心 理 学 报 2006,**38**(1):135～144
Acta Psychologica Sinica

项目反应理论框架下多级评分项目的信息函数

杜文久
(西南大学数学与财经学院,重庆 400715)

摘　要　目的是给出多级评分项目的信息函数计算公式,同时通过几个实例讨论了多级评分项目信息函数在实践中的应用。主要取得了如下成果:(1)首先通过一个例子给出了测验项目的样本空间;(2)以二参数逻辑斯蒂模型为基础,讨论了几种多级评分项目的概率函数,并在此基础上给出了多级评分项目的信息函数计算公式;(3)通过几个实例讨论了多级评分项目信息函数在实践中的应用。

关键词　IRT,信息函数,多级评分项目。
分类号　B841

1　引言

信息函数是项目反应理论的一个重要概念,它在挑选项目、编制测验等方面都有十分重要的应用。但迄今为止,信息函数的概念只适用于二级评分项目,而对于多级评分项目,由于其概率函数难以表达,参数估计十分复杂,因此一直未能得到很好的应用,这就使项目反应理论在实践中的应用受到很大限制。为克服这一缺陷,在本文中,我们将讨论各种类型多级评分项目的信息函数及其在实践中的一些应用。

2　测验项目的样本空间

多级评分项目的信息函数与测验项目的样本空间密切相关,因此,在给出多级评分项目的信息函数以前,我们首先来讨论测验项目的样本空间。先看下面的例子。

例 1　求满足不等式组:$\begin{cases}|2x-1|<4 & (1)\\ x^2-x>0 & (2)\end{cases}$的最小整数解。

我们首先来解答这个问题。

解:由不等式(1),解得

第一步:$-4<2x-1<4$,

第二步:$-3<2x<5$

第三步(A):$-\frac{3}{2}<x<\frac{5}{2}$　　1 分

第四步:又由不等式(2),得 $x(x-1)>0$

第五步(B):解得 $x<0$,或者 $x>1$　　2 分

第六步(C):它们的公共解为 $-\frac{3}{2}<x<0$ 或

$1<x<\frac{5}{2}$　　4 分

第七步(D):所以,满足不等式组的最小整数解为 $x=-1$。　　5 分

在上述解答过程中我们看到,为了求得问题的解,需要对问题进行一系列的变换,但这些变换不是随意的,而是要满足一定的规则。由于例 1 中的字母表示的是实数,因此变换应满足实数的运算规则。

一般情况下,任何一种测验都是由一系列的测验项目所组成,一个测验项目通常有三种状态:初始状态,目标状态,障碍。初始状态是指问题的出发点。目标状态则是指问题想要达到的目标。一个问题不能直接从初始状态达到目标状态,在初始状态与目标状态之间存在着障碍。要使问题从初始状态达到目标状态,就需要对问题进行一系列的变换。但这种变换并不是随意的,而要满足一定的"游戏规则"。

变换规则可以是一条定义,一条性质,一条运算律,一条公理、定理,或者一条定律等等,也可以是一些人为的规定。

变换规则也叫做算子,由变换规则组成的集合叫做算子集。

收稿日期:2004－12－18
通讯作者:杜文久,E-mail:duwenjiu@eduwest.com;电话:023－68213887

一个测验项目在从初始状态到达目标状态的变换过程中,常常有多种变换(多种解法),但在变换过程中,有些子目标可能是必经之点。也就是说,这些子目标必须要达到,问题解决才有可能继续进行下去。我们把这样的一些子目标叫做节点(或者叫做得分点)。

例如,在例 1 的解答过程中,总共进行了七次变换(还有一些变换未写下来),第一步,第二步不是节点,因为一个不等式的解法可以有多种。但第三步是一个节点,因为无论采用何种解法,其解集必然是相同的。

同样,第四步不是节点,但第五步是一个节点。

第六步与第七步也都是节点。

问题解决的目标状态本身也可以看作是一个节点,我们把这样的节点叫做原节点。

问题解决过程中的节点也叫做一个事件。

假设

A:=“$-\frac{3}{2}<x<\frac{5}{2}$”;B:=“$x<0$,或者 $x>1$”;

C:=“$-\frac{3}{2}<x<0,1<x<\frac{5}{2}$”;D:=“$x=-1$”。

通过对例 1 解答过程的分析我们不难发现,无论被试在不等式(1)上的解答是对是错,都不会影响到被试在不等式(2)上的解答。反过来说,无论被试在不等式(2)上的解答是对是错,也不会影响到被试在不等式(1)上的解答。由此可见,事件 A 与事件 B 是相互独立的。

但只有事件 A 与事件 B 同时发生以后,事件 C 才有可能发生。如果事件 A 或者事件 B 不发生,则事件 C 必不发生。因此有$(A\cap B)\supset C$。

同理,只有事件 C 发生以后,事件 D 才有可能发生,因此必有 $C\supset D$。

通过上面的分析我们不难发现,A、B、C、D 四个节点或者四个事件之间满足下列关系:

1)A、B 相互独立; 2)$(A\cap B)\supset C\supset D$。

根据 A、B、C、D 四个节点间的相互关系,问题解决过程中可能出现如下几种结果:

1)事件 A、B 均不发生,即被试在不等式(1)和不等式(2)上的解答都错,这时被试在 C、D 上的解答必然是错的。就是说,发生了事件:$\bar{A}\cdot\bar{B}$,记为 0 分;

2)事件 A 发生但事件 B 不发生;即发生了事件:$A\cdot\bar{B}$,记为 1 分;

3)事件 A 不发生,但事件 B 发生,即被试在不等式(1)上答错,但在不等式(2)上的解答是对的,即发生了事件:$\bar{A}\cdot B$,记为 2 分;

4)事件 A、B 均发生但 C 不发生;即发生了事件:$A\cdot B-C$,记为 3 分;

5)事件 A、B、C 发生,但 D 不发生。即发生了事件:$C-D$,记为 4 分;

6)事件 D 发生,这时事件 A,B,C 必发生,记为 5 分。

设 $\Omega=\{\bar{A}\cdot\bar{B},A\cdot\bar{B},\bar{A}\cdot B,A\cdot B-C,C-D,D\}$,则 Ω 包含了例 1 解答过程中可能发生的所有结果,由这些所有可能结果组成的集合 Ω 就构成了例 1 的样本空间。

Ω 的意义如图 1 所示。

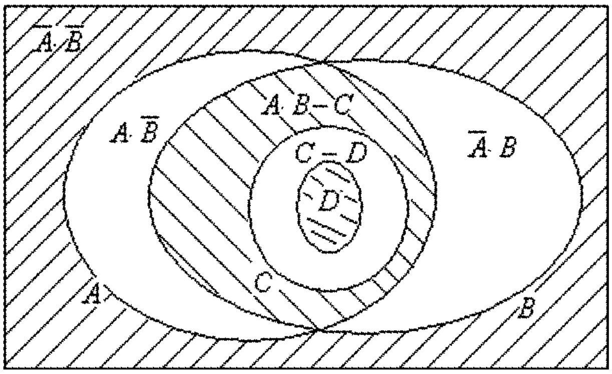

图 1

在图 2·1 中,矩形面积表示样本空间 Ω,四个圆(椭圆)分别表示四个事件 A、B、C、D,这四个事件将问题解答过程划分成 6 个部分,其中的每一个部分也叫做一个样本点,因此例 1 有 6 个样本点。这 6 个样本点的并就构成了例 1 的样本空间 Ω。

在实践中,用事件来表示测验项目的样本空间在操作上显得不够方便,为此可用人们习惯的“分数”来表示样本空间。

在例 1 中,设 $\chi=\{0,1,2,3,4,5\}$,并定义一一映射 $f:\Omega\to\chi$。

对应法则 f 定为 χ 中的每一个“分数”与 Ω 中处于相同位置的事件相对应。

于是在一一映射 f 下,例 1 的样本空间也可用 χ 表示。

从对例 1 的分析中我们看到,节点(或者等价点)是构成测验项目样本空间的基本元素,或者叫做基本事件。一般来说,这些基本事件具有两种基本关系:

1)事件与事件相互独立;2)一事件被另一事件

所包含。

节点之间的这些基本关系确定了测验项目解答过程的一个分划,这个分划将测验项目 的解答过程划分成若干种可能的结果,由这些所有可能结果组成的集合也就构成了测验项目的样本空间。

值得指出的是:

1)在项目解答过程中,如果某个子目标不是节点,那么与此有关的任何一种形式的样本空间都不存在。因为这样的子目标本身并不是一个事件,它不能确定测验项目的一个分划;

2)在项目解答过程中,包含原节点的任何一组节点都可以确定测验项目的一个分划,因此,一个项目的样本空间可以有多种形式。不同的分划,其样本空间也不相同。一般情况下,一个测验项目的节点越多,对样本空间的划分也越细密,这时包含在测验项目中的信息量也越多;

3)在项目解答过程中,有些节点被试可能未写下来,但这并不意味着被试未经过此点,节点是问题解决的必经之点,是绕不过去的。除非被试未能达到目标状态;

4)由同一个目标(或者子目标)状态出发,利用同一组算子达到的两个不同的子目标叫做等价点,例如"等腰三角形两底角相等","等腰三角形底边上的中线、底边上的高、顶角平分线三线合一"是由同一个初始状态:"等腰三角形"出发而得到的两个不同的子目标:"两底角相等"、"底边上的中线、底边上的高、顶角平分线三线合一",这样的两个子目标就叫做等价点,等价点可以看成是同一个节点。

在项目反应理论中,每一个节点的概率函数由二参数逻辑斯蒂模型定义,即

$$p(A_i/\theta) = \frac{1}{1+e^{Da_i(\theta-b_i)}}, i = 1,2\cdots,k$$

其中,$A_i(i=1,2,\cdots,k)$表示项目的第 i 个节点,$p(A_i/\theta)$表示节点(事件)A_i 发生的概率,θ 为被试的能力参数,a_i,b_i 为项目节点参数,D 为常数,通常 D 取 1·7。

于是,利用节点的概率函数以及节点之间的相互关系,就可以导出测验项目的概率函数。

设 X 表示被试在例 1 上的得分,这里,$X\in\chi$。则能力为 θ 的被试在例 1 上得 X 分的概率为

$P(X=0/\theta) = P(\bar{A}\cdot\bar{B}/\theta) = (1-P_1)\cdot(1-P_2)$;

$P(X=1/\theta) = P(A\cdot\bar{B}/\theta) = P_1\cdot(1-P_2)$;

$P(X=2/\theta) = P(\bar{A}\cdot B/\theta) = (1-P_1)\cdot P_2$;

$P(X=3/\theta) = P((A\cdot B-C)/\theta) = P_1\cdot P_2-P_3$;

$P(X=4/\theta) = P((C-D)/\theta) = P_3-P_4$;

$P(X=5/\theta) = P(D/\theta) = P_4$。

能力为 θ 的被试在例 1 上得 X 分的概率也可简单表示为

$$P(X,\theta) = \begin{cases} (1-p_1)\cdot(1-p_2) & X=0 \\ p_1\cdot(1-p_2) & X=1 \\ (1-p_1)\cdot p_2 & X=2 \\ p_1\cdot p_2-p_3 & X=3 \\ p_3-p_4 & X=4 \\ p_4 & X=5 \end{cases} \quad (2\cdot1)$$

其中,$p_i = \dfrac{1}{1+e^{Da_t(\theta-b_t)}}$,$t=1,2,3,4$。

不难验证,$p(\chi/\theta)=1$。

(2·1)式即为例 1 的得分概率函数。

在这里,我们所说的"分数"只是为了表述概率函数的方便而引入的,除此以外,"分数"没有其他的意义。在项目反应理论中,被试的能力水平主要由能力参数 θ 来表述,而不是用"分数"来表述的。不同的"分数"只表明被试在项目解答过程中达到的不同目标,因此,一般情况下,并不是"分数"越高,被试的能力也越高。为方便计,"分数"通常用非负整数表示。在以下的讨论中,"分数"也具有相同的意义。

然而在实践中,人们仍然希望分数的高低与能力参数的大小具有一致性,就是说,如果被试在项目上的得分越高,被试的能力也应当越高。在这种情况下,在例 1 中仅答对第三步(事件 $A\cdot\bar{B}$ 发生)得 1 分和仅答对第五步(事件 $\bar{A}\cdot B$ 发生)却得 2 分就显得不够合理了,因为这两步是独立的。为克服这一缺陷,可规定仅答对第三步或者仅答对第五步均只得 1 分,这时样本空间变形为:

$\Omega = \{\bar{A}\cdot\bar{B},\ A\cdot\bar{B}\cup\bar{A}\cdot B, A\cdot B-C, C-D, D\}$,数集 χ 规定为 $\chi=\{0,1,2,3,4\}$,一一映射 f: $\Omega\rightarrow\chi$ 仍然规定为 χ 中的每一个"分数"与 Ω 中处于相同位置的事件相对应。

在上述规定下,例 1 的得分概率函数变为

$$P(X,\theta)=\begin{cases}(1-P_1)\cdot(1-P_2) & X=0\\ p_1\cdot(1-p_2)+(1-p_1)\cdot p_2 & X=1\\ p_1\cdot p_2-p_3 & X=3\\ p_3-p_4 & X=3\\ p_4 & X=4\end{cases}\quad(2\cdot 2)$$

其中，$p_t=\frac{1}{1+e^{-Da_t(\theta-b_t)}}$，$t=1,2,3,4$。

上述结论可推广到一般情形。

设 $H(A_1,A_2,\cdots,A_k)$ 为定义在算子集 G 上的一个测验项目，$A_1,A_2,\cdots,A_k$ 为 H 的 k 个节点，这 k 个节点或者 k 个事件之间具有两种基本关系：

1）事件与事件相互独立；2）一事件被另一事件所包含。

这些基本关系确定了测验项目解答过程的一个分划，这个分划将测验项目的解答过程划分成 $m(m>1)$ 种可能的结果：$\omega_0,\omega_1,\cdots,\omega_{m-1}$，由这些所有可能结果组成的集合

$$\Omega=\{\omega_0,\omega_1,\cdots,\omega_{m-1}\}$$

也就构成了测验项目的样本空间。

又设 $\chi=\{0,w_1,w_2,\cdots,w_{m-1}\}$，其中，$\{0,w_1,w_2,\cdots,w_{m-1}\}$为不同的正实数，为方便记，$\{0,w_1,w_2,\cdots,w_{m-1}\}$可取为正整数。

定义一一映射 $f:\Omega\to\chi$。

对应法则 f 规定为 χ 中的每一个“分数”与 Ω 中处于相同位置的事件相对应。

在一一映射 f 下，测验项目 H 的样本空间也可用一个实数集 χ 表示。

每一个节点或每一个基本事件的概率函数由二参数逻辑斯蒂模型定义，即

$$p(A_i/\theta)=\frac{1}{1+e^{-Da_i(\theta-b_i)}},i=1,2,\cdots,k。$$

对任意 $X\in\chi$，利用节点的概率函数及其节点之间的相互关系不难算得 X 的概率函数 $p(X,a,b,\theta)$，并且 $p(\chi,a,b,\theta)=1$。

关于测验项目的概率函数 $p(X,a,b,\theta)$，作如下说明：

1）在概率函数 $p(X,a,b,\theta)$ 中，a,b 分别为 k 维向量，即 $a\in(R^+)^k$，$b\in R^k$。就是说，a,b 的维数等于节点的个数。而 θ 为一维变量，即 $\theta\in R$；

2）概率函数 $p(X,a,b,\theta)$ 是一个抽象的函数，它的具体表达式依赖于测验项目的节点个数以及节点之间的相互关系。因此对于不同的测验项目，其概率函数的表达式也是不相同的；

3）在定义了节点的概率函数后，一个测验项目的概率函数 $p(X,a,b,\theta)$ 由节点唯一确定。

3 参数估计

下面我们以例 1 为出发点来讨论多级评分项目的参数估计。

我们首先讨论在能力参数已知的情况下项目参数的估计。

注意到例 1 由四个节点组成，因此要求出例 1 的项目参数，只需求出每一个节点的参数就可以了。

在例 1 中我们不难发现，当 $X\in\{1,3,4,5\}$ 时，事件 A 发生，而当 $X\in\{0,2\}$ 时，事件 A 不发生。定义事件 A 的示性函数：$u_1=\begin{cases}1 & X\in\{1,3,4,5\}\\ 0 & X\in\{0,2\}\end{cases}$。于是根据示性函数 u_1 的定义，当 $u_1=1$ 时，事件 A 发生，而当 $u_1=0$ 时，事件 A 不发生。

同理可以定义事件 B、C、D 发生的示性函数分别为

$$u_2=\begin{cases}1 & X\in\{2,3,4,5\}\\ 0 & X\in\{0,1\}\end{cases},u_3=\begin{cases}1 & X\in\{4,5\}\\ 0 & X\in\{0,1,2,3\}\end{cases},$$

$$u_4=\begin{cases}1 & X\in\{5\}\\ 0 & X\in\{0,1,2,3,4\}\end{cases}。$$

假设 $X_1,X_2,\cdots,X_N$ 为能力值分别为 $\theta_1,\theta_2,\cdots,\theta_N$ 的被试在例 1 上的样本观察值，这里，$\theta_1,\theta_2,\cdots,\theta_N$ 已知。于是利用示性函数 u_1,u_2,u_3,u_4 的定义，可将 $X_1,X_2,\cdots,X_N$ 全部转换成 u 值：

$u_t(X_1),u(X_2),\cdots,u_t(X_N),t=1,2,3,4$

利用 $u_t(X_1),u_t(X_2),\cdots,u_t(X_N),t=1,2,3,4$，就可得到每一个节点的似然函数

$$L_t=\prod_{j=1}^{N}p_t^{u_t}\cdot(1-p_t)^{1-u_t},t=1,2,3,4\quad(3\cdot 1)$$

其中，$p_t=\frac{1}{1+e^{-Da_t(\theta_j-b_t)}}$，$t=1,2,3,4$。

对（3·1）式两边取对数，分别对 a,b 求导，除以 N 后令其等于零，得

$$\left\{\begin{aligned}&\frac{1}{N}\frac{\partial lnL_t}{\partial a_t}=0\\ &\frac{1}{N}\frac{\partial lnL_t}{\partial b_t}=0\quad t=1,2,3,4\end{aligned}\right\}\quad(3\cdot 2)$$

方程组(3·2)可分成四个独立的方程组：

1) $\begin{cases}\dfrac{1}{N}\dfrac{\partial lnL_1}{\partial a_1}=0\\ \dfrac{1}{N}\dfrac{\partial lnL_1}{\partial b_1}=0\end{cases}$；　2) $\begin{cases}\dfrac{1}{N}\dfrac{\partial lnL_2}{\partial a_2}=0\\ \dfrac{1}{N}\dfrac{\partial lnL_2}{\partial b_2}=0\end{cases}$；

3) $\begin{cases}\dfrac{1}{N}\dfrac{\partial lnL_3}{\partial a_3}=0\\ \dfrac{1}{N}\dfrac{\partial lnL_3}{\partial b_3}=0\end{cases}$；　4) $\begin{cases}\dfrac{1}{N}\dfrac{\partial lnL_4}{\partial a_4}=0\\ \dfrac{1}{N}\dfrac{\partial lnL_4}{\partial b_4}=0\end{cases}$

解上述 4 个方程组，就可分别解出四个节点的参数。

一般情况下，如果一个测验项目有 k 个节点，采用上述方法，就可得到 k 个方程组，分别解这 k 个方程组，就可求出全部的项目参数。

又假设一个测验由 n 个测验项目：$H_1, H_2, \cdots, H_n$ 组成，其中，$\mathrm{H_i}$ 的样本空间为 $\chi_i=\{0, w_1, w_2, \cdots, w_{m-1}\}$，$(i=1,2,\cdots,n)$，而 $X_1, X_2, \cdots, X_n$ 为能力为 θ_0 的被试在测验中的一组样本观察值，这里，$X_i \in \chi_i$，$X_1, X_2, \cdots, X_n$ 相互独立。且 X_i 具有概率函数 $p_i=(X_i, \theta, a_i, b_i)$，$i=1,2,\cdots,n$，项目参数 a_i, b_i 已知。

则随机变量 $X_1, X_2, \cdots, X_n$ 的似然函数为

$$L(X_1, X_2, \cdots, X_n, \theta)=\prod_{i=1}^{n} p_i(X_i, \theta, a_i, b_i) \quad (3\cdot3)$$

上式两边取对数，对 θ 求导，除以 n 后令其等于零，便得到

$$\frac{1}{n}\frac{\partial LnL}{\partial\theta}=0 \quad (3\cdot4)$$

解方程(3·4)，求出使(3·3)取得极大值的 $\hat{\theta}$，即可作为 θ_0 的极大似然估计值。

在实践中，通常项目参数与能力参数都是未知的，需要同时进行估计，在这种情况下，可采用 lord 的两阶段似然估计法来同时估计项目参数与能力参数，其具体做法是这样的：

假设一个测验由 n 个测验项目所组成，共有 N 个被试参加了测验，

1）首先将每一个被试在测验中的得分化为标准分，以此作为被试的能力初值，然后按照上面讨论的方法，逐一计算出每一个测验项目的节点参数；

2）将计算出的项目参数代入似然方程(3·4)，逐一计算出每一个被试的能力参数；

3）将算得的能力参数标准化然后重新计算项目参数，

这样反复迭代，直至项目参数与能力参数不再改进为止，这时得到的项目参数与能力参数就可以作为项目参数与能力参数的估计值。

在以上的参数估计法中，并未对参数作任何额外的限制，因此，当被试样本与项目样本的个数趋近于无穷大时，采用上述方法求得的项目参数与能力参数将依概率收敛于真值。

4　信息函数

下面我们来讨论多级评分项目的信息函数。

由数理统计理论知，在项目参数已知的情况下，能力参数的极大似然估计具有下述重要性质：

定理 1　假设

1）一个测验由 n 个测验项目：$H_1, H_2, \cdots, H_n$ 组成，其中，H_i 的概率空间为 $\chi_i=\{0, w_1, w_2, \cdots, w_{m-1}\}$，$(i=1,2,\cdots,n)$，而 $X_1, X_2, \cdots, X_n$ 为能力为 θ_0 的被试在测验中的一组样本观察值，这里，$X_i \in \chi$，$X_1, X_2, \cdots, X_n$ 相互独立。且 X_i 具有概率函数 $p_i(X_i, a_i, b_i, \theta)$，$i=1,2,\cdots,n$，项目参数 a_i, b_i 已知。

2）$\dfrac{\partial p_i(X_i,\theta)}{\partial\theta}$，$\dfrac{\partial^2 p_i(X,\theta)}{\partial\theta^2}$存在，且

$\left|\dfrac{\partial p_i(X_i,\theta)}{\partial\theta}\right|<M$（M 为常数），$i=1,2,\cdots,n$；

3）存在一充分小的正数 δ_0，使得对任意 $\theta\in\Theta$，$p_i(X_i,\theta)>\delta_0$，$E\left(\dfrac{\partial lnp_i(X_i,\theta)}{\partial\theta}\right)^2>\delta_0$；

4）积分与微分可交换，

则在 θ 的未知真值 θ_0 为 Θ 内点的情况下，当$n\to\infty$ 时，

①似然方程(3·4)以概率 1 存在唯一解 $\hat{\theta}$；

②$\hat{\theta}$ 是 θ_0 的相合估计；

③$\hat{\theta}\overset{L}{\backsim}(\theta_0, 1/I(\theta_0))$。

其中，

$$I(\theta)=\sum_{i=1}^{n}\int_{\chi i}\frac{p_i'(X_i, a_i, b_i, \theta))^2}{p_i}d\mu=\sum_{i=1}^{n}I_i(\theta) \quad (4\cdot1)$$

$$I_i(\theta)=\int_{\chi i}\frac{p_i'(X_i, a_i, b_i, \theta))^2}{p_i}d\mu。$$

从 $I(\theta)$ 的表达式中我们不难看出，$I(\theta)$ 越大，$D(\hat{\theta})$ 便越小，从而估计值 $\hat{\theta}$ 也就越精确。

我们把 $I(\theta)$ 叫做测验信息函数。而把 $I(\theta)$ 叫做项目信息函数。

如果知道了随机变量 X_i 的概率分布列，不难得出项目信息函数的计算公式。

设 H_i 为算子集 G 上的一个测验项目，具有概率空间(χ_i, B_i, P_i)，又设随机变量 $X_i \in \chi_i$，且具有概率分布列

X_i	0	w_{i1}	w_{i2}	…	w_{ik_i}
p_i	p_{i0}	p_{i1}	p_{i2}	…	p_{iki}

则 H_i 的项目信息函数为

$$I_i(\theta) = \sum_{t=0}^{k_i} \frac{p'_{it}}{p_{it}} \qquad (4 \cdot 2)$$

(4·2)式就是项目信息函数的一般计算公式。

利用(4·2)式,只要知道了一个测验项目的概率分布列,就可以很方便的计算出测验项目的信息函数。

下面我们来讨论几个多级评分项目信息函数的例子。

例 2. 化简:$\sqrt{3-2\sqrt{2}}-\sqrt{2}-1$。

我们首先来解答这个问题。

解:$\sqrt{3-2\sqrt{2}}-\sqrt{2}-1$。

A_1: $=\sqrt{(\sqrt{2}-1)^2}-\sqrt{2}-1$ 1 分

A_2: $=(\sqrt{2}-1)-\sqrt{2}-1$ 2 分

A_3: $=-2$。 3 分

由项目的解答过程不难发现,事件 A_1,A_2,A_3 之间具有如下的包含关系:$A_1 \supset A_2 \supset A_3$,这些包含关系将问题解决过程划分成如下四个部份:

$\overline{A}_1, A_1-A_2, A_2-A_3, A_3$。

令 $\Omega=\{\overline{A}_1, A_1-A_2, A_2-A_3, A_3\}$,$\chi=\{0,1,2,3\}$,

并定义一一映射 $f:\Omega \to \chi$。

对应法则 f 规定为 χ 中的每一个"分数"与 Ω 中处于相同位置的事件相对应。

设随机变量 $x \in \{0,1,2,3\}$,则 x 的分布列为

X	0	1	2	3
p	$1-p_1$	p_1-p_2	p_2-p_3	p_3

于是,由信息函数的一般计算公式(4·2)式,可得例 2 的项目信息函数为

$$I_i(\theta) = \frac{p_1'^2}{1-p_1} + \frac{(p_1'-p_2')^2}{p_1-p_2} + \frac{(p_2'-p_3')^2}{p_2-p_3} + \frac{p_3'^2}{p_3}。$$

这里,$p_t = \dfrac{1}{1+e^{-Da_t(\theta-b_t)}}$,$t=1,2,3$。

上述结果可推广到一般情形。

设 H 为定义在算子集 G 上的一个测验项目,$A_1,A_2,A_3,\cdots,A_k$ 为 H 的 k 个节点,这些节点具有包含关系:

$A_1 \supset A_2 \supset A_3, \cdots, \supset A_k$,

这些包含关系将 H 的问题解决过程划分成如下几种可能的结果:

$\overline{A}_1, A_1-A_2, A_2-A_3, \cdots, A_k$,

定义一一映射 $f:\{\overline{A}_1, A_1-A_2, A_2-A_3, \cdots, A_k\} \to \{0,1,2,\cdots,k\}$,又设 $X \in \{0,1,2,\cdots,k\}$,则 X 的分布列为

X	0	1	2	…	k
p	$1-p_1$	p_1-p_2	p_2-p_3	…	p_k

H 的项目信息函数为

$$I_i(\theta) = \frac{p_1'^2}{1-p_1} + \frac{(p_1'-p_2')^2}{p_1-p_2} + \cdots + \frac{p_k'^2}{p_k} \qquad (4 \cdot 3)$$

这里,$p_t = \dfrac{1}{1+e^{-Da_t(\theta-b_t)}}$,$t=1,2,\cdots,k$。

(4·3)式即为等级评分模型项目信息函数的计算公式。

在(4·3)式中,如果 $t=1$,这时(4·3)式变为

$$I_i(\theta) = \frac{p_1'^2}{1-p_1} + \frac{p_1'^2}{p_1} = \frac{p_1'^2}{p_1 q_1} (q_1 = 1-p_1)。$$

这正是二级评分项目的信息函数。

下面我们通过几个实际的例子,来讨论多级评分项目的信息函数在实践中的应用,这几个实例是从一次实际测验中得到的*,该次测验共有 24 个测验项目,有 1226 名被试参加了测验。

例 3·解方程 $\sqrt{x+10}+2=x$

首先我们来解这个方程。

解:$\sqrt{x+4}=x-2$

$(x+4)^2=(x-2)^2$

A_1:$x^2-5x-6=0$ 1 分

A_2:解得 $x_1=6$,$x_2=-1$ 2 分

经检验 $x_2=-1$ 是增根,舍去,

A_3:∴ 原方程的根为 $x=6$ 3 分

在例 3 的解答过程中,有三个主要的节点:A_1,A_2,A_3,且 A_1,A_2,A_3 之间具有如下的包含关系:$A_1 \supset A_2 \supset A_3$。

例 3 的项目参数估计值分别为:

$a_1=1.46$, $a_2=1.60$, $a_3=1.48$;

$b_1=-0.37$, $b_2=1.37$, $b_3=1.78$。

该项目的特征曲线如图 4·1。

在图 2a 中,从左至右依次为三个节点的概率函数曲线。它是依次将项目节点的参数估计值分别代入二参数逻辑斯蒂模型后得到的。

而图 2b 中所示的曲线是被试在项目上分别得 0 分、1 分、2 分、3 分的概率曲线,其中,$p(X=0/\theta)=1-p_1$;$p(X=1/\theta)=p_1-p_2$;

$p(X=2/\theta)=p_2-p_3$;$p(X=3/\theta)=p_3$。

这里，$p_t=\dfrac{1}{1+e^{-Da_t(\theta-b_t)}}$，$t=1,2,3$。

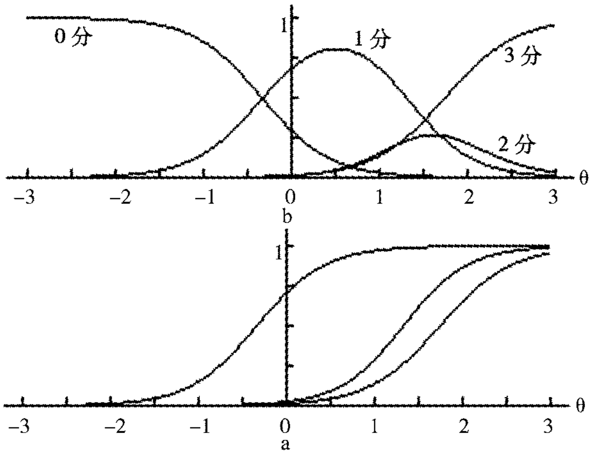

图　2

从该项目的难度参数估计值中我们看到，该项目从第一步到第二步的跨度较大，$b_2-b_1=1.37+0.37=1.74$，而第二步与第三步的跨度较小，$b_3-b_2=1.78-1.37=0.41$，因此，答对第一步的被试，有相当一部份不能在第二步上作出正确反应，而答对第二步的被试，大部分都能够正确回答第三步。

从图 2b 中的得分概率曲线中我们看到，当被试的能力参数 $\theta<-0.5$ 时，被试在项目上得 0 分的概率最大，而得 1 分、2 分、3 分的概率几乎为零。随着被试能力的提高，被试在项目上得零分的概率开始降低，而得其它分数的概率增大，在区间$(-0.5,1.5)$内，被试得 1 分的概率最大，而得 0 分、2 分、3 分的概率相对较小。当被试的能力参数继续增大时，被试得 3 分的概率开始增大，而得其它分数的概率开始减小。当 $\theta>1.5$ 时，被试得 3 分的概率最大。

由信息函数的一般计算公式(4・2)式，可得到例 3 的项目信息函数为

$$I_i(\theta)=\frac{p_1'^2}{1-p_1}+\frac{(p_1'-p_2')^2}{p_1-p_2}+\frac{(p_2'-p_3')^2}{p_2-p_3}+\frac{p_3'^2}{p_3}。$$

这里，$p_t=\dfrac{1}{1+e^{-Da_t(\theta-b_t)}}$，$t=1,2,3$。

例 3 的信息函数如图 3 中的上曲线。

从图 3 中我们看到，该项目对能力值位于$(-1,2.5)$这一区间的被试，提供的信息量较大，而在这一区间外，该项目提供的信息量很少。

从图 3 中我们看到，例 3 的项目信息函数呈双峰形，在 $\theta=-0.5$ 及 $\theta=1.5$ 附近，该项目的信息量分别达到两个不同的峰值。在这两个峰值之间，呈现一个凹区间。因此，在能力值 $\theta=-0.5$ 及 $\theta=1.5$ 附近，该项目提供了较大的信息量，而在两个峰值之间，项目提供的信息量相对较少。

如果将例 3 编成二级评分项目，如填空题，则其参数估计值就是原节点的参数，即

$a=a_3=1.48$，$b=b_3=1.78$。

其信息函数的计算公式为

$$I_i(\theta)=\frac{p'^2}{pq}(q=1-p)$$

它的信息函数曲线如图 3 中的下曲线。

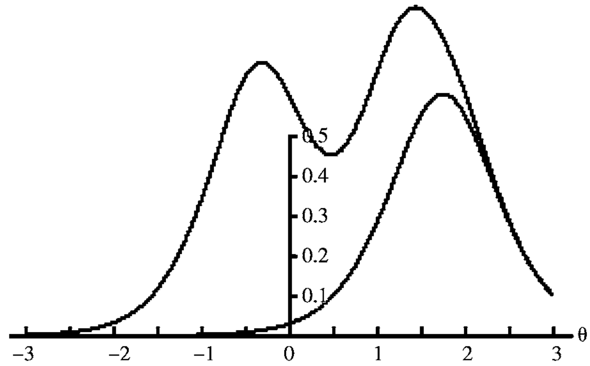

图　3

从图 3 中我们看到，当将项目编制成二级评分项目后，它只在 $\theta=1.78$ 附近提供了较大的信息，而在其它地方，它所提供的信息量很少。特别当 $\theta<0.5$ 时，它所提供的信息量几乎为零。因此该项目对能力参数小于 0.5 的被试几乎不提供任何信息。一个测验项目的信息总量可用信息函数曲线与 θ 轴所围的图形面积来表示。从图 3 中两条曲线与 θ 轴所围图形面积的对比中我们可以清楚地看到，当把例 3 转化成一道二级评分试题时，有大量的信息被丢失了。

由此可见，一道编制得好的多级评分试题，不仅比二级评分项目提供的信息量要大，而且它所提供的信息量的范围也比二级评分项目要宽。

例 3 的马鞍形曲线值得引起注意。这一情况表明，一个测验项目各个节点之间的难度参数的跨度不应过大，如果难度跨度过大，会在两个难度之间出现一个凹区间，从而降低能力位于这两个难度之间的信息。

我们再来看一个例子。

例 4・填空：满足不等式 $x^2+3x+2\geqslant 0$ 的解集为 ＿＿＿＿＿＿，满足不等式 $log_2(x^2-x-4)\leqslant 1$ 的解集为＿＿＿＿＿＿，同时满足这两个不等式的所有整数解为 ＿＿＿＿＿＿。

例 4 的第一个空填 $A: x \leq -2$ 或者 $x \geq -1$，

第二个空填 $B: -2 \leq x < \frac{1-\sqrt{17}}{2}$ 或者 $\frac{1+\sqrt{17}}{2} < x \leq 3$，

第三个空填 $C: \{-2, 3\}$。

从项目的解答过程中我们不难发现，A, B, C 三者之间具有下述关系：

1）A, B 相互独立；2）$A \cap B \supset C$。

A, B, C 之间的这些关系将问题解决过程划分为如下几个部分：

$\bar{A} \cap \bar{B}, A \cap \bar{B}, \bar{A} \cap B, A \cap B - C, C$，

定义一一映射 $f: \{\bar{A} \cap \bar{B}, A \cap \bar{B}, \bar{A} \cap B, A \cap B - C, C\} \to \{0, 1, 2, 3, 4\}$，设 $x \epsilon \{0, 1, 2, 3, 4\}$，则不难得到 x 的分布列为

X	0	1	2	3	4
p_i	$(1-p_1)(1-p_2)$	$p_1(1-p_2)$	$(1-p_1)p_2$	$p_1p_2-p_3$	p_3

于是，由信息函数的通用计算公式（4·2），我们不难得到例 4 的信息函数为

$$I_i(\theta) = \frac{[(1-p_1)(1-p_2)]'^2}{(1-p_1)(1-p_2)} + \frac{[p_1(1-p_2)]'^2}{p_1(1-p_2)} + \frac{[(1-p_1)p_2]'^2}{(1-p_1)p_2} + \frac{(p_1p_2-p_3)'^2}{p_1p_2-p_3} + \frac{p'^2_3}{p_3}。$$

这里，$p_t = \frac{1}{1+e^{-Da_t(\theta-b_t)}}, t = 1, 2, 3$。

例 4 的信息函数图象如图 4 中的上曲线。

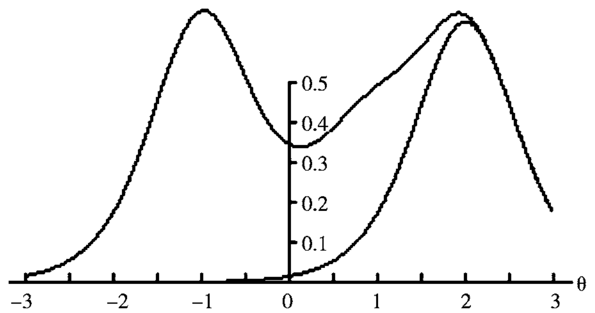

图 4

其中，各节点的参数值分别为：

$a_1 = 1.5, a_2 = 1.2, a_3 = 1.5$，

$b_1 = -1, b_2 = 1, b_3 = 2$。

从图 4 中我们看到，在区间（$-1.5, 2.5$）内，该项目提供了较大的信息，而在这一区间外，该项目提供的信息量较少。

注意到图 4 呈双峰形，在能力值分别为 -1 和 2 附近，该项目的信息量分别达到两个不同的峰值。而在能力值为 -1 和 2 之间出现了一个凹区间。因此，在区间（$-1, 2$）内，项目提供的信息量相对要少一些。

如果将这一项目编制成一个二级评分项目，则例 4 变为"同时满足不等式 $x^2+3x+2 \geq 0$ 和 $log_2(x^2-x-4) \leq 1$ 的所有整数解为__________"它的参数值等于例 4 中第三个节点 C 的参数，即 $a = a_3 = 1.5, b = b_3 = 2$。

其信息函数如图 4 中的下曲线。

其信息函数的计算公式为 $I_i(\theta) = \frac{p'^2}{pq}(q = 1-p)$。

从图 4 中我们看到，当将这一测验项目编制成一个二级评分项目的时候，该项目只在 $b_3 = 2$ 附近提供了较大的能力信息，在距 $b_3 = 2$ 较远的地方，该项目提供的能力信息很少。

综合上述例子的讨论我们不难发现，与二级评分项目相比，多级评分项目的优越性是十分明显的，一道编制得好的多级评分项目，它所提供的信息量远比二级评分项目要多，而且多级评分项目提供的信息量的范围通常也比二级评分项目要宽。

那么，多级评分项目是否一定要比二级评分项目好呢？为了回答这一问题，我们再来看另外一个例子。

例 5. 计算：$[5^3 \times (-\frac{1}{5}) - 3] \div (-4)$

解：原式

$A_1: = [-25-3] \div (-4)$ （1 分）

$A_2: = (-28) \div (-4)$ （2 分）

$A_3: = 7$ （3 分）

在例 5 的解答过程中，有三个主要的节点：A_1, A_2, A_3，且 A_1, A_2, A_3 之间具有如下的包含关系：$A_1 \supset A_2 \supset A_3$，这些包含关系将问题解决过程划分成如下四个部份：

$\bar{A}_1, A_1 - A_2, A_2 - A_3, A_3$。

各节点的参数估计值分别为

$a_1 = 0.66, a_2 = 0.65, a_3 = 0.69$；

$b_1 = -1.18, b_2 = -0.93, b_3 = -0.72$。

各节点的项目特征曲线如图 5。

从图 5 中我们看到，考生在项目各个节点的正

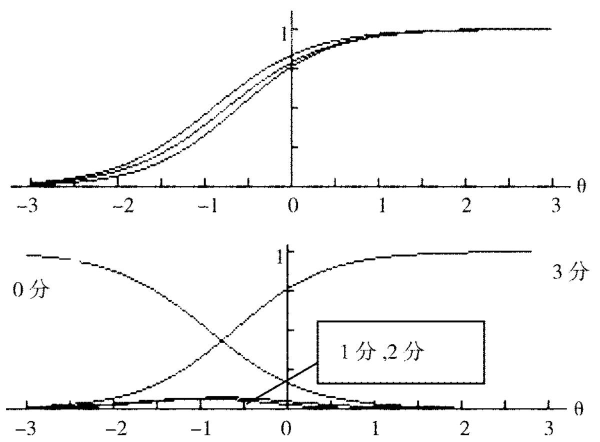

图　5

确反应概率差异很小,在第一步上答对的考生,几乎都能答对第三步。考生在项目上要么得零分,要么得满分 3 分,在中间步骤上的得分概率很低。

例 5 的信息函数计算公式与例 3 相同,信息函数曲线如图 6 中的上曲线。

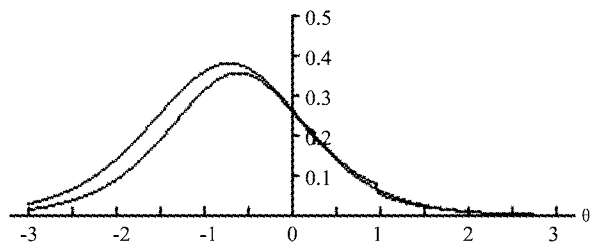

图　6

从图 6 中我们看到,在区间($-1.5, 0.5$)内,该项目提供了较大信息,其峰值出现在 $\theta = -0.7$ 附近,在其余地方,该项目提供的信息很少。

若将例 5 变为一道二级评分试题(如填空题),则其对应的项目参数分别为 $a = a_3 = 0.69$, $b = b_3 = -0.72$,。其信息函数曲线如图 6 中的下曲线。

从图 6 中两条曲线的对比中我们看到,两条曲线的差异很小,二者几乎重合。因此,从信息函数的角度来说,不能认为这是一道编制得很好的多级评分试题,因为它所提供的信息量与一道二级评分项目相比几乎没有什么差别。

例 5 的结果表明,我们不能笼统地说多级评分项目要比二级评分项目好,多级评分项目优于二级评分项目是有条件的。只有当一个多级评分项目各节点之间的难度参数差异足够大时(但也不能相差太大),多级评分项目才能充分体现出它的优越性。如果一个多级评分项目各节点之间的难度参数过于接近,那么这样的多级评分项目与一道二级评分项目相比几乎没有什么差别。这样的项目与其编制成多级评分项目,到不如将其编制成二级评分项目更方便一些。

5　结语

从以上的讨论中我们看到:适当提高测验项目各节点的 a 参数;适当加大各节点的难度参数;适当增加测验项目的节点数,均可有效地提高测验项目的信息量,从而可有效地提高整个测验的估计精度。这在测验项目有限的情况下,是很有实践意义的。项目信息函数是项目反应理论的一个十分有力的工具,它在各类测验中都有广泛的用途。利用这一工具,我们可以精心设计出一个测验,这个测验就象是一把精确的“尺子”,利用这把“尺子”,就可以对被试的“能力参数”进行测量,并且可以达到任意指定的精度。

致谢:本文在写作过程中,审稿人给笔者提出了许多很好的修改意见,谨此致谢。

参　考　文　献

1　Hilin C L, Drasgow F, Parsons C K. Item Response Theory Application to Psychological Measurement (in Chinese). Hubei education publishing house, 1990

2　Yu J Y. Item Response Theory Application (in Chinese). Nanjing: Jiangshu Education Publishing House, 1992
(余嘉元. 项目反应理论及其应用. 南京:江苏教育出版社, 1992)

3　Qi S Q, Dai H Q. Item Response Theory (in Chinese). Jiangxi College Publishing House, 1992
(漆书青,戴海畸. 项目反应理论. 南昌:江西高校出版社, 1992)

Information of IRT Multilevel Item

Du Wenjiu
(*College of Mathmatics and Finance Ecomomics ,Southwest University, Chongqing ,400715*)

Abstract

The objective was to give out calculate formula of multilevel item information function, and give some examples of multilevel item information. Results was: 1) at first discuss the sample sky of the test item by a example; 2) give out the probability function and information function of multilevel item; 3) discuss the application of multilevel item information function in practice by a few examples.
Key words IRT, information, multilevel item.

第三届全国心理学学术刊物研讨会纪要

第三届全国心理学学术刊物研讨会于2005年11月4日~8日在浙江宁波召开。《心理学报》、《心理科学》、《心理发展与教育》、《心理科学进展》、《心理学探新》、《应用心理学》、《心理与行为研究》等国内多家专业心理学学术刊物的负责同志出席会议。会议由天津师范大学心理与行为研究中心、《心理与行为研究》编辑部主办,宁波大学承办。

国务院学位委员会心理学科评议组召集人、教育部人文社会科学重点研究基地——天津师范大学心理与行为研究中心主任沈德立教授,中国心理学会原理事长、《心理学报》主编、中国科学院心理研究所陈永明研究员,国务院学位委员会心理学科评议组成员、中国心理学会原副理事长、《心理科学》主编、华东师范大学杨治良教授,国务院学位委员会心理学科评议组成员、中国科学院心理研究所所长、《心理科学进展》主编杨玉芳研究员,国务院学位委员会心理学科评议组成员、中国心理学会副理事长、《应用心理学》主编、浙江大学沈模卫教授,《心理发展与教育》副主编、北京师范大学方晓义教授,《心理学探新》主编、江西师范大学胡竹箐教授,中国人民大学复印报刊资料《心理学》责编解军副编审,宁波大学副校长叶飞帆教授以及宁波大学师范学院的领导和教师等与会。

开幕式由天津师范大学心理与行为研究中心副主任、《心理与行为研究》编辑部主任白学军教授主持。叶飞帆教授代表宁波大学党政领导和全体师生员工对与会的各家专业心理学学术刊物的负责同志表示热烈欢迎,他简要介绍了宁波大学的发展及办学情况。沈德立教授对全国专业心理学学术刊物的发展现状进行了总结,提出了今后的办刊方向。指出中国心理学学术期刊在我国心理学事业快速发展的时期,应巩固和发展自己的优势,进一步提高质量:要加强学术期刊的规范化建设;要注重刊物的特色,强化精品意识;要兼顾学术性和普及性;要注重专业期刊发展的国际化。

会议期间,七家刊物负责同志介绍了两年来各自的办刊经验,并对以下四个方面的问题进行了研讨,这四个问题是:第一,中国心理学学术刊物质量的保证与提高;第二,中国心理学学术刊物的办刊特色;第三,中国心理学学术刊物编辑与出版的现代化;第四,中国心理学学术刊物的国际化。

为了繁荣我国心理学事业和树立良好的学风,七家刊物达成了以下两点共识:第一,从2006年开始,如果一篇文章有多位作者,投稿必须提交每位作者的亲笔签名,以便使每位作者对文章负责;第二,从2006年起,将对一稿多投现象进行严肃处理,即一旦发现后,七家刊物将互通信息,并在两年内不再刊发其文章。