随机化试验, 将对象随机分为 4 个大小相等的组: $A = 0$ 且 $B = 0$ 组, $A = 1$ 且 $B = 0$ 组, $A = 0$ 且 $B = 1$ 组, $A = 1$ 且 $B = 1$ 组。我们考虑两种假设检验的方法: 第一种按某个因素暴露与否分组, 如果

暴露 A 和不暴露 A 的患病率相等,即 $P(Y=1|A=1)=P(Y=1|A=0)$;且暴露 B 和不暴露 B 的患病率也相等,即 $P(Y=1|B=1)=P(Y=1|B=0)$;可以证明这时 A 和 B 都不暴露的患病率一定等于 A 和 B 都暴露的患病率,即 $P(Y=1|A=0, B=0)=P(Y=1|A=1, B=1)$ 。因此,这种假设检验方法不能用于确定协同交互作用。第二种按单独一个因素暴露与否分组,在不暴露 B 情况下单独暴露 A 和不暴露 A 的患病率相等,即 $P(Y=1|A=1, B=0)=P(Y=1|A=0, B=0)$;在不暴露 A 情况下单独暴露 B 和不暴露 B 的患病率也相等,即 $P(Y=1|A=0, B=1)=P(Y=1|A=0, B=0)$;假若同时 A 和 B 都不暴露的患病率与同时 A 和 B 都暴露的患病率不相等,即 $P(Y=1|A=0, B=0) \neq P(Y=1|A=1, B=1)$,那么,可以得到:

$$P_{11} - P_{00} > (P_{10} - P_{00}) + (P_{01} - P_{00}) \quad (6)$$

因此,存在生物协同交互作用,即 $P(S) > 0$ 。但是,不等式(6)不成立时,不能确定无生物协同交互作用。

3. 生物交互作用的上下界估计 :Miettinen^[3]提出的生物交互作用的上下界估计方法,在暴露因素都是危险因素的情况下,患病概率应有如下关系: $\pi_{00} \leq \pi_{10} \leq \pi_{11}, \pi_{00} \leq \pi_{01} \leq \pi_{11}$ 。在总体中 6 类子总体所占的比率之和为 1,所以, A, B, S 和 P 类中每一类子总体在总体中所占比率一定在 0 和 1 之间,即:

$$\begin{aligned} 0 \leq P(A) = \pi_{10} - \mu \leq 1, \quad 0 \leq P(B) = \pi_{01} - \mu \leq 1, \\ 0 \leq P(S) = \pi_{11} - \pi_{10} - \pi_{01} + \mu \leq 1, \quad 0 \leq P(P) = \mu - \pi_{00} \leq 1. \end{aligned}$$

根据上式可以得到: $\max(\pi_{00}, \pi_{01} + \pi_{10} - \pi_{11}) \leq \mu \leq \min(\pi_{01}, \pi_{10})$ 。由此,我们可以得到 $P(S)$ 的上下界为:

$$\max(\pi_{11} + \pi_{00} - \pi_{10} - \pi_{01}, 0) \leq P(S) \leq \min(\pi_{11} - \pi_{01}, \pi_{11} - \pi_{10})$$

当 π_{01} 和 π_{10} 中某一个接近 π_{00} 或 π_{11} 时, $P(S)$ 的区间缩短。特别是当 π_{01} 或 π_{10} 中有一个等于 π_{00} 或 π_{11} 时, $P(S)$ 的上下界相等,因此,区间缩短到一个点,如果 π_{01} 和 π_{10} 都位于 π_{00} 与 π_{11} 的中间部位时, $P(S)$ 的区间可能会很宽。

4. 生物交互作用的最大熵估计:由以上的讨论可知,各个不可估计的概率之间相互依赖,知道了一个便是知道了其他。而 μ 正好是它们之间的联系纽带。Darroch, Borkent^[4]提出生物交互作用的最大熵估计,即是寻找 μ 值在各个概率 $P(r)$ 满足观测的约束条件下,使熵函数 $-\sum P \log P$ 达到最大。图 1 列出了这些约束条件。无约束条件下的最大熵估计的分布即是均匀分布,而在限制条件下的最大熵估计可以

看作在限制条件下的最均匀分布。二维列联表情形下的最大熵估计即是认为在两个因素相互独立的条件下的最大似然估计。

$P(S)$	$P(B)$	$\pi_{11} - \pi_{10}$
$P(A)$	$P(P)$	$\pi_{10} - \pi_{00}$
$\pi_{11} - \pi_{01}$	$\pi_{01} - \pi_{00}$	$\pi_{11} - \pi_{00}$

图1 各类概率的边缘约束条件

由熵函数最大可以得到:

$$\hat{\mu} = \pi_{00} + \frac{(\pi_{01} - \pi_{00})(\pi_{10} - \pi_{00})}{\pi_{11} - \pi_{00}}$$

反解出的最大熵估计有如下的乘积形式:

$$\left\{ \frac{P(S)P(B)}{P(A)P(P)} \right\} = \frac{1}{\pi_{11} - \pi_{00}} \left\{ \frac{\pi_{11} - \pi_{10}}{\pi_{10} - \pi_{00}} \right\}^{\pi_{11} - \pi_{01}} \left\{ \frac{\pi_{01} - \pi_{00}}{\pi_{10} - \pi_{00}} \right\}^{\pi_{10} - \pi_{00}}$$

5. 生物交互作用的贝叶斯估计:我们提出生物交互作用的贝叶斯方法。假若专家对各类概率具有先验概率,比如 $P(S)$ 具有 Beta 先验分布,参数为 α 和 β 。可以类似地讨论 $P(A), P(B)$ 或 $P(P)$ 有先验分布的情况。可以采用 MCMC (markov chain monte carlo) 计算 $P(S)$ 的后验分布。

在给定的先验分布情形下,已知观测数据 D , 参数 θ 的后验分布密度为:

$$P(\theta|D) = \frac{f(\theta)g(D|\theta)}{\int f(\theta)g(D|\theta)d\theta}$$

其中 $\theta = [P(\Phi), P(A), P(B), P(S), P(P), P(\Omega)]$, $f(\theta)$ 为 $P(S)$ 的先验分布, $g(D|\theta)$ 为似然函数。 $P(S)$ 的后验分布可由上式计算边缘分布得到,可以使用统计软件来计算。

参 考 文 献

- 1 胡永华, 耿直. 因果推断与混杂概念的讨论. 中华流行病学杂志, 2001, 22: 459-461.
- 2 Rothman KJ, Greenland S. Modern epidemiology. Boston: lippincott-raven. 1998. 329-342.
- 3 Miettinen OS. Causal and preventive interdependence: elementary principles. Scand. J Work Environ. Health, 1982, 8: 159-169.
- 4 Darroch JN, Borkent M. Synergism, attributable risk and interaction for two binary exposure factors. Biometrika, 1994, 81: 259-270.
- 5 Darroch JN. Biologic synergism and parallelism. Am J Epid, 1997, 145: 661-668.

(收稿日期 2001-10-30)

(本文编辑: 段江媚)