高校应用数学学报
2024, 39(4): 426-438

基于交互作用的生存函数中混杂因子的判断准则

范凯旋[1], 李开灿[2*]

(1. 北京工商大学 数学与统计学院, 北京 102488;

2. 湖北师范大学 文理学院, 湖北黄石 435109)

摘　要: 该文利用协变量在总体与子群体的潜在分布中的差异性, 讨论了在生存函数中, 协变量是连续型的条件下, 基于多暴露且存在交互作用的混杂因子判断准则问题, 得到了协变量是一维情况下一致无混杂和一致无关因子的充要条件, 同时也得到了协变量是多维情况下条件一致无混杂和条件一致无关因子向量的充分条件, 从而给出了基于交互作用的生存函数中判断单个混杂因子和多重混杂因子的准则.

关键词: 交互作用; 混杂偏倚; 混杂因子; 一致无混杂; 一致无关因子

中图分类号: O212.4

文献标识码: A　　**文章编号**: 1000-4424(2024)04-0426-13

§1　引　言

在处理效应的统计推断研究中, 如果某种暴露(或者处理)对某种疾病的因果作用受到协变量不同取值的影响而不同, 这种现象称为混杂, 能引起混杂且不受暴露影响的协变量称为混杂因子. 由于可能存在混杂, 在估计因果作用之前, 首先要判断哪些协变量是混杂因子, 目前其判断准则主要有可压缩性准则和可比较性准则两类. 可压缩性准则依赖于变量间的关联测度, 可压缩一般对应无混杂, 从而在估计因果作用时可以忽略掉不会引起混杂的协变量[1]. 然而关联测度的选择往往使得混杂因子的判断准则具有较强的主观性, 且有时不可压缩却对应无混杂[2], 因此在处理效应的统计推断中基于潜在结果模型的可比较性准则的应用更广泛. 文[3]利用很多流行病学的案例归纳总结出判断混杂因子的准则, 文[4]在潜在混杂因子集是混杂的充分控制集的假设下推广了Miettinen-Cook的准则, 文[5]利用标准化的方法给出了离散型协变量中混杂因子的判断准则, 文[6]提出了离散型协变量中判断多重混杂因子的方法. 以上研究均考虑的是单一暴露情况下混杂因子的判断准则, 但是在现实情况中可能存在多个暴露, 且多个暴露之间存在交互

收稿日期: 2022-09-22　　修回日期: 2024-09-29

*通讯作者, E-mail: lkstat@163.com

基金项目: 国家自然科学基金(11471105)

作用. 在这种情况下, 文[7]给出了基于优势比测度下Logistic回归模型中具有交互作用的混杂因子判别条件. 从文[8-10]的讨论可以知道, 这些研究在处理效应的统计推断中是十分有意义的.

在生存分析中, 常常用一个非负随机变量T来表示寿命, 在考虑系统性因素对寿命的影响时, 按照文[11]的描述方式, 假设X是暴露(处理)变量, V是协变量, 在一定的条件下, 考虑协变量可能使寿命和暴露之间的关系产生完全相反的结论, 即出现Simpson悖论. 根据处理效应统计推断的相关描述, 可能存在混杂, 即V可能是混杂因子. 上述学者给出的是流行病学中基于离散型协变量判断混杂因子的准则, 而在生存分析的范围内, 协变量是连续型的. 另外, 文[2]基于因果网络(因果图模型)给出的后门准则也可以判断混杂因子, 这种方法依赖于路径分析. 其主要问题是因果网络路径的学习是困难的, 研究者需要接受因果关系假设的正确性, 这导致因果网络的方法可能不适合在生存函数中判断混杂因子. 笔者利用一致无关因子给出了一个间接判断混杂因子的准则, 从而解决了该问题. 考虑到可能存在不止一个暴露的影响, 本文将致力于研究寿命关于多个变量暴露且有交互作用, 连续型协变量在生存函数中判断混杂的准则问题, 这对生存分析中影响寿命的混杂因子的辨别和因果作用的估计具有极其重要的意义.

§2给出了暴露变量、协变量、潜在生存函数等的基本概念和记号. §3在考虑多种暴露影响且存在交互作用, 以及协变量是连续型的条件下, 利用协变量在总体与子群体的潜在分布中的差异性, 通过遍历协变量值域的任意非空可测子集, 得到了协变量是一维情况下因果寿命分布一致无混杂和一致无关因子的充要条件, 同时也得到了协变量是多维情况下因果寿命分布条件一致无混杂和条件一致无关因子向量的充分条件, 进而给出了生存函数中基于交互作用的混杂因子的判断准则. §4对本文研究的内容进行了总结和讨论.

§2 基本概念与记号

记T为寿命变量, $F(t)=P(T\le t)$是T的分布函数, $f(t)$是T的概率密度函数, $S(t)=P(T>t)$为T的生存函数, 如果寿命T受到暴露(处理)变量X和协变量V的作用时, 则给定$X=x$, $V=v$的条件概率密度函数和条件生存函数分别记为$f(t|x,v)$, $S(t|x,v)=P(T>t|X=x,V=v)$, 而$f(t|x)$, $S(t|x)=P(T>t|X=x)$分别有类似的称呼.

令X为离散型多值暴露变量, T_{x_0}表示在暴露$X=x_0$中群体对应的寿命. 假设对于一个总体而言, 暴露的任意取值$X=x$的子群体均存在, 则$S_{x_0}(t|x_0)=P(T_{x_0}>t|X=x_0)$表示在暴露$X=x_0$中子群体对应的潜在生存函数. $S_{x_0}(t)=P(T_{x_0}>t)$表示总体在暴露$X=x_0$中对应的潜在生存函数, 称为T_{x_0}的因果寿命分布. 考虑连续型协变量V时, $S_{x_0}(t|x_0,v)=P(T_{x_0}>t|X=x_0,V=v)$和$S_{x_0}(t|v)=P(T_{x_0}>t|V=v)$分别表示给定$V=v$条件下暴露取值为$X=x_0$的子群体对应的潜在生存函数和暴露取值为$X=x_0$的总体对应的潜在生存函数. 假设暴露变量和协变量对寿命T有某种影响, 且协变量不受暴露变量的影响, 作者想要研究的是存在多种暴露变量且有交互作用时, 连续型协变量是否为混杂因子, 判断准则是什么? 截止目前, 在生存函数的因果推断中, 笔者还没有发现有这方面的研究文献.

本文采用文[12]的记号, $X\perp\!\!\!\perp V$表示随机变量(或随机向量)X与V相互独立, $T\perp\!\!\!\perp V|X$表示给定X的条件下, T与V相互独立.

§3 基于交互作用的生存函数中判断混杂因子的准则

本节主要在多暴露变量存在交互作用的条件下, 从协变量是一维连续型和多维连续型两种情况分别给出单个混杂因子和多重混杂因子的判断准则.

为了简化讨论, 假设存在两个暴露X_1和X_2的共同影响, 研究的暴露变量均为二值型, 则$S_{x_1,x_2}(t|X_1=x_1,X_2=x_2)=P(T_{x_1,x_2}>t|X=x_1,X_2=x_2)$表示在暴露$X_1=x_1$和$X_2=x_2$共同作用中子群体对应的潜在生存函数. 不失一般性, 设协变量V的密度函数$f(v)$的支撑集为$\mathbf{R}$, 以下先给出交互作用的基本概念.

定义3.1 设X_1和X_2是两个二值暴露(处理)变量, 分别均取值为1, 0, 表示暴露与不暴露, 如果X_1对寿命T的因果作用在X_2分别取0和1时不同, 即

$$\begin{aligned}&S_{1,1}(t|X_1=1,X_2=1)-S_{0,1}(t|X_1=0,X_2=1)\\ \neq&S_{1,0}(t|X_1=1,X_2=0)-S_{0,0}(t|X_1=0,X_2=0),\end{aligned}$$

称X_1和X_2之间存在交互作用.

由于在使用交互作用概念的时候必须说明所使用的关联测度[13], 比如文[7]讨论的是基于优势比测度下Logistic回归模型中具有交互作用的混杂因子判别条件, 然而关联测度的选择具有主观上的局限性, 而且根据反事实结局, 受到单一暴露变量影响的总体中全部个体的响应类型有四种: 劣势型、优势型、受益型和受害型(如表1所示), 如果存在两种暴露及其交互作用的影响, 情况如文[14]首次描述的16种不同响应类型, 由此可见多种暴露存在交互作用的情况十分复杂. 文[15]给出了一个新的研究角度, 考虑将两个暴露变量看作有四种不同取值的单一暴露变量$X=(X_1,X_2)$, 取值分别为(0,0), (0,1), (1,0), (1,1), 这样既可以忽略所选的关联测度, 又可以避免讨论复杂的交互作用类型的问题. 这时, 基于交互作用的多暴露变量生存函数中混杂因子的判断准则问题转化成多值的单一暴露生存函数中混杂因子的判断准则问题, 同理对于超过两个暴露且存在交互作用的情况也有类似的考虑方式.

表 1 单一暴露下的四种响应类型

响应类型	$X=0$	$X=1$
劣势型	$T_0<t$	$T_1<t$
优势型	$T_0>t$	$T_1>t$
受益型	$T_0<t$	$T_1>t$
受害型	$T_0>t$	$T_1<t$

不失一般性, 设协变量V的密度函数$f(v)$的支撑集为R, R是实数集$\mathbf{R}$的子集, 以下给出混杂偏倚等概念.

定义3.2 对于给定的暴露$X=x_0$, 寿命T_{x_0}的潜在生存函数在暴露$X=x_0$作用下的混杂偏倚表示为$B=S_{x_0}(t)-S_{x_0}(t|x_0)$. 设$\omega$为$R$的可测子集, $P(V\in\omega)>0$, 令

$$B_\omega=S_{x_0}(t|V\in\omega)-S_{x_0}(t|x_0,V\in\omega).$$

称B_ω为潜在生存函数在暴露$X=x_0$作用下, 协变量V在ω上产生的混杂偏倚.

注记3.1 当$B\neq 0$时，则存在混杂因子.

定义3.3 给定暴露$X=x_0$, 如果

$$S_{x_0}(t)=S_{x_0}(t|x_0),\forall t,$$

即$B=0$, 称T_{x_0}的因果寿命分布在给定暴露$X=x_0$中是无混杂的.

定义3.4 对于给定的暴露$X=x_0$, 令
$$B_v = S_{x_0}(t|v) - S_{x_0}(t|x_0, v).$$
称B_v是潜在生存函数在暴露$X=x_0$作用下, 协变量在$V=v$处的混杂偏倚. 如果
$$S_{x_0}(t|v) = S_{x_0}(t|x_0, v), \forall t, v,$$
即$B_v=0$, 称T_{x_0}的因果寿命分布关于协变量V在给定暴露$X=x_0$中是条件无混杂的. 如果存在x', t', v', 使得$S_{x'}(t'|v') \neq S_{x'}(t'|x', v')$, 则称$T_{x'}$的因果寿命分布关于协变量$V$在给定暴露$X=x'$中是条件混杂的, 此时$V$就是混杂因子.

注记3.2 若T_{x_0}的因果寿命分布在给定暴露$X=x_0$中是无混杂的, 则不需要根据协变量V对总体进行分层, 即T_{x_0}的因果寿命分布关于协变量V在给定暴露$X=x_0$中一定是条件无混杂的, 从而不存在混杂因子.

文[16]提出的处理分配机制的可忽略性假定是观察性研究中判断混杂因子的最重要的假定. 与之类似, 本文给出生存分析中基于生存函数的暴露分配机制可忽略性假定, 下文对混杂因子判断准则的讨论也多次运用到这个假定, 现在给出这个假定的表述.

假定3.1(暴露分配机制的可忽略性假定) 对于暴露的任意取值$X=x$, 如果满足

(i) $T_x \perp\!\!\!\perp X|V$, 且

(ii) $0 < P(X=x|V=v) < 1$,

那么称暴露分配机制是可忽略的.

这里给出的暴露分配机制可忽略性假定是一种弱可忽略性, 即对于暴露的四个取值$(0,0), (0,1), (1,0), (1,1)$分别有$T_{00} \perp\!\!\!\perp X|V$, $T_{01} \perp\!\!\!\perp X|V$, $T_{10} \perp\!\!\!\perp X|V$和$T_{11} \perp\!\!\!\perp X|V$都成立, 与文[16]中的二值处理下强可忽略性不同.

3.1 单个混杂因子的判断准则

文[5]在暴露变量和响应变量都是二值且协变量是离散型的时候给出了混杂因子的判别方法, 而在本文中寿命变量T和协变量V都是连续型, 且暴露变量转化为多值型. 一般而言, 将离散型变量过渡到连续型时, 相应的结论往往是很难得到, 例如文[17]在证明分布相依度的可压缩性条件时, 发现连续型变量远比离散型变量的情况更复杂. 本节通过遍历连续型协变量值域的任意非空可测子集去定义一致无混杂和一致无关因子, 探索了协变量在暴露的子群体和暴露的总体时的分布差异, 从而获得了多个暴露且存在交互作用时单个混杂因子的判断准则.

定义3.5 给定暴露$X=x_0$, 如果对于R的任意非空可测子集ω总有
$$S_{x_0}(t|V\in\omega) = S_{x_0}(t|x_0, V\in\omega), \forall t,$$
即$B_\omega=0$, 称T_{x_0}的因果寿命分布关于协变量V在给定暴露$X=x_0$中是一致无混杂的.

不失一般性, 以下均假设$P(X=x, V\in\omega) > 0, \forall x, \omega$. 为了给出单个混杂因子的判断准则, 下面先给出一个引理.

引理3.1 设非空可测集$\omega_1 \subseteq \omega, \omega_2 \subseteq \omega$, 使得$\omega_1 \cup \omega_2 = \omega$且$\omega_1 \cap \omega_2 = \varnothing$. 如果$T_{x_0}$的因果寿命分布关于协变量$V$在给定暴露$X=x_0$中是一致无混杂的, 则

(i) $P(T_{x_0} > t, V\in\omega|X=x_0) = P(T_{x_0} > t|X=x_0)P(V\in\omega|X=x_0)$, 或者

(ii) $P(V\in\omega|X=x_0) = P(V\in\omega)$.

证 因为T_{x_0}的因果寿命分布关于协变量V在给定暴露$X=x_0$中是一致无混杂的, 故
$$P(T_{x_0} > t|V\in\omega) = P(T_{x_0} > t|X=x_0, V\in\omega), \forall t, \omega\subseteq R, \tag{1}$$

同时$P(T_{x_0} > t|V \in \omega_i) = P(T_{x_0} > t|X = x_0, V \in \omega_i), i = 1, 2$也成立. 又由全概率公式, 方程(1)可以转化为

$$\begin{aligned}&P(T_{x_0} > t|X = x_0, V \in \omega_1)P(V \in \omega_1|V \in \omega)+\\&P(T_{x_0} > t|X = x_0, V \in \omega_2)P(V \in \omega_2|V \in \omega) =\\&P(T_{x_0} > t|X = x_0, V \in \omega_1)P(V \in \omega_1|X = x_0, V \in \omega)+\\&P(T_{x_0} > t|X = x_0, V \in \omega_2)P(V \in \omega_2|X = x_0, V \in \omega),\end{aligned} \tag{2}$$

因为

$$P(V \in \omega_1|X = x_0, V \in \omega) + P(V \in \omega_2|X = x_0, V \in \omega) = 1,$$

所以方程(2)可以转化成

$$\begin{aligned}&[P(T_{x_0} > t|X = x_0, V \in \omega_1) - P(T_{x_0} > t|X = x_0, V \in \omega_2)]\times\\&[P(V \in \omega_1|V \in \omega, X = x_0) - P(V \in \omega_1|V \in \omega)] = 0,\end{aligned}$$

则

$$P(T_{x_0} > t|X = x_0, V \in \omega_1) = P(T_{x_0} > t|X = x_0, V \in \omega_2), \tag{3}$$

或者

$$P(V \in \omega_1|V \in \omega, X = x_0) = P(V \in \omega_1|V \in \omega). \tag{4}$$

(i) 当方程(3)成立时, 即

$$\frac{P(T_{x_0} > t, X = x_0, V \in \omega_1)}{P(X = x_0, V \in \omega_1)} = \frac{P(T_{x_0} > t, X = x_0, V \in \omega_2)}{P(X = x_0, V \in \omega_2)},$$

则

$$\begin{aligned}\frac{P(T_{x_0} > t, X = x_0, V \in \omega_1)}{P(X = x_0, V \in \omega_1)} =&\frac{P(T_{x_0} > t, X = x_0, V \in \omega_1) + P(T_{x_0} > t, X = x_0, V \in \omega_2)}{P(X = x_0, V \in \omega_1) + P(X = x_0, V \in \omega_2)} =\\&\frac{P(T_{x_0} > t, X = x_0, V \in \omega)}{P(X = x_0, V \in \omega)} = \frac{P(T_{x_0} > t, X = x_0, V \in \omega_2)}{P(X = x_0, V \in \omega_2)},\end{aligned}$$

即

$$P(T_{x_0} > t|X = x_0, V \in \omega_1) = P(T_{x_0} > t|X = x_0, V \in \omega_2) = P(T_{x_0} > t|X = x_0, V \in \omega).$$

由ω的任意性可知, 取$\omega_1 = \omega$，$\omega_2 = \overline{\omega}$, 则

$$P(T_{x_0} > t|X = x_0, V \in \omega) = P(T_{x_0} > t|X = x_0, V \in \omega + \overline{\omega}) = P(T_{x_0} > t|X = x_0),$$

于是

$$\frac{P(T_{x_0} > t, X = x_0, V \in \omega)}{P(X = x_0, V \in \omega)} = \frac{P(T_{x_0} > t, V \in \omega|X = x_0)}{P(V \in \omega|X = x_0)} = P(T_{x_0} > t|X = x_0),$$

即

$$P(T_{x_0} > t, V \in \omega|X = x_0) = P(T_{x_0} > t|X = x_0)P(V \in \omega|X = x_0).$$

(ii) 当方程(3)不成立时, 则方程(4)成立. 因为

$$P(V \in \omega_1|V \in \omega, X = x_0) = \frac{P(V \in \omega_1|X = x_0)}{P(V \in \omega|X = x_0)}, P(V \in \omega_1|V \in \omega) = \frac{P(V \in \omega_1)}{P(V \in \omega)},$$

故方程(4)等价于

$$\frac{P(V \in \omega_1|X = x_0)}{P(V \in \omega_1)} = \frac{P(V \in \omega|X = x_0)}{P(V \in \omega)},$$

由于ω_1和ω_2地位的对称性, 则

$$\frac{P(V \in \omega_1|X = x_0)}{P(V \in \omega_1)} = \frac{P(V \in \omega|X = x_0)}{P(V \in \omega)} = \frac{P(V \in \omega_2|X = x_0)}{P(V \in \omega_2)},$$

由ω的任意性可知, 取$\omega_1=\omega$, $\omega_2=\overline{\omega}$, 则

$$\frac{P(V\in\omega|X=x_0)}{P(V\in\omega)}=\frac{P(V\in\omega|X=x_0)+P(V\in\overline{\omega}|X=x_0)}{P(V\in\omega)+P(V\in\overline{\omega})}=\frac{P(V\in R|X=x_0)}{P(V\in R)}=1,$$

即

$$P(V\in\omega|X=x_0)=P(V\in\omega).$$

定理3.1 在暴露分配机制的可忽略性假定下, T_{x_0}的因果寿命分布关于协变量V在给定暴露$X=x_0$中一致无混杂的充要条件是

(i) $T_{x_0}\perp\!\!\!\perp V|X=x_0$, 或者

(ii) $f(v)=f(v|x_0)$.

其中$f(v|x_0)$为给定$X=x_0$下, V的条件分布函数$P(V<v|X=x_0)$的条件概率密度函数.

证 *充分性* 为了证明充分性, 只需证明

$$S_{x_0}(t|V\in\omega)=S_{x_0}(t|x_0,V\in\omega),\forall t,\omega\subseteq R$$

成立即可.

$$\begin{aligned}S_{x_0}(t|V\in\omega)=&\frac{P(T_{x_0}>t,V\in\omega)}{P(V\in\omega)}=\frac{\int_t^{+\infty}\int_\omega f(u,v)\mathrm{d}u\mathrm{d}v}{P(V\in\omega)}=\\&\frac{\int_t^{+\infty}\int_\omega f(u|v)f(v)\mathrm{d}u\mathrm{d}v}{P(V\in\omega)}=\int_\omega P(T_{x_0}>t|V=v)\frac{f(v)\mathrm{d}v}{P(V\in\omega)}.\end{aligned}$$

由暴露分配机制的可忽略性假定3.1可知, $P(T_{x_0}>t|V=v)=P(T_{x_0}>t|X=x_0,V=v)$, 则

$$S_{x_0}(t|V\in\omega)=\int_\omega P(T_{x_0}>t|X=x_0,V=v)\frac{f(v)\mathrm{d}v}{P(V\in\omega)}.\tag{5}$$

(i) 设$T_{x_0}\perp\!\!\!\perp V|X=x_0$, 则$P(T_{x_0}>t|X=x_0,V=v)=P(T_{x_0}>t|X=x_0)$, 代入(5)式可得

$$\begin{aligned}S_{x_0}(t|V\in\omega)=&\int_\omega P(T_{x_0}>t|X=x_0)\frac{f(v)\mathrm{d}v}{P(V\in\omega)}=\\&P(T_{x_0}>t|X=x_0)\int_\omega\frac{f(v)\mathrm{d}v}{P(V\in\omega)}=P(T_{x_0}>t|X=x_0),\end{aligned}\tag{6}$$

再利用$T_{x_0}\perp\!\!\!\perp V|X=x_0$, 有$P(T_{x_0}>t|X=x_0)=P(T_{x_0}>t|X=x_0,V\in\omega)$, 所以

$$S_{x_0}(t|V\in\omega)=P(T_{x_0}>t|X=x_0,V\in\omega)=S_{x_0}(t|x_0,V\in\omega),$$

由定义3.5可知, T_{x_0}的因果寿命分布关于协变量V在给定暴露$X=x_0$中是一致无混杂的.

(ii) 设$f(v)=f(v|x_0)$, 则$P(V\in\omega|X=x_0)=P(V\in\omega)$, 那么

$$\begin{aligned}&S_{x_0}(t|x_0,V\in\omega)=\frac{P(T_{x_0}>t,X=x_0,V\in\omega)}{P(X=x_0,V\in\omega)}=\\&\frac{P(T_{x_0}>t,V\in\omega|X=x_0)}{P(V\in\omega|X=x_0)}=\frac{\int_t^{+\infty}\int_\omega f(u,v|x_0)\mathrm{d}u\mathrm{d}v}{P(V\in\omega|X=x_0)}=\frac{\int_t^{+\infty}\int_\omega f(u|v,x_0)f(v|x_0)\mathrm{d}u\mathrm{d}v}{P(V\in\omega|X=x_0)}=\\&\int_\omega P(T_{x_0}>t|X=x_0,V=v)\frac{f(v|x_0)\mathrm{d}v}{P(V\in\omega|X=x_0)}=\\&\int_\omega P(T_{x_0}>t|X=x_0,V=v)\frac{f(v)\mathrm{d}v}{P(V\in\omega)},\end{aligned}$$

由(5)式得

$$S_{x_0}(t|V\in\omega)=S_{x_0}(t|x_0,V\in\omega),$$

由定义3.5可知, T_{x_0}的因果寿命分布关于协变量V在给定暴露$X=x_0$中是一致无混杂的.

必要性 因为T_{x_0}的因果寿命分布关于协变量V在给定暴露$X=x_0$中是一致无混杂的,

故由引理1可知, $P(T_{x_0} > t, V \in \omega|X = x_0) = P(T_{x_0} > t|X = x_0)P(V \in \omega|X = x_0)$或者$P(V \in \omega|X = x_0) = P(V \in \omega)$成立.

对于$P(T_{x_0} > t, V \in \omega|X = x_0) = P(T_{x_0} > t|X = x_0)P(V \in \omega|X = x_0)$, 由条件独立性的定义知, $T_{x_0} \perp\!\!\!\perp V|X = x_0$成立.

当$P(V \in \omega|X = x_0) = P(V \in \omega)$时, 即$\forall\omega$, 有

$$\int_{\omega} f(v|x_0)\mathrm{d}v = \int_{\omega} f(v)\mathrm{d}v.$$

由ω的任意性, 根据文[18]第71页定理3.2.2(2)可知, $f(v|x_0) = f(v)$ a.e., 从而必要性得证.

这里是对于暴露的特定取值$X = x_0$, 有$T_{x_0} \perp\!\!\!\perp V|X = x_0$或者$f(v) = f(v|x_0)$成立, 而对于$X = x'', x'' \neq x_0$时, 可能使得一致无混杂不成立, 即$S_{x''}(t|V \in \omega) \neq S_{x''}(t|x'', V \in \omega)$, 但如果对于暴露的任意取值$X = x$, 都有定理3.1的两个条件成立, 则这两个条件可以写成$T_x \perp\!\!\!\perp V|X = x$或者$V \perp\!\!\!\perp X$, 下面给出一致无关因子的定义.

定义3.6 对于任意的暴露取值$X = x$和R的任意非空可测子集ω总有

$$S_x(t|V \in \omega) = S_x(t|x, V \in \omega), \forall t,$$

称V为一个一致无关因子.

注记3.3 V为一个一致无关因子意味着T_x的因果寿命分布关于协变量V在任意暴露$X = x$中是一致无混杂的.

定理3.2 在暴露分配机制的可忽略性假定下, 一个协变量V是一致无关因子的充要条件是

(i) $T_x \perp\!\!\!\perp V|X = x$, 或者

(ii) $V \perp\!\!\!\perp X$.

证 充分性由定理3.1及暴露取值的任意性易得.

必要性 因为协变量V是一致无关因子, 由定理3.1和注记3.3可得$T_x \perp\!\!\!\perp V|X = x$或者$V \perp\!\!\!\perp X$, 所以结论成立.

定义3.7 设$\omega_1, \omega_2, \cdots, \omega_s (s \geq 2)$是$R$的$s$个非空可测子集, 若同时满足

(i) $\omega_i \cap \omega_j = \varnothing, i \neq j$,

(ii) $P(V \in \omega_i) > 0, i = 1, 2, \cdots, s$,

(iii) $\bigcup\limits_{i} \omega_i = R$,

则称$\mathscr{P} = \{\omega_1, \omega_2, \cdots, \omega_s\}$为$R$关于$V$的一个分割.

定理3.3 在暴露分配机制的可忽略性假定下, 如果一个协变量V是一致无关因子, 则T_x的因果寿命分布关于协变量V在任意暴露$X = x$中是无混杂的.

证 因为协变量V是一个一致无关因子, 故

$$S_x(t|V \in \omega) = S_x(t|x, V \in \omega), \forall t, \omega \subseteq R,$$

由定理3.2可知, 协变量V是一个一致无关因子可以推出$T_x \perp\!\!\!\perp V|X = x$或者$V \perp\!\!\!\perp X$.

(i) 当$T_x \perp\!\!\!\perp V|X = x$成立时, 可得与方程(6)类似的形式, 即$S_x(t|V \in \omega) = P(T_x > t|X = x)$, 设对于$R$关于$V$的任意一个分割$\mathscr{P} = \{\omega_1, \omega_2, \cdots, \omega_s\}$使得$\sum\limits_{i=1}^{s} P(V \in \omega_i) = 1$, 则

$$S_x(t) = \sum_{i=1}^{s} S_x(t|V \in \omega_i)P(V \in \omega_i) = P(T_x > t|X = x)\sum_{i=1}^{s} P(V \in \omega_i) = S_x(t|x),$$

由定义3.3及暴露取值的任意性知, T_x的因果寿命分布关于协变量V在任意暴露$X = x$中是无混杂的.

(ii) 当$V \perp\!\!\!\perp X$成立时, 可得$P(V \in \omega|X = x) = P(V \in \omega), \forall \omega \subseteq R$, 则对于$R$关于$V$的任意一个分割$\mathscr{P} = \{\omega_1, \omega_2, \cdots, \omega_s\}$,

$$S_x(t) = \sum_{i=1}^{s} S_x(t|V \in \omega_i) P(V \in \omega_i) = \sum_{i=1}^{s} S_x(t|x, V \in \omega_i) P(V \in \omega_i|X = x) = S_x(t|x),$$

由定义3.3及暴露取值的任意性知, T_x的因果寿命分布关于协变量V在任意暴露$X = x$中是无混杂的.

由注记3.2和定理3.3可知, 如果协变量V是一个一致无关因子, 则它不是混杂因子, 从而间接得到了一维协变量情况下判断混杂因子的准则.

3.2 多重混杂因子的判断准则

协变量是多维时, 设协变量$V = (V_1, V_2, \cdots, V_n)$, 协变量的某些分量可能不会引起混杂, 如果能控制那些可以引起混杂的分量, 即混杂因子向量, 则其他分量均可忽略掉, 这时称混杂因子向量为充分控制向量, 显然最大的充分控制向量是协变量本身. 若对协变量进行控制, 则控制了混杂因子向量, 从而有$T_x \perp\!\!\!\perp X|V$, 这与暴露机制的可忽略性假定3.1是对应的. 文[6]在讨论多维离散型协变量是否为混杂因子集的判断准则时, 否认了弱可忽略性的可用性, 而是利用因果图的方式得到了协变量给定的条件下, 响应变量与处理变量独立的结论, 但给出的暴露机制的可忽略性假定3.1在响应变量和协变量都是连续型时是可用的. 本小节将给出协变量是多维连续型情况下多重混杂因子的判断准则.

令R_k表示协变量V的分量V_k的密度函数的支撑集, ω_k表示R_k的一个非空可测子集, 则$R = R_1 \times R_2 \times \cdots \times R_n$, $\omega = (\omega_1, \omega_2, \cdots, \omega_n)^\top$. 令$V'$ 和V''表示V的分块, 即$V = (V', V'')$, 其中V'的维数是q, $0 < q < n$.

定义3.8 给定暴露$X = x_0$, 如果对于V各个分量的支撑集的任意非空可测子集$\omega_k \subseteq R_k$, $k = 1, 2, \cdots, n$, 总有

$$S_{x_0}(t|V_1 \in \omega_1, \cdots, V_n \in \omega_n) = S_{x_0}(t|x_0, V_1 \in \omega_1, \cdots, V_n \in \omega_n), \forall t,$$

即

$$S_{x_0}(t|V \in \omega) = S_{x_0}(t|x_0, V \in \omega), \forall t,$$

那么称T_{x_0}的因果寿命分布关于协变量V在给定暴露$X = x_0$中是一致无混杂的.

定义3.9 对于暴露$X = x_0$, 设$R' = R_1 \times R_2 \times \cdots \times R_q$, 如果对于任意$V'' = v''$和$\omega' \subseteq R'$, 总有

$$S_{x_0}(t|V' \in \omega', v'') = S_{x_0}(t|x_0, V' \in \omega', v''), \forall t,$$

称T_{x_0}的因果寿命分布在V''条件下关于V'在给定暴露$X = x_0$中是一致无混杂的.

定义3.10 对于任意的暴露取值$X = x$, $V'' = v''$ 和$\omega' \subseteq R'$, 总有

$$S_x(t|V' \in \omega', v'') = S_x(t|x, V' \in \omega', v''), \forall t,$$

称V'为一个给定V''的一致无关因子向量.

定理3.4 在暴露分配机制的可忽略性假定下, 令$f(v'|x_0, v'')$为给定$X = x_0$和$V'' = v''$下, V'的条件分布函数$P(V' < v'|X = x_0, V'' = v'')$的条件概率密度函数, 如果

(i) $T_{x_0} \perp\!\!\!\perp V'|(X = x_0, V'')$, 或者

(ii) $f(v'|v'') = f(v'|x_0, v'')$,

则T_{x_0}的因果寿命分布在V''条件下关于V'在给定暴露$X = x_0$中是一致无混杂的.

证 由定义3.9可知, 只需证明

$$S_{x_0}(t|V'\in\omega',v'')=S_{x_0}(t|x_0,V'\in\omega',v''),\forall t,v'',\omega'\subseteq R'$$

成立即可. 显然

$$S_{x_0}(t|V'\in\omega',v'')=\frac{P(T_{x_0}>t,V'\in\omega'|V''=v'')}{P(V'\in\omega'|V''=v'')}=\frac{\int_t^{+\infty}\int_{\omega'}f(u,v'|v'')\mathrm{d}u\mathrm{d}v'}{P(V'\in\omega'|V''=v'')}=$$

$$\frac{\int_t^{+\infty}\int_{\omega'}f(u|v',v'')f(v'|v'')\mathrm{d}u\mathrm{d}v'}{P(V'\in\omega'|V''=v'')}=\frac{\int_{\omega'}P(T_{x_0}>t|V'=v',V''=v'')f(v'|v'')\mathrm{d}v'}{P(V'\in\omega'|V''=v'')},$$

由暴露分配机制的可忽略性假定3.1可知

$$P(T_{x_0}>t|V'=v',V''=v'')=P(T_{x_0}>t|X=x_0,V'=v',V''=v''),$$

则

$$S_{x_0}(t|V'\in\omega',v'')=\frac{\int_{\omega'}P(T_{x_0}>t|X=x_0,V'=v',V''=v'')f(v'|v'')\mathrm{d}v'}{P(V'\in\omega'|V''=v'')}. \tag{7}$$

(i) 设$T_{x_0}\perp\!\!\!\perp V'|(X=x_0,V'')$, 则

$$P(T_{x_0}>t|X=x_0,V'=v',V''=v'')=P(T_{x_0}>t|X=x_0,V''=v''),$$

代入(7)式可得

$$\begin{aligned}S_{x_0}(t|V'\in\omega',v'')=&\frac{\int_{\omega'}P(T_{x_0}>t|X=x_0,V''=v'')f(v'|v'')\mathrm{d}v'}{P(V'\in\omega'|V''=v'')}=\\&\frac{P(T_{x_0}>t|X=x_0,V''=v'')\int_{\omega'}f(v'|v'')\mathrm{d}v'}{P(V'\in\omega'|V''=v'')}=\\&P(T_{x_0}>t|X=x_0,V''=v''),\end{aligned}$$

再利用$T_{x_0}\perp\!\!\!\perp V'|(X=x_0,V'')$, 有

$$P(T_{x_0}>t|X=x_0,V''=v'')=P(T_{x_0}>t|X=x_0,V'\in\omega',V''=v''),$$

所以

$$S_{x_0}(t|V'\in\omega',v'')=P(T_{x_0}>t|X=x_0,V'\in\omega',V''=v'')=S_{x_0}(t|x_0,V'\in\omega',v''),$$

由定义3.9可知, T_{x_0}的因果寿命分布在V''条件下关于V'在给定暴露$X=x_0$中是一致无混杂的.

(ii) 设$f(v'|v'')=f(v'|x_0,v'')$, 则$P(V'\in\omega'|V''=v'')=P(V'\in\omega'|X=x_0,V''=v'')$, 则

$$S_{x_0}(t|x_0,V'\in\omega',v'')=\frac{P(T_{x_0}>t,V'\in\omega'|X=x_0,V''=v'')}{P(V'\in\omega'|X=x_0,V''=v'')}=$$

$$\frac{\int_t^{+\infty}\int_{\omega'}f(u,v'|x_0,v'')\mathrm{d}u\mathrm{d}v'}{P(V'\in\omega'|X=x_0,V''=v'')}=\frac{\int_t^{+\infty}\int_{\omega'}f(u|x_0,v',v'')f(v'|x_0,v'')\mathrm{d}u\mathrm{d}v'}{P(V'\in\omega'|X=x_0,V''=v'')}=$$

$$\int_{\omega'}P(T_{x_0}>t|X=x_0,V'=v',V''=v'')\frac{f(v'|x_0,v'')\mathrm{d}v'}{P(V'\in\omega'|X=x_0,V''=v'')}=$$

$$\int_{\omega'}P(T_{x_0}>t|X=x_0,V'=v',V''=v'')\frac{f(v'|v'')\mathrm{d}v'}{P(V'\in\omega'|V''=v'')},$$

由(7)式得

$$S_{x_0}(t|V'\in\omega',v'')=S_{x_0}(t|x_0,V'\in\omega',v''),$$

由定义3.9可知, T_{x_0}的因果寿命分布在V''条件下关于V'在给定暴露$X=x_0$中是一致无混杂的.

如果对于暴露的任意取值$X=x$, 均有定理3.4的充分条件成立, 则这两个条件可以写成$T_x\perp\!\!\!\perp Y'|(X=x,V'')$或者$V'\perp\!\!\!\perp X|V''$. 令$V^-(i)=V\backslash V_i$表示$V$剔除一个特定的分量$V_i$后的协变量, 于是就有如下的定理3.5.

定理3.5 在暴露分配机制的可忽略性假定下, 对于暴露的任意取值$X=x$, 如果

(i) $T_x \perp\!\!\!\perp V_i|(X=x, V^-(i))$, 或者

(ii) $V_i \perp\!\!\!\perp X|V^-(i)$,

则V_i为一个给定$V^-(i)$的一致无关因子向量.

证 由定理3.4及暴露取值的任意性易得.

由于给定V''条件下的一致无关因子向量一定不是给定相同条件下的混杂因子向量, 如果将协变量V中不引起混杂的分量剔除, 则剩余的分量构成了混杂因子向量. 根据定理5的条件, 将V中的分量分成

$$V^{(1)} = \{V_i : T_x \perp\!\!\!\perp V_i|(X=x, V^-(i))\},$$
$$V^{(2)} = \{V_i : V_i \perp\!\!\!\perp X|V^-(i)\},$$
$$V^{(3)} = V\backslash(V^{(1)} \cup V^{(2)})$$

这三类. 需要注意的是$V^{(1)}$与$V^{(2)}$可能有相同的分量. 下面给出一个推论.

推论3.1 对于暴露的任意取值$X=x$, 如果V被分成以上三类$V^{(1)}$, $V^{(2)}$和$V^{(3)}$, 则

(i) $V^{(1)}$为一个给定$V\backslash V^{(1)}$的一致无关因子向量;

(ii) $V^{(2)}$为一个给定$V\backslash V^{(2)}$的一致无关因子向量.

证 对于任意$V_i \in V^{(1)}$, $T_x \perp\!\!\!\perp V^{(1)}|(X=x, V\backslash V^{(1)})$等价于$T_x \perp\!\!\!\perp V_i|(X=x, V^-(i))$, 由定理3.5(i)可得条件(i)成立. 同理可证, 由定理3.5(ii)可得条件(ii)成立.

$V^{(1)}$和$V^{(2)}$分别是给定相应条件下的一致无关因子向量, 但是$V^{(1)} \cup V^{(2)}$可能是一个给定$V^{(3)}$的混杂因子向量, 从而不能通过直接剔除$V^{(1)} \cup V^{(2)}$来获得混杂因子向量. 下面再给出一个定理来判断这种情形.

定理3.6 对于暴露的任意取值$X=x$, V被分成以上三类$V^{(1)}$, $V^{(2)}$和$V^{(3)}$, 在暴露分配机制的可忽略性假定下, 如果存在两个不相交的子向量$V'(\subseteq V^{(1)})$和$V''(\subseteq V^{(2)})$使得$V' \perp\!\!\!\perp V''|[V\backslash(V', V'')]$, 则$(V', V'')$为一个给定$V\backslash(V', V'')$的一致无关因子向量.

证 令$V^c = V\backslash(V', V'')$, 由定义3.10可知, 只需证明

$$S_x(t|V' \in \omega', V'' \in \omega'', v^c) = S_x(t|x, V' \in \omega', V'' \in \omega'', v^c), \forall t, v^c, \omega' \subseteq R', \omega'' \subseteq R''$$

成立即可. 由暴露分配机制的可忽略性假定3.1可知

$$P(T_x > t|V'=v', V''=v'', V^c=v^c) = P(T_x > t|X=x, V'=v', V''=v'', V^c=v^c),$$

则

$$\begin{aligned}
&P(T_x > t|V' \in \omega', V'' \in \omega'', V^c=v^c) = \\
&\frac{P(T_x > t, V' \in \omega', V'' \in \omega''|V^c=v^c)}{P(V' \in \omega', V'' \in \omega''|V^c=v^c)} = \\
&\frac{\int_t^{+\infty}\int_{\omega'}\int_{\omega''} f(u, v', v''|v^c)\mathrm{d}u\mathrm{d}v'\mathrm{d}v''}{P(V' \in \omega', V'' \in \omega''|V^c=v^c)} = \\
&\frac{\int_t^{+\infty}\int_{\omega'}\int_{\omega''} f(u|v', v'', v^c)f(v', v''|v^c)\mathrm{d}u\mathrm{d}v'\mathrm{d}v''}{P(V' \in \omega', V'' \in \omega''|V^c=v^c)} = \\
&\iint_{\omega'\times\omega''} P(T_x > t|V'=v', V''=v'', V^c=v^c)\frac{f(v', v''|v^c)\mathrm{d}v'\mathrm{d}v''}{P(V' \in \omega', V'' \in \omega''|V^c=v^c)} = \\
&\iint_{\omega'\times\omega''} P(T_x > t|X=x, V'=v', V''=v'', V^c=v^c)\frac{f(v', v''|v^c)\mathrm{d}v'\mathrm{d}v''}{P(V' \in \omega', V'' \in \omega''|V^c=v^c)},
\end{aligned} \tag{8}$$

由$V^{(1)}$的定义知$T_x \perp\!\!\!\perp V'|(X=x,V'',V^c)$, 则

$$P(T_x>t|X=x,V'=v',V''=v'',V^c=v^c)=P(T_x>t|X=x,V''=v'',V^c=v^c),$$

(8)式可以表示为

$$\begin{aligned}&P(T_x>t|V'\in\omega',V''\in\omega'',V^c=v^c)=\\&\iint_{\omega'\times\omega''}P(T_x>t|X=x,V''=v'',V^c=v^c)\frac{f(v',v''|v^c)\mathrm{d}v'\mathrm{d}v''}{P(V'\in\omega',V''\in\omega''|V^c=v^c)},\end{aligned}\tag{9}$$

又因为$V' \perp\!\!\!\perp V''|V^c$, 则

$$f(v',v''|v^c)=f(v'|v^c)f(v''|v^c),$$
$$P(V'\in\omega',V''\in\omega''|V^c=v^c)=P(V'\in\omega'|V^c=v^c)P(V''\in\omega''|V^c=v^c),$$

所以(9)式转变为

$$\begin{aligned}&P(T_x>t|V'\in\omega',V''\in\omega'',V^c=v^c)=\\&\int_{\omega''}P(T_x>t|X=x,V''=v'',V^c=v^c)f(v''|v^c)\mathrm{d}v''\times\\&\frac{\int_{\omega'}f(v'|v^c)\mathrm{d}v'}{P(V'\in\omega'|V^c=v^c)P(V''\in\omega''|V^c=v^c)}=\\&\int_{\omega''}\frac{P(T_x>t|X=x,V''=v'',V^c=v^c)}{P(V''\in\omega''|V^c=v^c)}f(v''|v^c)\mathrm{d}v''.\end{aligned}\tag{10}$$

另一方面, 由暴露分配机制的可忽略性假定3.1和$T_x \perp\!\!\!\perp V'|(X=x,V'',V^c)$可得到与(9)式类似的形式

$$\begin{aligned}&P(T_x>t|X=x,V'\in\omega',V''\in\omega'',V^c=v^c)=\\&\iint_{\omega'\times\omega''}P(T_x>t|X=x,V''=v'',V^c=v^c)\frac{f(v',v''|x,v^c)\mathrm{d}v'\mathrm{d}v''}{P(V'\in\omega',V''\in\omega''|X=x,V^c=v^c)}.\end{aligned}\tag{11}$$

由条件概率的性质知

$$f(v',v''|x,v^c)=f(v''|x,v',v^c)f(v'|x,v^c),$$
$$\begin{aligned}&P(V'\in\omega',V''\in\omega''|X=x,V^c=v^c)=\\&P(V''\in\omega''|X=x,V'\in\omega',V^c=v^c)P(V'\in\omega'|X=x,V^c=v^c),\end{aligned}$$

则(11)式可以转换成

$$\begin{aligned}&P(T_x>t|X=x,V'\in\omega',V''\in\omega'',V^c=v^c)=\\&\int_{\omega''}P(T_x>t|X=x,V''=v'',V^c=v^c)\times\\&\frac{\int_{\omega'}f(v'|x,v^c)\mathrm{d}v'f(v''|x,v',v^c)\mathrm{d}v''}{P(V''\in\omega''|X=x,V'\in\omega',V^c=v^c)P(V'\in\omega'|X=x,V^c=v^c)},\end{aligned}\tag{12}$$

由$V^{(2)}$的定义知$V'' \perp\!\!\!\perp X|(V',V^c)$, 结合$V' \perp\!\!\!\perp V''|V^c$, 则

$$f(v''|x,v',v^c)=f(v''|v',v^c)=f(v''|v^c),$$
$$P(V''\in\omega''|X=x,V'=v',V^c=v^c)=P(V''\in\omega''|V'=v',V^c=v^c)=P(V''\in\omega''|V^c=v^c),$$

所以(12)式可以转化为

$$P(T_x > t|X=x, V'\in\omega', V''\in\omega'', V^c=v^c) =$$
$$\int_{\omega''} P(T_x > t|X=x, V''=v'', V^c=v^c)\frac{\int_{\omega'} f(v'|x,v^c)\mathrm{d}v' f(v''|v^c)\mathrm{d}v''}{P(V''\in\omega''|V^c=v^c)P(V'\in\omega'|X=x,V^c=v^c)} =$$
$$\int_{\omega''} P(T_x > t|X=x, V''=v'', V^c=v^c)\frac{P(V'\in\omega'|X=x,V^c=v^c)f(v''|v^c)\mathrm{d}v''}{P(V''\in\omega''|V^c=v^c)P(V'\in\omega'|X=x,V^c=v^c)} =$$
$$\int_{\omega''} P(T_x > t|X=x, V''=v'', V^c=v^c)\frac{f(v''|v^c)\mathrm{d}v''}{P(V''\in\omega''|V^c=v^c)}, \tag{13}$$

由(10)式和(13)式可知

$$S_x(t|V'\in\omega', V''\in\omega'', v^c) = S_x(t|x, V'\in\omega', V''\in\omega'', v^c),$$

所以(V', V'')为一个给定$V\backslash(V', V'')$的一致无关因子向量.

定理3.6保证了(V', V'')是协变量V中不引起混杂的变量, 如果剔除(V', V''), 余下的变量则构成了混杂因子向量, 从而间接得到了多维协变量情况下判断多重混杂因子的准则.

§4 讨论

本文研究了基于交互作用的生存函数中判断混杂因子的准则. 给出的单个混杂因子和多重混杂因子的判断准则, 均是在暴露分配机制的可忽略性假定(弱可忽略性)下通过判断协变量是否为一致无关因子和条件一致无关因子向量来实现的, 这是一种间接判断混杂因子的准则. 定理3.1和定理3.2给出了判断一致无混杂和一致无关因子的准则, 其中判断一致无关因子的准则, 即$T_x \perp\!\!\!\perp V|X=x$或者$V \perp\!\!\!\perp X$, 在形式上与文[5]在响应变量和协变量均为离散型条件下混杂因子的判断准则相似. 定理3.5, 推论3.1和定理3.6给出了判断多重混杂因子的一个具体方法. 然而检验随机变量的条件独立性较复杂, 所以要找出一致无关因子和条件一致无关因子向量并不是一件很容易的事. 此理论在形式上包含了生存函数中关于离散型协变量的混杂因子判断准则的情况, 在实证方面还未发现关于生存函数中混杂的统计检验研究, 这是一个比较有趣的研究问题.

另一方面, 本文考虑的多暴露变量是离散型的, 且只考虑了一维寿命系统, 而当暴露变量是连续型或者寿命系统是多维时, 生存函数中混杂因子的判断准则又是怎样的, 这也是一个十分值得研究的问题.

参考文献:

[1] Geng Zhi, Liu Yue, Liu Chunchen, et al. Evaluation of causal effects and local structure learning of causal networks[J]. Annual Review of Statistics and Application, 2019, 6(1): 103-124.

[2] Pearl J. Causality: Models, Reasoning, and Inference[M]. Cambridge: Cambridge University Press, 2009.

[3] Miettinen O S, Cook E F. Confounding: essence and detection[J]. American Journal of Epidemiology, 1981, 114(4): 593-603.

[4] Robins J M. Causal inference from complex longitudinal data[C]. In: Latent Variable Modeling with Applications to Causality, New York: Springer, 1997, 69-117.

[5] Geng Zhi, Guo Jianhua, Fung Wing-Kam. Criteria for confounders in epidemiological studies[J]. Journal of the Royal Statistical Society, Series B, 2002, 64(1): 3-15.
[6] Wang Xueli, Geng Zhi, Chen Hua, et al. Detecting multiple confounders[J]. Journal of Statistical Planning and Inference, 2009, 139(3): 1073-1081.
[7] Liu Aihua, Abrahamowicz M, Siemiatycki J. Conditions for confounding of interactions[J]. Pharmacoepidemiology and Drug Safety, 2016, 25(3): 287-296.
[8] Kleinbaum D G, Kupper L L, Morgenstern H. Epidemiologic Research: Principles and Quantitative Methods[M]. New York: Van Nostrand Reinhold, 1982.
[9] Rothman K J. Modern Epidemiology[M]. Boston: Little, Brown and Company, 1986.
[10] Greenland S, Robins J M, Pearl J. Confounding and collapsibility in causal inference[J]. Statistical Science, 1999, 14(1): 29-46.
[11] Serio C D, Rinott Y, Scarsini M. Simpson's paradox in survival models[J]. Scandinavian Journal of Statistics, 2009, 36(3): 463-480.
[12] Dawid A P. Conditional independence in statistical theory[J]. Journal of the Royal Statistical Society, Series B, 1979, 41(1): 1-31.
[13] Vanderweele T J. Explanation in Causal Inference[M]. New York: Oxford University Press, 2015.
[14] Miettinen O S. Causal and preventive interdependence: elementary principles[J]. Scandinavian Journal of Work Environment & Health, 1982, 8(3): 159-168.
[15] Hernán M A, Robins J M. Causal Inference: What If[M]. Boca Raton FL: Chapman & Hall/CRC, 2020.
[16] Rosenbaum P R, Rubin D B. The central role of the propensity score in observational studies for causal effects[J]. Biometrika, 1983, 70(1): 41-55.
[17] Ma Zongming, Xie Xianchao, Geng Zhi. Uniform collapsibility of distribution dependence over a nominal, ordinal or continuous background[J]. Journal of the Royal Statistical Society, Series B, 2006, 68(1): 127-133.
[18] 程士宏. 测度论与概率论基础[M]. 北京: 北京大学出版社, 2004.

Criteria for confounders in survival function based on the interaction

FAN Kai-xuan[1], LI Kai-can[2]

(1. School of Mathematics and Statistics, Beijing Technology and Business University, Beijing 102488, China;

2. College of Arts and Sciences of Hubei Normal University, Huangshi 435109, China)

Abstract: In this paper, the criteria for confounders in survival function based on the interaction under the condition that the covariate is continuous are discussed. When the covariate is one-dimensional, the necessary and sufficient conditions for the uniform non-confounding and uniformly irrelevant factor are obtained by utilizing the potential distribution difference of the covariate between the population and subpopulation. Then the sufficient conditions for the conditional uniform non-confounding and conditional uniformly irrelevant factor vector are also obtained when the covariate is multidimensional. Further, the criteria for detecting a single confounder and multiple confounders in survival function based on the interaction are proposed.

Keywords: interaction; confounding bias; confounder; uniform non-confounding; uniformly irrelevant factor

MR Subject Classification: 62N05; 62H99