



# [12] 发明专利申请公开说明书

[21] 申请号 02812200.3

[43] 公开日 2004 年 7 月 28 日

[11] 公开号 CN 1516859A

[22] 申请日 2002.4.19 [21] 申请号 02812200.3

[30] 优先权

[32] 2001.4.20 [33] US [31] 09/838,129

[86] 国际申请 PCT/US2002/012424 2002.4.19

[87] 国际公布 WO2002/086841 英 2002.10.31

[85] 进入国家阶段日期 2003.12.18

[71] 申请人 教育考试服务中心

地址 美国新泽西州

[72] 发明人 威廉·F·斯托特 萨拉·M·哈茨

[74] 专利代理机构 北京同立钧成知识产权代理有限公司

代理人 刘芳

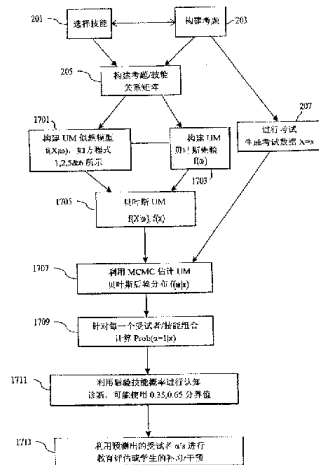
代理人 刘芳

权利要求书 22 页 说明书 37 页 附图 18 页

[54] 发明名称 潜在属性的诊断方法

[57] 摘要

一种方法，可以用于：对智力技巧进行认知诊断，对病症和疾病进行医学和精神病学诊断，以及一般来讲，对一组对象(通常是人)的潜在属性进行诊断(1711)，其中，可以得到关于所述对象的、有两种取值的(二值性)多项信息，例如，采用对/错方式进行记分的考题的考试。可以应用本发明的场景包括但不限于：各种水平的课堂，基于网络的教学，公司内部培训，大规模标准化考试，以及医学和精神病学场景。用途(1713)包括但不限于：个体学员的反馈，学员的补习，群体水平的教育评估，以及医学和精神病学治疗。



## 1. 一种方法，包括：

构建一项其中包括若干考题的考试，并选择一个技能集合，所述技能集合的设计目的是：测定参加所述考试的受试者的能力水平，并测定每一个受试者已经掌握还是尚未掌握所述技能；

构建一个以数学方式表达的模型，包括所述考题和所述选定的技能，所述选定的技能是一个更大的技能组的子集，所述更大的技能组中的技能可以影响受试者的考试表现，所述更大的技能组中未规定的剩余部分，通过剩余能力参数的形式，在所述模型中得到体现；所述模型中包括参数，所述参数描述了所述考题如何依赖于所述选定的技能集合，也描述了所述考题如何依赖于所述剩余能力参数，其中，受试者对考题的反应提供了针对每一个参数的估计信息，使得可以对所述参数进行标定，也提供了对所述受试者掌握还是未掌握哪些技能的预测；所述模型还解释了如下事件的概率，即：对每一道考题而言，每一名受试者可能已掌握了所述考题所要求的全部技能，但未能在所述考题上正确地运用这些必需的、已经掌握的技能中的至少一项，因此答错了这道题，其中，所述考题所要求的全部技能是所述选定的技能集合的一个子集中的；以及如下事件的概率：对每一道考题而言，每一名受试者可能未掌握所述考题所要求的技能中的至少一项，但仍然在所述考题上正确地运用了这些必需、但没有掌握的技能中的每一项，同时，也在所述考题上正确地运用了剩余的、必需、同时也掌握了、选自所述选定的技能集合中的技能，从而答对了所述考题；所述模型将每种技能的掌握定义为一个设定的水平，表示一名超过所述水平的受试者已经具备了相应的技能水平；所述模型针对所述选定的技能集合中的技能对，表达了在每一个所述技能对的两个成员之间的正联系，还表达了对每一个技能对正联系大小的估量，其中，针对每一个技能对的所述正联系可以由受试者对具体考题的反应估计出来；以及

将根据受试者反应所获得的考试结果用于标定所述模型的具体考题，以得到技能已掌握的预测，或者技能未掌握的预测，或者针对每一个由具体受试者和具体技能构成的组合不做出任何预测。

## 2. 如权利要求 1 所述的方法，包括：

构建一项其中包括若干考题的考试，当受试者  $j$  答错或答对考题  $i$  时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个技能集合  $\{\alpha_{jk}\}$ ，当受试者  $j$  未掌握或已掌握技能  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，包括可识别的、因此能够被标定的参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述考题如何依赖于所述选定的技能集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{i1}} \times (r_{i2}^*)^{1-\alpha_{i2}} \times \dots \times (r_{im}^*)^{1-\alpha_{im}}$$

其中， $S_{ij}$  是正确运用全部必需技能的概率，由受试者掌握及未掌握这些所需技能所确定；在

$S_{ij}$  中针对考题/技能关系矩阵中所规定的、考题  $i$  所要求的  $m$  项技能对  $r^*$ 's 求积;  $\pi^*_i = \prod (\pi_{ik})$ , 对  $k$  求积;  $r^*_{ik} = r_{ik} / \pi_{ik}$ , 其中,  $r_{ik} = \text{Prob}$  (在受试者未掌握技能  $k$  的情况下, 将技能  $k$  正确应用于考题  $i$ ),  $\pi_{ik} = \text{Prob}$  (在受试者已掌握技能  $k$  的情况下, 将技能  $k$  正确应用于考题  $i$ ); 间接表达了对每一个技能对正联系大小的估量, 正如二值技能  $\alpha_k, \alpha_{k'}$  的连续二变量正态技能前身  $\alpha'_k, \alpha'_{k'}$  之间的相关性  $\sigma_{kk'}$  所确定的那样, 然后, 所需的二值技能对  $\alpha_k, \alpha_{k'}$  可以通过如下方式产生: 将每一个  $\alpha'_k$  在指定的掌握水平分界点处进行分界, 从而使得技能  $k$  的掌握概率  $p_k(\alpha_k = 1)$  被定义为等于  $\text{Prob}(\alpha'_k \geq \text{分界点})$ , 其中, 这样定义的技能掌握水平是通过选择  $\alpha'_k$  的分界点而实现的,  $p_k$  是由使用者确定的、被判断为已掌握技能  $k$  的受试者比例, 从而使得每一项技能通过连续的  $\alpha'_k$  及其分界点而获得了被赋予的水平, 表示一名超过该水平的受试者已经掌握该项技能。

### 3. 一种方法, 包括:

构建一项其中包括若干考题的考试, 并选择一个技能集合, 所述技能集合的设计目的是: 测定参加所述考试的受试者的能力水平, 并测定每一个受试者已经掌握还是尚未掌握所述技能;

构建一个以数学方式表达的模型, 包括所述考题和所述选定的技能, 所述选定的技能是一个更大的技能组的子集, 所述更大的技能组中的技能可以影响受试者的考试表现, 所述更大的技能组中未规定的剩余部分, 通过剩余能力参数的形式, 在所述模型中被体现; 所述模型中包括参数, 所述参数描述了所述考题如何依赖于所述选定的技能集合, 也描述了所述考题如何依赖于所述剩余能力参数, 其中, 受试者对考题的反应提供了针对每一个参数的估计信息, 使得可以对所述参数进行标定, 也提供了对所述受试者掌握还是未掌握哪些技能的预测; 所述模型还解释了如下事件的概率, 即: 对每一道考题而言, 每一名受试者可能已掌握了所述考题所要求的全部技能, 但未能在所述考题上正确地运用这些必需的、已经掌握的技能中的至少一项, 因此答错了这道题, 其中, 所述考题所要求的全部技能是所述选定的技能集合的一个子集中的; 以及如下事件的概率: 对每一道考题而言, 每一名受试者可能未掌握所述考题所要求的技能中的至少一项, 但仍然在所述考题上正确地运用了这些必需、但没有掌握的技能中的每一项, 同时, 也在所述考题上正确地运用了剩余的、必需、同时也掌握了、选自所述选定的技能集合中的技能, 从而答对了所述考题; 所述模型针对所述选定的技能集合中的技能对, 表达了在每一个所述技能对的两个成员之间的正联系, 还表达了对每一个技能对正联系大小的估量, 其中, 针对每一个技能对的所述正联系可以由受试者对具体考题的反应估计出来; 以及

将根据受试者反应所获得的考试结果用于标定所述模型的具体考题, 以得到技能已掌握的预测, 或者技能未掌握的预测, 或者针对每一个由具体受试者和具体技能构成的组合不做任何预测。

4. 如权利要求3所述的方法，包括：

构建一项其中包括若干考题的考试，当受试者  $j$  答错或答对考题  $i$  时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个技能集合  $\{\alpha_{jk}\}$ ，当受试者  $j$  未掌握或已掌握技能  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方法表达的模型，包括可识别的、因此能够被标定的参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述考题如何依赖于所述选定的技能集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{i1}} \times (r_{i2}^*)^{1-\alpha_{i2}} \times \dots \times (r_{im}^*)^{1-\alpha_{im}}$$

其中， $S_{ij}$  是正确运用全部必需技能的概率，由受试者掌握及未掌握这些必需技能所确定；在  $S_{ij}$  中针对考题/技能关系矩阵中所规定的、考题  $i$  所要求的  $m$  项技能对  $r^*$ 's 求积； $\pi_i^* = \prod (\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在受试者未掌握技能  $k$  的情况下，将技能  $k$  正确应用于考题  $i$ )， $\pi_{ik} = \text{Prob}$  (在受试者已掌握技能  $k$  的情况下，将技能  $k$  正确应用于考题  $i$ )；间接表达了对每一个技能对正联系大小的估量，正如二值技能  $\alpha_k$ ， $\alpha_k$  的连续二变量正态技能前身  $\alpha'_k$ ， $\alpha'_k$  之间的相关性  $\sigma_{kk'}$  所确定的那样，然后，所需的二值技能对  $\alpha_k$ ， $\alpha_k$  可以通过在一个分界点处对每一个  $\alpha'_k$  进行分界而得到。

5. 一种方法，包括：

构建一项其中包括若干考题的考试，并选择一个技能集合，所述技能集合的设计目的是：测定参加所述考试的受试者的能力水平，并测定每一个受试者已经掌握还是尚未掌握所述技能；

构建一个以数学方式表达的模型，包括所述考题和所述选定的技能，所述选定的技能是一个更大的技能组的子集，所述更大的技能组中的技能可以影响受试者的考试表现，所述更大的技能组中未规定的剩余部分，通过剩余能力参数的形式，在所述模型中被体现；所述模型中包括参数，所述参数描述了所述考题如何依赖于所述选定的技能集合，也描述了所述考题如何依赖于所述剩余能力参数，其中，受试者对考题的反应提供了针对每一个参数的估计信息，使得可以对所述参数进行标定，也提供了对所述受试者掌握还是未掌握哪些技能的预测；所述模型还解释了如下事件的概率，即：对每一道考题而言，每一名受试者可能已掌握了所述考题所要求的全部技能，但未能在所述考题上正确地运用这些必需的、已经掌握的技能中的至少一项，因此答错了这道题，其中，所述考题所要求的全部技能是所述选定的技能集合的一个子集中的；以及如下事件的概率：对每一道考题而言，每一名受试者可能未掌握所述考题所要求的技能中的至少一项，但仍然在所述考题上正确地运用了这些必需、但没有掌握的技能中的每一项，同时，也在所述考题上正确地运用了剩余的、必需、同时也掌握了、选自所述选定的技能集合中的技能，从而答对了所述考题；所述模型定义了有待于赋以水平值的每种技能的掌握程度，表示一名超过所述水平的受试者已经具备了相应的技能水平；以及

将根据受试者反应所获得的考试结果用于标定所述模型的具体考题，以得到技能已掌握

的预测，或者技能未掌握的预测，或者针对每一个由具体受试者和具体技能构成的组合不做出任何预测。

6. 如权利要求 5 所述的方法，包括：

构建一项其中包括若干考题的考试，当受试者  $j$  答错或答对考题  $i$  时，相应地分别有  $X_{ij} = 0$  或 1；选择一个技能集合  $\{\alpha_{jk}\}$ ，当受试者  $j$  未掌握或已掌握技能  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或 1；以及

构建一个以数学方式表达的模型，包括可识别的、因此能够被标定的参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述考题如何依赖于所述选定的技能集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中， $S_{ij}$  是正确运用全部必需技能的概率，由受试者掌握及未掌握这些必需技能所确定；在  $S_{ij}$  中针对考题/技能关系矩阵中所规定的、考题  $i$  所要求的  $m$  项技能对  $r^*$ 's 求积； $\pi_i^* = \prod (\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在受试者未掌握技能  $k$  的情况下，将技能  $k$  正确应用于考题  $i$ )， $\pi_{ik} = \text{Prob}$  (在受试者已掌握技能  $k$  的情况下，将技能  $k$  正确应用于考题  $i$ )，其中， $p_k$  是由使用者确定的、被判断为已掌握技能  $k$  的受试者比例。

7. 一种方法，包括：

构建一项其中包括若干考题的考试，并选择一个技能集合，所述技能集合的设计目的是：测定参加所述考试的受试者的能力水平，并测定每一个受试者已经掌握还是尚未掌握所述技能；

构建一个以数学方式表达的模型，包括所述考题和所述选定的技能，所述选定的技能是一个更大的技能组的子集，所述更大的技能组中的技能可以影响受试者的考试表现，所述更大的技能组中未规定的剩余部分，通过剩余能力参数的形式，在所述模型中被体现；所述模型中包括参数，所述参数描述了所述考题如何依赖于所述选定的技能集合，也描述了所述考题如何依赖于所述剩余能力参数，还提供了对所述受试者掌握还是未掌握哪些技能的预测；所述模型还解释了如下事件的概率，即：对每一道考题而言，每一名受试者可能已掌握了所述考题所要求的全部技能，但未能在所述考题上正确地运用这些必需的、已经掌握的技能中的至少一项，因此答错了这道题，其中，所述考题所要求的全部技能是所述选定的技能集合的一个子集中的；以及如下事件的概率：对每一道考题而言，每一名受试者可能未掌握所述考题所要求的技能中的至少一项，但仍然在所述考题上正确地运用了这些必需、但没有掌握的技能中的每一项，同时，也在所述考题上正确地运用了剩余的、必需、同时也掌握了、选自所述选定的技能集合中的技能，从而答对了所述考题；所述模型将每种技能的掌握程度定义为一个设定的水平，表示一名超过所述水平的受试者已经具备了相应的技能水平；所述模型针对所述选定的技能集合中的技能对，表达了在每一个所述技能对的两个成员之间的正联系，还表达了对每一个技能对正联系大小的估量，其中，针对每一个技能对的所述正联系

可以由受试者对具体考题的反应估计出来；以及

将根据受试者反应所获得的考试结果用于标定所述模型的具体考题，以得到技能已掌握的预测，或者技能未掌握的预测，或者针对每一个由具体受试者和具体技能构成的组合不做出任何预测。

8. 如权利要求 7 所述的方法，包括：

构建一项其中包括若干考题的考试，当受试者  $j$  答错或答对考题  $i$  时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个技能集合  $\{\alpha_{jk}\}$ ，当受试者  $j$  未掌握或已掌握技能  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，包括参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述考题如何依赖于所述选定的技能集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{i1}} \times (r_{i2}^*)^{1-\alpha_{i2}} \times \dots \times (r_{im}^*)^{1-\alpha_{im}}$$

其中， $S_{ij}$  是正确运用全部必需技能的概率，由受试者掌握及未掌握这些必需技能所确定；在  $S_{ij}$  中针对考题/技能关系矩阵中所规定的、考题  $i$  所要求的  $m$  项技能对  $r^*$ 's 求积； $\pi_i^* = \prod (\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在受试者未掌握技能  $k$  的情况下，将技能  $k$  正确应用于考题  $i$ )， $\pi_{ik} = \text{Prob}$  (在受试者已掌握技能  $k$  的情况下，将技能  $k$  正确应用于考题  $i$ )；间接表达了对每一个技能对正联系大小的估量，正如二值技能  $\alpha_k, \alpha_{k'}$  的连续二变量正态技能前身  $\alpha'_k, \alpha'_{k'}$  之间的相关性  $\sigma_{kk'}$  所确定的那样，然后，所需的二值技能对  $\alpha_k, \alpha_{k'}$  可以通过如下方式产生：将每一个  $\alpha'_k$  在指定的掌握水平分界点处进行分界，从而使得技能  $k$  的掌握概率  $p_k(\alpha_k = 1)$  被定义为等于  $\text{Prob}(\alpha'_k \geq \text{分界点})$ ，其中，这样定义的技能掌握水平是通过选择  $\alpha'_k$  的分界点而实现的， $p_k$  是由使用者确定的、被判断为已掌握技能  $k$  的受试者比例，从而使得每一项技能通过连续的  $\alpha'_k$  及其分界点而获得了被赋予的水平，表示一名超过该水平的受试者已经掌握该项技能。

9. 一种方法，包括：

构建一项其中包括若干考题的考试，并选择一个技能集合，所述技能集合的设计目的是：测定参加所述考试的受试者的能力水平，并测定每一个受试者已经掌握还是尚未掌握所述技能；

构建一个以数学方式表达的模型，包括所述考题和所述选定的技能，所述选定的技能是一个更大的技能组的子集，所述更大的技能组中的技能可以影响受试者的考试表现，所述更大的技能组中未规定的剩余部分，通过剩余能力参数的形式，在所述模型中被体现；所述模型中包括参数，所述参数描述了所述考题如何依赖于所述选定的技能集合，也描述了所述考题如何依赖于所述剩余能力参数，其中，受试者对考题的反应提供了针对每一个参数的估计信息，使得可以对所述参数进行标定，也提供了对所述受试者掌握还是未掌握哪些技能的预测；所述模型还解释了如下事件的概率，即：对每一道考题而言，每一名受试者可能已掌握

了所述考题所要求的全部技能，但未能在所述考题上正确地运用这些必需的、已经掌握的技能中的至少一项，因此答错了这道题，其中，所述考题所要求的全部技能是所述选定的技能集合的一个子集中的；以及如下事件的概率：对每一道考题而言，每一名受试者可能未掌握所述考题所要求的技能中的至少一项，但仍然在所述考题上正确地运用了这些必需、但没有掌握的技能中的每一项，同时，也在所述考题上正确地运用了剩余的、必需、同时也掌握了、选自所述选定的技能集合中的技能，从而答对了所述考题；将根据受试者反应所获得的考试结果用于标定所述模型的具体考题，以得到技能已掌握的预测，或者技能未掌握的预测，或者针对每一个由具体受试者和具体技能构成的组合不做出任何预测。

10. 如权利要求 1 所述的方法，包括：

构建一项其中包括若干考题的考试，当受试者  $j$  答错或答对考题  $i$  时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个技能集合  $\{\alpha_{jk}\}$ ，当受试者  $j$  未掌握或已掌握技能  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，包括可识别的、因此能够被标定的参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述考题如何依赖于所述选定的技能集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中， $S_{ij}$  是正确运用全部必需技能的概率，由受试者掌握及未掌握这些必需技能所确定；在  $S_{ij}$  中针对考题/技能关系矩阵中所规定的、考题  $i$  所要求的  $m$  项技能对  $r^*$ 's 求积； $\pi_i^* = \prod (\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在受试者未掌握技能  $k$  的情况下，将技能  $k$  正确应用于考题  $i$ )， $\pi_{ik} = \text{Prob}$  (在受试者已掌握技能  $k$  的情况下，将技能  $k$  正确应用于考题  $i$ )。

11. 一种方法，包括：

构建一项其中包括若干考题的考试，并选择一个技能集合，所述技能集合的设计目的是：测定参加所述考试的受试者的能力水平，并测定每一个受试者已经掌握还是尚未掌握所述技能；

构建一个以数学方式表达的模型，包括所述考题和所述选定的技能，所述选定的技能是一个更大的技能组的子集，所述更大的技能组中的技能可以影响受试者的考试表现，所述更大的技能组中未规定的剩余部分，通过剩余能力参数的形式，在所述模型中被体现；所述模型中包括参数，所述参数描述了所述考题如何依赖于所述选定的技能集合，也描述了所述考题如何依赖于所述剩余能力参数，还提供了对所述受试者掌握还是未掌握哪些技能的预测；所述模型还解释了如下事件的概率，即：对每一道考题而言，每一名受试者可能已掌握了所述考题所要求的全部技能，但未能在所述考题上正确地运用这些必需的、已经掌握的技能中的至少一项，因此答错了这道题，其中，所述考题所要求的全部技能是所述选定的技能集合的一个子集中的；以及如下事件的概率：对每一道考题而言，每一名受试者可能未掌握所述

考题所要求的技能中的至少一项，但仍然在所述考题上正确地运用了这些必需、但没有掌握的技能中的每一项，同时，也在所述考题上正确地运用了剩余的、必需、同时也掌握了、选自所述选定的技能集中的技能，从而答对了所述考题；所述模型将每种技能的掌握程度定义为一个设定的水平，表示一名超过所述水平的受试者已经具备了相应的技能水平；以及将根据受试者反应所获得的考试结果用于标定所述模型的具体考题，以得到技能已掌握的预测，或者技能未掌握的预测，或者针对每一个由具体受试者和具体技能构成的组合不做出任何预测。

12. 如权利要求 11 所述的方法，包括：

构建一项其中包括若干考题的考试，当受试者  $j$  答错或答对考题  $i$  时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个技能集合  $\{\alpha_{jk}\}$ ，当受试者  $j$  未掌握或已掌握技能  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，包括参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述考题如何依赖于所述选定的技能集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中， $S_{ij}$  是正确运用全部必需技能的概率，由受试者掌握及未掌握这些必需技能所确定；在  $S_{ij}$  中针对考题/技能关系矩阵中所规定的、考题  $i$  所要求的  $m$  项技能对  $r^*$ 's 求积； $\pi_i^* = \prod (\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在受试者未掌握技能  $k$  的情况下，将技能  $k$  正确应用于考题  $i$ )， $\pi_{ik} = \text{Prob}$  (在受试者已掌握技能  $k$  的情况下，将技能  $k$  正确应用于考题  $i$ )，其中， $p_k$  是由使用者确定的、被判断为已掌握技能  $k$  的受试者比例。

13. 一种方法，包括：

构建一项其中包括若干考题的考试，并选择一个技能集合，所述技能集合的设计目的是：测定参加所述考试的受试者的能力水平，并测定每一个受试者已经掌握还是尚未掌握所述技能；

构建一个以数学方式表达的模型，包括所述考题和所述选定的技能，所述选定的技能是一个更大的技能组的子集，所述更大的技能组中的技能可以影响受试者的考试表现，所述更大的技能组中未规定的剩余部分，通过剩余能力参数的形式，在所述模型中被体现；所述模型中包括参数，所述参数描述了所述考题如何依赖于所述选定的技能集合，也描述了所述考题如何依赖于所述剩余能力参数，还提供了对所述受试者掌握还是未掌握哪些技能的预测；所述模型还解释了如下事件的概率，即：对每一道考题而言，每一名受试者可能已掌握了所述考题所要求的全部技能，但未能能在所述考题上正确地运用这些必需的、已经掌握的技能中的至少一项，因此答错了这道题，其中，所述考题所要求的全部技能是所述选定的技能集合的一个子集中的；以及如下事件的概率：对每一道考题而言，每一名受试者可能未掌握所述考题所要求的技能中的至少一项，但仍然在所述考题上正确地运用了这些必需、但没有掌握



的技能中的每一项，同时，也在所述考题上正确地运用了剩余的、必需、同时也掌握了、选自所述选定的技能集合中的技能，从而答对了所述考题；所述模型针对所述选定的技能集合中的技能对，表达了在每一个所述技能对的两个成员之间的正联系，还表达了对每一个技能对正联系大小的估量，其中，针对每一个技能对的所述正联系可以由受试者对具体考题的反应估计出来；以及

将根据受试者反应所获得的考试结果用于标定所述模型的具体考题，以得到技能已掌握的预测，或者技能未掌握的预测，或者针对每一个由具体受试者和具体技能构成的组合不做出任何预测。

#### 14. 如权利要求 7 所述的方法，包括：

构建一项其中包括若干考题的考试，当受试者  $j$  答错或答对考题  $i$  时，相应地分别有  $X_{ij} = 0$  或 1；选择一个技能集合  $\{\alpha_{jk}\}$ ，当受试者  $j$  未掌握或已掌握技能  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或 1；以及

构建一个以数学方式表达的模型，包括参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述考题如何依赖于所述选定的技能集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{i1}} \times (r_{i2}^*)^{1-\alpha_{i2}} \times \dots \times (r_{im}^*)^{1-\alpha_{im}}$$

其中， $S_{ij}$  是正确运用全部必需技能的概率，由受试者掌握及未掌握这些必需技能所确定；在  $S_{ij}$  中针对考题/技能关系矩阵中所规定的、考题  $i$  所要求的  $m$  项技能对  $r^*$ 's 求积； $\pi_i^* = \prod (\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在受试者未掌握技能  $k$  的情况下，将技能  $k$  正确应用于考题  $i$ )， $\pi_{ik} = \text{Prob}$  (在受试者已掌握技能  $k$  的情况下，将技能  $k$  正确应用于考题  $i$ )；间接表达了对每一个技能对正联系结构大小的测量，正如二值技能  $\alpha_k, \alpha_k$  的连续二变量正态技能前身  $\alpha'_k, \alpha'_k$  之间的相关性  $\sigma_{kk}$  所确定的那样，然后，所需的二值技能对  $\alpha_k, \alpha_k$  可以通过在一个指定的分界点处对每一个  $\alpha'_k$  进行分界而得到。

#### 15. 一种方法，包括：

构建一个集合，其中包括在医学上或精神病学上所关注的二值性患者症状或二值性个人特征，并为了评估的目的选择一个规定的集合，其中包括每位患者具有或不具有的、可能的医学或精神病学上的疾病，每位患者可能具有多种疾病；

构建一个以数学方式表达的模型，包括所述症状或特征以及所述为了评估而选出的规定疾病；其中，潜在健康或生活质量参数代表了未包括在所述规定的疾病集合中的患者的潜在方面；所述模型中包括参数，所述参数描述了所述症状或特征如何依赖于所述规定的疾病集合，也描述了所述症状或特征如何依赖于潜在一般性健康或生活质量参数，其中，所述患者的症状或特征提供了针对每一个参数的估计信息，使得可以对所述参数进行标定，也提供了对所述可能疾病的似然预测；所述模型还解释了如下事件的概率，即：一名患者可能具有一种疾病的典型症状或特征的集合，但是所述患者并未患有该疾病；以及，一名患者可能缺少

所述疾病的典型症状或特征中的至少一个，但所述患者却患有所述疾病；对于某些精神病学或医学疾病而言，所述模型针对所述某些精神病学或医学疾病中的每一种，定义了被判断为构成具有该疾病的水平；在所述选定疾病集合的所有疾病对中，所述模型针对其中的某些疾病对表达了同样疾病对中的每一个之间的联系，可能是正联系或负联系的；还表达了对所述选定疾病集合的每一个疾病对之间联系的大小的估量，所述的联系可以由所述患者对所述疾病或特征的反应而估计出来；以及，利用应用了患者数据的所述模型，以得到对患者具有所述规定的可能疾病集合中的每一种疾病的概率。

16. 如权利要求 15 所述的方法，包括：

构建一项医学或精神病学诊断，其中包括观测到的症状/特征，当患者  $j$  未表现出或表现出症状/特征  $i$  时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个可能疾病的集合  $\{\alpha_{jk}\}$ ，当患者  $j$  具有或不具有疾病  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，其中包括可识别的、因此能够被标定的参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述症状/特征如何依赖于所述选定的疾病集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中， $S_{ij}$  是表现出症状/特征  $i$  的概率，由患者  $j$  具有或不具有所述疾病所确定，并且采用了关于潜在健康/生活质量变量的完全性；针对关系矩阵中所规定的、与症状/特征  $i$  相联系的  $m$  种疾病对  $r^*$ 's 求积； $\pi_i^* = \prod(\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在所述患者具有疾病  $k$  的情况下，症状/特征  $i$ )， $\pi_{ik} = \text{Prob}$  (在所述患者不具有疾病  $k$  的情况下，症状/特征  $i$ )；间接表达了每一个疾病对之间联系的大小的估量，正如二值疾病  $\alpha_k, \alpha_{k'}$  的连续二变量正态疾病前身  $\alpha'_k, \alpha'_{k'}$  之间的相关性  $\sigma_{kk'}$  所确定的那样，然后，所需的二值疾病对  $\alpha_k, \alpha_{k'}$  可以通过如下方式产生：将每一个  $\alpha'_k$  在指定的疾病分界点处进行分界，从而使得疾病  $k$  的患病概率  $p_k$  被定义为等于  $\text{Prob}(\alpha'_k \geq \text{分界点})$ ，其中，这样定义的、被判断为构成具有所述疾病 ( $\alpha_k = 0$ ) 的所述疾病的水平  $\alpha'_k$ ，是通过选择  $\alpha'_k$  的分界点而实现的，其中， $p_k$  是被判断为具有所述疾病的患者的比例，所述的比例是通过设定  $\alpha'_k$  的分界点而确定的，从而使得每一种疾病通过连续的  $\alpha'_k$  及其分界点而获得了被赋予的水平，表示一名超过该水平的患者具有所述疾病。

17. 一种方法，包括：

构建一个集合，其中包括在医学上或精神病学上关注的二值性患者症状或二值性个人特征，并为了评估的目的选择一个规定的集合，其中包括每位患者具有或不具有的、可能的医学或精神病学上的疾病，每位患者可能具有多种疾病；

构建一个以数学方式表达的模型，包括所述症状或特征以及所述为了评估而选出的规定疾病；其中，潜在健康或生活质量参数代表了未包括在所述规定的疾病集合中的患者的潜在方面；所述模型中包括参数，所述参数描述了所述症状或特征如何依赖于所述规定的疾病集

合，也描述了所述症状或特征如何依赖于潜在一般性健康或生活质量参数，其中，所述患者的症状或特征提供了针对每一个参数的估计信息，使得可以对所述参数进行标定，也提供了对所述可能疾病的似然预测；所述模型还解释了如下事件的概率，即：一名患者可能具有一种疾病的典型症状或特征的集合，但是所述患者并未患有该疾病；以及，一名患者可能缺少所述疾病的典型症状或特征中的至少一个，但所述患者却患有所述疾病；在所述选定疾病集合的所有疾病对中，所述模型针对其中的某些疾病对表达了同样疾病对中的每一个之间的联系，可能是正联系或负联系的；还表达了对所述选定疾病集合的每一个疾病对之间联系大小的估量，所述的联系可以由所述患者对所述疾病或特征的反应而估计出来；以及，利用应用了患者数据的所述模型，以得到对患者具有所述规定的可能疾病集合中的每一种疾病的概率。

18. 如权利要求 17 所述的方法，包括：

构建一项医学或精神病学诊断，其中包括观测到的症状/特征，当患者  $j$  未表现出或表现出症状/特征  $i$  时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个可能疾病的集合  $\{\alpha_{jk}\}$ ，当患者  $j$  具有或不具有疾病  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，其中包括可识别的、因此能够被标定的参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述症状/特征如何依赖于所述选定的疾病集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中， $S_{ij}$  是表现出症状/特征  $i$  的概率，由患者  $j$  具有或不具有所述疾病所确定，并且采用了关于潜在健康/生活质量变量的完全性；针对关系矩阵中所规定的、与症状/特征  $i$  相联系的  $m$  种疾病对  $r^*$ 's 求积； $\pi_i^* = \prod(\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在所述患者具有疾病  $k$  的情况下，症状/特征  $i$ )， $\pi_{ik} = \text{Prob}$  (在所述患者不具有疾病  $k$  的情况下，症状/特征  $i$ )；间接表达了每一个疾病对之间联系的大小的测量，正如二值疾病  $\alpha_k, \alpha_k'$  的连续二变量正态疾病前身  $\alpha'_k, \alpha'_k'$  之间的相关性  $\sigma_{kk'}$  所确定的那样，然后，所需的二值疾病对  $\alpha_k, \alpha_k'$  可以通过将每一个  $\alpha'_k$  在一个分界点处进行分界而得到。

19. 一种方法，包括：

构建一个集合，其中包括在医学上或精神病学上所关注的二值性患者症状或二值性个人特征，并为了评估的目的选择一个规定的集合，其中包括每位患者具有或不具有的、可能的医学或精神病学上的疾病，每位患者可能具有多种疾病；

构建一个以数学方式表达的模型，包括所述症状或特征以及所述为了评估而选出的规定疾病；其中，潜在健康或生活质量参数代表了未包括在所述规定的疾病集合中的患者的潜在方面；所述模型中包括参数，所述参数描述了所述症状或特征如何依赖于所述规定的疾病集合，也描述了所述症状或特征如何依赖于潜在一般性健康或生活质量参数，其中，所述患者的症状或特征提供了针对每一个参数的估计信息，使得可以对所述参数进行标定，也提供了

对所述可能疾病的似然预测；所述模型还解释了如下事件的概率，即：一名患者可能具有一种疾病的典型症状或特征的集合，但是所述患者并未患有该疾病；以及，一名患者可能缺少所述疾病的典型症状或特征中的至少一个，但所述患者却患有所述疾病；对于某些精神病学或医学疾病而言，所述模型针对所述某些精神病学或医学疾病中的每一种，定义了被判断为构成具有该疾病的水平；以及，

利用应用了患者数据的所述模型，以得到对患者具有所述规定的可能疾病集合中的每一种疾病的概率。

20. 如权利要求 19 所述的方法，包括：

构建一项医学或精神病学诊断，其中包括观测到的症状/特征，当患者  $j$  未表现出或表现出症状/特征  $i$  时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个可能疾病的集合  $\{\alpha_{jk}\}$ ，当患者  $j$  具有或不具有疾病  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，其中包括可识别的、因此能够被标定的参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述症状/特征如何依赖于所述选定的疾病集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中， $S_{ij}$  是表现出症状/特征  $i$  的概率，由患者  $j$  具有或不具有所述疾病所确定，并且采用了关于潜在健康/生活质量变量的完全性；针对关系矩阵中所规定的、与症状/特征  $i$  相联系的  $m$  种疾病对  $r^*$ 's 求积； $\pi_i^* = \prod(\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在所述患者具有疾病  $k$  的情况下，症状/特征  $i$ )， $\pi_{ik} = \text{Prob}$  (在所述患者不具有疾病  $k$  的情况下，症状/特征  $i$ )，其中， $p_k$  是被判断为具有所述疾病的患者的比例，是由定义所述疾病的使用者确定的。

21. 一种方法，包括：

构建一个集合，其中包括在医学上或精神病学上所关注的二值性患者症状或二值性个人特征，并为了评估的目的选择一个规定的集合，其中包括每位患者具有或不具有的、可能的医学或精神病学上的疾病，每位患者可能具有多种疾病；

构建一个以数学方式表达的模型，包括所述症状或特征以及所述为了评估而选出的规定疾病；其中，潜在健康或生活质量参数代表了未包括在所述规定的疾病集合中的患者的潜在方面；所述模型中包括参数，所述参数描述了所述症状或特征如何依赖于所述规定的疾病集合，也描述了所述症状或特征如何依赖于潜在一般性健康或生活质量参数；所述模型还解释了如下事件的概率，即：一名患者可能具有一种疾病的典型症状或特征的集合，但是所述患者并未患有该疾病；以及，一名患者可能缺少所述疾病的典型症状或特征中的至少一个，但所述患者却患有所述疾病；对于某些精神病学或医学疾病而言，所述模型针对所述某些精神病学或医学疾病中的每一种，定义了被判断为构成具有该疾病的水平；在所述选定疾病集合的所有疾病对中，所述模型针对其中的某些疾病对表达了同样疾病对中的每一个之间的联

系，可能是正联系或负联系的；还表达了对所述选定疾病集合的每一个疾病对之间联系大小的估量，所述的联系可以由所述患者对所述疾病或特征的反应而估计出来；以及，

利用应用了患者数据的所述模型，以得到对患者具有所述规定的可能疾病集合中的每一种疾病的概率。

22. 如权利要求 21 所述的方法，包括：

构建一项医学或精神病学诊断，其中包括观测到的症状/特征，当患者  $j$  未表现出或表现出症状/特征  $i$  时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个可能疾病的集合  $\{\alpha_{jk}\}$ ，当患者  $j$  具有或不具有疾病  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，其中包括参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述症状/特征如何依赖于所述选定的疾病集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{i1}} \times (r_{i2}^*)^{1-\alpha_{i2}} \times \dots \times (r_{im}^*)^{1-\alpha_{im}}$$

其中， $S_{ij}$  是表现出症状/特征  $i$  的概率，由患者  $j$  具有或不具有所述疾病所确定，并且采用了关于潜在健康/生活质量变量的完全性；针对关系矩阵中所规定的、与症状/特征  $i$  相联系的  $m$  种疾病对  $r^*$ 's 求积； $\pi_i^* = \prod(\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik}/\pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在所述患者具有疾病  $k$  的情况下，症状/特征  $i$ )， $\pi_{ik} = \text{Prob}$  (在所述患者不具有疾病  $k$  的情况下，症状/特征  $i$ )；间接表达了每一个疾病对之间联系的大小的测量，正如二值疾病  $\alpha_k, \alpha_{k'}$  的连续二变量正态疾病前身  $\alpha'_k, \alpha'_{k'}$  之间的相关性  $\sigma_{kk'}$  所确定的那样，然后，所需的二值疾病对  $\alpha_k, \alpha_{k'}$  可以通过如下方式产生：将每一个  $\alpha'_k$  在指定的疾病分界点处进行分界，从而使得疾病  $k$  的患病概率  $p_k$  被定义为等于  $\text{Prob}(\alpha'_k \geq \text{分界点})$ ，其中，这样定义的、被判断为构成具有所述疾病 ( $\alpha_k = 0$ ) 的所述疾病的水平  $\alpha'_k$ ，是通过选择  $\alpha'_k$  的分界点而实现的，其中， $p_k$  是被判断为具有所述疾病的患者的比例，所述的比例是通过设定  $\alpha'_k$  的分界点而确定的，从而使得每一种疾病通过连续的  $\alpha'_k$  及其分界点而获得了被赋予的水平，表示一名超过该水平的患者具有所述疾病。

23. 一种方法，包括：

构建一个集合，其中包括在医学上或精神病学上所关注的二值性患者症状或二值性个人特征，并为了评估的目的选择一个规定的集合，其中包括每位患者具有或不具有的、可能的医学或精神病学上的疾病，每位患者可能具有多种疾病；

构建一个以数学方式表达的模型，包括所述症状或特征以及所述为了评估而选出的规定疾病；其中，潜在健康或生活质量参数代表了未包括在所述规定的疾病集合中的患者的潜在方面；所述模型中包括参数，所述参数描述了所述症状或特征如何依赖于所述规定的疾病集合，也描述了所述症状或特征如何依赖于潜在一般性健康或生活质量参数，其中，所述患者的症状或特征提供了针对每一个参数的估计信息，使得可以对所述参数进行标定，也提供了对所述可能疾病的似然预测；所述模型还解释了如下事件的概率，即：一名患者可能具有一

种疾病的典型症状或特征的集合，但是所述患者并未患有该疾病；以及，一名患者可能缺少所述疾病的典型症状或特征中的至少一个，但所述患者却患有所述疾病；以及，

利用应用了患者数据的所述模型，以得到对患者具有所述规定的可能疾病集合中的每一种疾病的概率。

24. 如权利要求 23 所述的方法，包括：

构建一项医学或精神病学诊断，其中包括观测到的症状/特征，当患者  $j$  未表现出或表现出症状/特征  $i$  时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个可能疾病的集合  $\{\alpha_{jk}\}$ ，当患者  $j$  具有或不具有疾病  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，其中包括可识别的、因此能够被标定的参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述症状/特征如何依赖于所述选定的疾病集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中， $S_{ij}$  是表现出症状/特征  $i$  的概率，由患者  $j$  具有或不具有所述疾病所确定，并且采用了关于潜在健康/生活质量变量的完全性；针对关系矩阵中所规定的、与症状/特征  $i$  相联系的  $m$  种疾病对  $r^*$ 's 求积； $\pi_i^* = \prod(\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在所述患者具有疾病  $k$  的情况下，症状/特征  $i$ )， $\pi_{ik} = \text{Prob}$  (在所述患者不具有疾病  $k$  的情况下，症状/特征  $i$ )。

25. 一种方法，包括：

构建一个集合，其中包括在医学上或精神病学上所关注的二值性患者症状或二值性个人特征，并为了评估的目的选择一个规定的集合，其中包括每位患者具有或不具有的、可能的医学或精神病学上的疾病，每位患者可能具有多种疾病；

构建一个以数学方式表达的模型，包括所述症状或特征以及所述为了评估而选出的规定疾病；其中，潜在健康或生活质量参数代表了未包括在所述规定的疾病集合中的患者的潜在方面；所述模型中包括参数，所述参数描述了所述症状或特征如何依赖于所述规定的疾病集合，也描述了所述症状或特征如何依赖于潜在一般性健康或生活质量参数；所述模型还解释了如下事件的概率，即：一名患者可能具有一种疾病的典型症状或特征的集合，但是所述患者并未患有该疾病；以及，一名患者可能缺少所述疾病的典型症状或特征中的至少一个，但所述患者却患有所述疾病；对于某些精神病学或医学疾病而言，所述模型针对所述某些精神病学或医学疾病中的每一种，定义了被判断为构成具有该疾病的水平；以及，

利用应用了患者数据的所述模型，以得到对患者具有所述规定的可能疾病集合中的每一种疾病的概率。

26. 如权利要求 25 所述的方法，包括：

构建一项医学或精神病学诊断，其中包括观测到的症状/特征，当患者  $j$  未表现出或表现

出症状/特征  $i$  时, 相应地分别有  $X_{ij} = 0$  或  $1$ ; 选择一个可能疾病的集合  $\{\alpha_{jk}\}$ , 当患者  $j$  具有或不具有疾病  $k$  时, 相应地分别有  $\alpha_{jk} = 0$  或  $1$ ; 以及

构建一个以数学方式表达的模型, 其中包括参数  $\{\pi^*, r^*\}$ , 所述参数按照下述概率描述了所述症状/特征如何依赖于所述选定的疾病集合:

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中,  $S_{ij}$  是表现出症状/特征  $i$  的概率, 由患者  $j$  具有或不具有所述疾病所确定, 并且采用了关于潜在健康/生活质量变量的完全性; 针对关系矩阵中所规定的、与症状/特征  $i$  相联系的  $m$  种疾病对  $r^*$ 's 求积;  $\pi_i^* = \prod(\pi_{ik})$ , 对  $k$  求积;  $r_{ik}^* = r_{ik}/\pi_{ik}$ , 其中,  $r_{ik} = \text{Prob}$  (在所述患者具有疾病  $k$  的情况下, 症状/特征  $i$ ),  $\pi_{ik} = \text{Prob}$  (在所述患者不具有疾病  $k$  的情况下, 症状/特征  $i$ ), 其中,  $p_k$  是被判断为具有所述疾病的患者的比例, 是由定义所述疾病的使用者确定的。

27. 一种方法, 包括:

构建一个集合, 其中包括在医学上或精神病学上所关注的二值性患者症状或二值性个人特征, 并为了评估的目的选择一个规定的集合, 其中包括每位患者具有或不具有的、可能的医学或精神病学上的疾病, 每位患者可能具有多种疾病;

构建一个以数学方式表达的模型, 包括所述症状或特征以及所述为了评估而选出的规定疾病; 其中, 潜在健康或生活质量参数代表了未包括在所述规定的疾病集合中的患者的潜在方面; 所述模型中包括参数, 所述参数描述了所述症状或特征如何依赖于所述规定的疾病集合, 也描述了所述症状或特征如何依赖于潜在一般性健康或生活质量; 所述模型还解释了如下事件的概率, 即: 一名患者可能具有一种疾病的典型症状或特征的集合, 但是所述患者并未患有该疾病; 以及, 一名患者可能缺少所述疾病的典型症状或特征中的至少一个, 但所述患者却患有所述疾病; 在所述选定疾病集合的所有疾病对中, 所述模型针对其中的某些疾病对表达了同样疾病对中的每一个之间的联系, 可能是正联系或负联系的; 还表达了对所述选定疾病集合的每一个疾病对之间相关的大小的估量, 所述的联系可以由所述患者对所述疾病或特征的反应而估计出来; 以及,

利用应用了患者数据的所述模型, 以得到对患者具有所述规定的可能疾病集合中的每一种疾病的概率。

28. 如权利要求 27 所述的方法, 包括:

构建一项医学或精神病学诊断, 其中包括观测到的症状/特征, 当患者  $j$  未表现出或表现出症状/特征  $i$  时, 相应地分别有  $X_{ij} = 0$  或  $1$ ; 选择一个可能疾病的集合  $\{\alpha_{jk}\}$ , 当患者  $j$  具有或不具有疾病  $k$  时, 相应地分别有  $\alpha_{jk} = 0$  或  $1$ ; 以及

构建一个以数学方式表达的模型, 其中包括参数  $\{\pi^*, r^*\}$ , 所述参数按照下述概率描述了所述症状/特征如何依赖于所述选定的疾病集合:

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{i1}} \times (r_{i2}^*)^{1-\alpha_{i2}} \times \dots \times (r_{im}^*)^{1-\alpha_{im}}$$

其中,  $S_{ij}$  是表现出症状/特征  $i$  的概率, 由患者  $j$  具有或不具有所述疾病所确定, 并且采用了关于潜在健康/生活质量变量的完全性; 针对关系矩阵中所规定的、与症状/特征  $i$  相联系的  $m$  种疾病对  $r^*$ 's 求积;  $\pi_i^* = \prod(\pi_{ik})$ , 对  $k$  求积;  $r_{ik}^* = r_{ik} / \pi_{ik}$ , 其中,  $r_{ik} = \text{Prob}$  (在所述患者具有疾病  $k$  的情况下, 症状/特征  $i$ ),  $\pi_{ik} = \text{Prob}$  (在所述患者不具有疾病  $k$  的情况下, 症状/特征  $i$ ); 间接表达了每一个疾病对之间联系的大小的测量, 正如二值疾病  $\alpha_k, \alpha_k'$  的连续二变量正态疾病前身  $\alpha'_k, \alpha'_k'$  之间的相关性  $\sigma_{kk'}$  所确定的那样, 然后, 所需的二值疾病对  $\alpha_k, \alpha_k'$  可以通过将每一个  $\alpha'_k$  在一个分界点处进行分界而得到。

#### 29. 一种方法, 包括:

构建一个其中包括若干二值记分的探针的集合, 并选择一个集合, 其中包括每个对象所具有或不具有的、观测不到的若干二值属性, 其目的是评估每一个被探测的所述对象的潜在状态;

构建一个以数学方式表达的模型, 包括所述探针和所述选定的属性, 所述选定的属性是一个更大的属性组的子集, 所述更大的属性组中的属性可以影响对象对探针的反应, 所述更大的属性组中未规定的剩余部分, 通过剩余状态参数的形式, 在所述模型中被体现; 所述模型中包括参数, 所述参数描述了所述探针如何依赖于所述选定的属性集合, 也描述了所述探针如何依赖于所述剩余状态参数, 其中, 对象对所述探针的反应提供了针对每一个参数的估计信息, 使得可以对所述参数进行标定, 也提供了对所述对象具有或不具有哪些属性的预测; 所述模型还解释了如下事件的概率, 即: 对每一个具体探针而言, 一个对象可能具有对所述具体探针做出正反应所需的全部属性, 但未能在所述的具体探针上适当地应用至少一项必需属性, 从而对所述具体探针做出了负反应, 其中, 所述对具体探针做出正反应所需的全部属性是所述选定的属性集合的一个子集中的; 以及如下事件的概率: 对每一个具体探针而言, 一个对象可能未具有对所述探针做出正反应所需的规定属性中的至少一项, 但仍然在所述探针上适当地应用了所述未具有的、规定的必需属性, 同时, 也适当地应用了剩余的、必需、同时也具有的、选自所述选定属性集合中的属性, 从而对所述具体探针做出了正反应; 所述模型将具有每种属性的水平定义为一个设定的水平, 所述设定的水平被判断为可令对象具有所述的具体属性; 所述模型针对所述选定属性集合的某些属性对, 表达了每一个属性对的两个成员之间的联系, 所述联系可能是正的或负的; 还表达了对每一个属性对的所述正的或负的联系大小的估量, 每一个属性对的所述联系可以由所述对象对所述具体探针的反应而估计出来;

将根据所述对象的反应而获得的综合探测结果用于标定所述模型的具体探针; 以及

得到具有所述属性的预测, 或者不具有所述属性的预测, 或者针对每一个由对象和规定属性所构成的组合不做出任何预测。



30. 如权利要求 29 所述的方法，包括：

构建一个集合，其中包括若干二值记分的探针，当对象  $j$  对探针  $i$  做出负反应或正反应时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个潜在属性的集合  $\{\alpha_{jk}\}$ ，当对象  $j$  不具有或具有属性  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，包括可识别的、因此能够被标定的参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述探针如何依赖于所述选定的潜在属性集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中， $S_{ij}$  是对探针  $i$  做出正反应的概率，由对象  $j$  具有及不具有所述规定属性所确定，并且采用了关于所述剩余状态的完全性；针对对探针  $i$  做出正反应所要求的  $m$  项属性对  $r^*$ 's 求积： $\pi_i^* = \prod(\pi_{ik})$ ，对  $k$  求积： $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在所述对象不具有属性  $k$  的情况下，对探针  $i$  做出正反应)， $\pi_{ik} = \text{Prob}$  (在所述对象具有属性  $k$  的情况下，对探针  $i$  做出正反应)；间接表达了对每一个属性对正联系结构大小的测量，正如二值属性  $\alpha_k, \alpha_{k'}$  的连续二变量正态属性前身  $\alpha'_k, \alpha'_{k'}$  之间的相关性  $\sigma_{kk'}$  所确定的那样，然后，所需的二值属性对  $\alpha_k, \alpha_{k'}$  可以通过如下方式产生：将每一个  $\alpha'_k$  在指定的具有水平分界点处进行分界，从而使得具有属性  $k$  的概率  $p_k(\alpha_k = 1)$  被定义为等于  $\text{Prob}(\alpha'_k \geq \text{分界点})$ ，其中，这样定义的具有属性的水平是通过选择  $\alpha'_k$  的分界点而实现的， $p_k$  是由使用者确定的、被判断为具有属性  $k$  的对象的比例，从而使得每一种属性通过连续的  $\alpha'_k$  及其分界点而获得了被赋予的水平，表示一名超过该水平的对象具有所述属性。

31. 一种方法，包括：

构建一个其中包括若干二值记分的探针的集合，并选择一个集合，其中包括每个对象所具有或不具有的、观测不到的若干二值属性，其目的是评估每一个被探测的所述对象的潜在状态；

构建一个以数学方式表达的模型，包括所述探针和所述选定的属性，所述选定的属性是一个更大的属性组的子集，所述更大的属性组中的属性可以影响对象对探针的反应，所述更大的属性组中未规定的剩余部分，通过剩余状态参数的形式，在所述模型中被体现；所述模型中包括参数，所述参数描述了所述探针如何依赖于所述选定的属性集合，也描述了所述探针如何依赖于所述剩余状态参数，其中，对象对所述探针的反应提供了针对每一个参数的估计信息，使得可以对所述参数进行标定，也提供了对所述对象具有或不具有哪些属性的预测；所述模型还解释了如下事件的概率，即：对每一个具体探针而言，一个对象可能具有对所述具体探针做出正反应所需的全部属性，但未能在所述的具体探针上适当地应用至少一项必需属性，从而对所述具体探针做出了负反应，其中，所述对具体探针做出正反应所需的全部属性是所述选定的属性集合的一个子集中的；以及如下事件的概率：对每一个具体探针而言，一个对象可能未具有对所述探针做出正反应所需的规定属性中的至少一项，但仍然在所述探针上适当地应用了所述未具有的、选定的必需属性，同时，也适当地应用了所述选定属

性集合中的、未规定的剩余属性，从而对所述具体探针做出了正反应；所述模型针对所述选定属性集合的某些属性对，表达了每一个属性对的两个成员之间的联系，所述联系可能是正的或负的；还表达了对每一个属性对的所述正的或负的联系的大小的估量，每一个属性对的所述联系可以由所述对象对所述具体探针的反应而估计出来；

将根据所述对象的反应而获得的综合探测结果用于标定所述模型的具体探针；以及得到具有所述属性的预测，或者不具有所述属性的预测，或者针对每一个由对象和规定属性所构成的组合不做出任何预测。

32. 如权利要求 29 所述的方法，包括：

构建一个集合，其中包括若干二值记分的探针，当对象  $j$  对探针  $i$  做出负反应或正反应时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个潜在属性的集合  $\{\alpha_{jk}\}$ ，当对象  $j$  不具有或具有属性  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，包括可识别的、因此能够被标定的参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述探针如何依赖于所述选定的潜在属性集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中， $S_{ij}$  是对探针  $i$  做出正反应的概率，由对象  $j$  具有及不具有所述规定属性所确定，并且采用了关于所述剩余状态的完全性；针对对探针  $i$  做出正反应所要求的  $m$  项属性对  $r^*$ 's 求积； $\pi_i^* = \prod(\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik}/\pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在所述对象不具有属性  $k$  的情况下，对探针  $i$  做出正反应)， $\pi_{ik} = \text{Prob}$  (在所述对象具有属性  $k$  的情况下，对探针  $i$  做出正反应)；间接表达了对每一个属性对正联系结构大小的测量，正如二值属性  $\alpha_k, \alpha_{k'}$  的连续二变量正态属性前身  $\alpha'_k, \alpha'_{k'}$  之间的相关性  $\sigma_{kk'}$  所确定的那样，然后，所需的二值属性对  $\alpha_k, \alpha_{k'}$  可以通过将每一个  $\alpha'_k$  在某一水平进行分界而获得。

33. 一种方法，包括：

构建一个其中包括若干二值记分的探针的集合，并选择一个集合，其中包括每个对象所具有或不具有的、观测不到的若干二值属性，其目的是评估每一个被探测的所述对象的潜在状态；

构建一个以数学方式表达的模型，包括所述探针和所述选定的属性，所述选定的属性是一个更大的属性组的子集，所述更大的属性组中的属性可以影响对象对探针的反应，所述更大的属性组中未规定的剩余部分，通过剩余状态参数的形式，在所述模型中被体现；所述模型中包括参数，所述参数描述了所述探针如何依赖于所述选定的属性集合，也描述了所述探针如何依赖于所述剩余状态参数，其中，对象对所述探针的反应提供了针对每一个参数的估计信息，使得可以对所述参数进行标定，也提供了对所述对象具有或不具有哪些属性的预测；所述模型还解释了如下事件的概率，即：对每一个具体探针而言，一个对象可能具有对所述具体探针做出正反应所需的全部属性，但未能在所述的具体探针上适当地应用至少一项

必需属性，从而对所述具体探针做出了负反应，其中，所述对具体探针做出正反应所需的全部属性是所述选定的属性集合的一个子集中的；以及如下事件的概率：对每一个具体探针而言，一个对象可能未具有对所述探针做出正反应所需的规定属性中的至少一项，但仍然在所述探针上适当地应用了所述未具有的、选定的必需属性，同时，也适当地应用了所述选定属性集合中的、未规定的剩余属性，从而对所述具体探针做出了正反应；所述模型将具有每种属性的水平定义为一个设定的水平，所述设定的水平被判断为可令对象具有所述的具体属性；

将根据所述对象的反应而获得的综合探测结果用于标定所述模型的具体探针；以及得到具有所述属性的预测，或者不具有所述属性的预测，或者针对每一个由对象和规定属性所构成的组合不做出任何预测。

34. 如权利要求 29 所述的方法，包括：

构建一个集合，其中包括若干二值记分的探针，当对象  $j$  对探针  $i$  做出负反应或正反应时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个潜在属性的集合  $\{\alpha_{jk}\}$ ，当对象  $j$  不具有或具有属性  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，包括可识别的、因此能够被标定的参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述探针如何依赖于所述选定的潜在属性集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中， $S_{ij}$  是对探针  $i$  做出正反应的概率，由对象  $j$  具有及不具有所述规定属性所确定，并且采用了关于所述剩余状态的完全性；针对对探针  $i$  做出正反应所要求的  $m$  项属性对  $r^*$ 's 求积； $\pi_i^* = \prod(\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在所述对象不具有属性  $k$  的情况下，对探针  $i$  做出正反应)， $\pi_{ik} = \text{Prob}$  (在所述对象具有属性  $k$  的情况下，对探针  $i$  做出正反应)，其中， $p_k$  是由使用者确定的、被判断为具有属性  $k$  的对象的比例。

35. 一种方法，包括：

构建一个其中包括若干二值记分的探针的集合，并选择一个集合，其中包括每个对象所具有或不具有的、观测不到的若干二值属性，其目的是评估每一个被探测的所述对象的潜在状态；

构建一个以数学方式表达的模型，包括所述探针和所述选定的属性，所述选定的属性是一个更大的属性组的子集，所述更大的属性组中的属性可以影响对象对探针的反应，所述更大的属性组中未规定的剩余部分，通过剩余状态参数的形式，在所述模型中被体现；所述模型中包括参数，所述参数描述了所述探针如何依赖于所述选定的属性集合，也描述了所述探针如何依赖于所述剩余状态参数，所述模型中也包括对所述对象具有或不具有哪些属性的预测；所述模型还解释了如下事件的概率，即：对每一个具体探针而言，一个对象可能具有对所述具体探针做出正反应所需的全部属性，但未能在所述的具体探针上适当地应用至少一项

必需属性，从而对所述具体探针做出了负反应，其中，所述对具体探针做出正反应所需的全部属性是所述选定的属性集合的一个子集中的；以及如下事件的概率：对每一个具体探针而言，一个对象可能未具有对所述探针做出正反应所需的规定属性中的至少一项，但仍然在所述探针上适当地应用了所述未具有的、选定的必需属性，同时，也适当地应用了所述选定属性集合中的、未规定的剩余属性，从而对所述具体探针做出了正反应；所述模型将具有每种属性的水平定义为一个设定的水平，所述设定的水平被判断为可令对象具有所述的具体属性；所述模型针对所述选定属性集合的某些属性对，表达了每一个属性对的两个成员之间的联系，所述联系可能是正的或负的；还表达了对每一个属性对的所述正的或负的联系的大小的测量，每一个属性对的所述联系可以由所述对象对所述具体探针的反应而估计出来；

将根据所述对象的反应而获得的综合探测结果用于标定所述模型的具体探针；以及得到具有所述属性的预测，或者不具有所述属性的预测，或者针对每一个由对象和规定属性所构成的组合不做出任何预测。

36. 如权利要求 29 所述的方法，包括：

构建一个集合，其中包括若干二值记分的探针，当对象  $j$  对探针  $i$  做出负反应或正反应时，相应地分别有  $X_{ij} = 0$  或 1；选择一个潜在属性的集合  $\{\alpha_{jk}\}$ ，当对象  $j$  不具有或具有属性  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或 1；以及

构建一个以数学方式表达的模型，包括参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述探针如何依赖于所述选定的潜在属性集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{i1}} \times (r_{i2}^*)^{1-\alpha_{i2}} \times \dots \times (r_{im}^*)^{1-\alpha_{im}}$$

其中， $S_{ij}$  是对探针  $i$  做出正反应的概率，由对象  $j$  具有及不具有所述规定属性所确定，并且采用了关于所述剩余状态的完全性；针对对探针  $i$  做出正反应所要求的  $m$  项属性对  $r^*$ 's 求积； $\pi_i^* = \prod(\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik}/\pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在所述对象不具有属性  $k$  的情况下，对探针  $i$  做出正反应)， $\pi_{ik} = \text{Prob}$  (在所述对象具有属性  $k$  的情况下，对探针  $i$  做出正反应)；间接表达了对每一个属性对正联系结构大小的测量，正如二值属性  $\alpha_k, \alpha_k'$  的连续二变量正态属性前身  $\alpha'_k, \alpha'_k'$  之间的相关性  $\sigma_{kk'}$  所确定的那样，然后，所需的二值属性对  $\alpha_k, \alpha_k'$  可以通过如下方式产生：将每一个  $\alpha'_k$  在指定的具有水平分界点处进行分界，从而使得具有属性  $k$  的概率  $p_k(\alpha_k = 1)$  被定义为等于  $\text{Prob}(\alpha'_k \geq \text{分界点})$ ，其中，这样定义的具有属性的水平是通过选择  $\alpha'_k$  的分界点而实现的， $p_k$  是由使用者确定的、被判断为具有属性  $k$  的对象的比例，从而使得每一种属性通过连续的  $\alpha'_k$  及其分界点而获得了被赋予的水平，表示一名超过该水平的对象具有所述属性。

37. 一种方法，包括：

构建一个其中包括若干二值记分的探针的集合，并选择一个集合，其中包括每个对象所具有或不具有的、观测不到的若干二值属性，其目的是评估每一个被探测的所述对象的潜在

状态;

构建一个以数学方式表达的模型, 包括所述探针和所述选定的属性, 所述选定的属性是一个更大的属性组的子集, 所述更大的属性组中的属性可以影响对象对探针的反应, 所述更大的属性组中未规定的剩余部分, 通过剩余状态参数的形式, 在所述模型中被体现; 所述模型中包括参数, 所述参数描述了所述探针如何依赖于所述选定的属性集合, 也描述了所述探针如何依赖于所述剩余状态参数, 其中, 对象对所述探针的反应提供了针对每一个参数的估计信息, 使得可以对所述参数进行标定, 也提供了对所述对象具有或不具有哪些属性的预测; 所述模型还解释了如下事件的概率, 即: 对每一个具体探针而言, 一个对象可能具有对所述具体探针做出正反应所需的全部属性, 但未能在所述的具体探针上适当地应用至少一项必需属性, 从而对所述具体探针做出了负反应, 其中, 所述对具体探针做出正反应所需的全部属性是所述选定的属性集合的一个子集中的; 以及如下事件的概率: 对每一个具体探针而言, 一个对象可能未具有对所述探针做出正反应所需的规定属性中的至少一项, 但仍然在所述探针上适当地应用了所述未具有的、选定的必需属性, 同时, 也适当地应用了所述选定属性集合中的、未规定的剩余属性, 从而对所述具体探针做出了正反应;

将根据所述对象的反应而获得的综合探测结果用于标定所述模型的具体探针; 以及

得到具有所述属性的预测, 或者不具有所述属性的预测, 或者针对每一个由对象和规定属性所构成的组合不做出任何预测。

38. 如权利要求 29 所述的方法, 包括:

构建一个集合, 其中包括若干二值记分的探针, 当对象  $j$  对探针  $i$  做出负反应或正反应时, 相应地分别有  $X_{ij} = 0$  或  $1$ ; 选择一个潜在属性的集合  $\{\alpha_{jk}\}$ , 当对象  $j$  不具有或具有属性  $k$  时, 相应地分别有  $\alpha_{jk} = 0$  或  $1$ ; 以及

构建一个以数学方式表达的模型, 包括可识别的、因此能够被标定的参数  $\{\pi^*, r^*\}$ , 所述参数按照下述概率描述了所述探针如何依赖于所述选定的潜在属性集合:

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中,  $S_{ij}$  是对探针  $i$  做出正反应的概率, 由对象  $j$  具有及不具有所述规定属性所确定, 并且采用了关于所述剩余状态的完全性; 针对对探针  $i$  做出正反应所要求的  $m$  项属性对  $r^*$ 's 求积;  $\pi_i^* = \prod(\pi_{ik})$ , 对  $k$  求积;  $r_{ik}^* = r_{ik} / \pi_{ik}$ , 其中,  $r_{ik} = \text{Prob}$  (在所述对象不具有属性  $k$  的情况下, 对探针  $i$  做出正反应),  $\pi_{ik} = \text{Prob}$  (在所述对象具有属性  $k$  的情况下, 对探针  $i$  做出正反应)。

39. 一种方法, 包括:

构建一个其中包括若干二值记分的探针的集合, 并选择一个集合, 其中包括每个对象所具有或不具有的、观测不到的若干二值属性, 其目的是评估每一个被探测的所述对象的潜在状态;

构建一个以数学方式表达的模型，包括所述探针和所述选定的属性，所述选定的属性是一个更大的属性组的子集，所述更大的属性组中的属性可以影响对象对探针的反应，所述更大的属性组中未规定的剩余部分，通过剩余状态参数的形式，在所述模型中被体现；所述模型中包括参数，所述参数描述了所述探针如何依赖于所述选定的属性集合，也描述了所述探针如何依赖于所述剩余状态参数，以及对所述对象具有或不具有哪些属性的预测；所述模型还解释了如下事件的概率，即：对每一个具体探针而言，一个对象可能具有对所述具体探针做出正反应所需的全部属性，但未能在所述的具体探针上适当地应用至少一项必需属性，从而对所述具体探针做出了负反应，其中，所述对具体探针做出正反应所需的全部属性是所述选定的属性集合的一个子集中的；以及如下事件的概率：对每一个具体探针而言，一个对象可能未具有对所述探针做出正反应所需的规定属性中的至少一项，但仍然在所述探针上适当地应用了所述未具有的、规定的必需属性，同时，也适当地应用了所述选定属性集合中的、未规定的剩余属性，从而对所述具体探针做出了正反应；所述模型将具有每种属性的水平定义为一个设定的水平，所述设定的水平被判断为可令对象具有所述的具体属性；

将根据所述对象的反应而获得的综合探测结果用于标定所述模型的具体探针；以及得到具有所述属性的预测，或者不具有所述属性的预测，或者针对每一个由对象和规定属性所构成的组合不做出任何预测。

40. 如权利要求 39 所述的方法，包括：

构建一个集合，其中包括若干二值记分的探针，当对象  $j$  对探针  $i$  做出负反应或正反应时，相应地分别有  $X_{ij} = 0$  或 1；选择一个潜在属性的集合  $\{\alpha_{jk}\}$ ，当对象  $j$  不具有或具有属性  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或 1；以及

构建一个以数学方式表达的模型，包括参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述探针如何依赖于所述选定的潜在属性集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中， $S_{ij}$  是对探针  $i$  做出正反应的概率，由对象  $j$  具有及不具有所述规定属性所确定，并且采用了关于所述剩余状态的完全性；针对对探针  $i$  做出正反应所要求的  $m$  项属性对  $r^*$ 's 求积； $\pi_i^* = \prod(\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在所述对象不具有属性  $k$  的情况下，对探针  $i$  做出正反应)， $\pi_{ik} = \text{Prob}$  (在所述对象具有属性  $k$  的情况下，对探针  $i$  做出正反应)，其中， $p_k$  是由使用者确定的、被判断为具有属性  $k$  的对象的比例。

41. 一种方法，包括：

构建一个其中包括若干二值记分的探针的集合，并选择一个集合，其中包括每个对象所具有或不具有的、观测不到的若干二值属性，其目的是评估每一个被探测的所述对象的潜在状态；

构建一个以数学方式表达的模型，包括所述探针和所述选定的属性，所述选定的属性是

一个更大的属性组的子集，所述更大的属性组中的属性可以影响对象对探针的反应，所述更大的属性组中未规定的剩余部分，通过剩余状态参数的形式，在所述模型中被体现；所述模型中包括参数，所述参数描述了所述探针如何依赖于所述选定的属性集合，也描述了所述探针如何依赖于所述剩余状态参数，以及对所述对象具有或不具有哪些属性的预测；所述模型还解释了如下事件的概率，即：对每一个具体探针而言，一个对象可能具有对所述具体探针做出正反应所需的全部属性，但未能在所述的具体探针上适当地应用至少一项必需属性，从而对所述具体探针做出了负反应，其中，所述对具体探针做出正反应所需的全部属性是所述选定的属性集合的一个子集中的；以及如下事件的概率：对每一个具体探针而言，一个对象可能未具有对所述探针做出正反应所需的规定属性中的至少一项，但仍然在所述探针上适当地应用了所述未具有的、规定的必需属性，同时，也适当地应用了所述选定属性集合中的、未规定的剩余属性，从而对所述具体探针做出了正反应；所述模型针对所述选定属性集合的某些属性对，表达了每一个属性对的两个成员之间的联系，所述联系可能是正的或负的；还表达了对每一个属性对的所述正的或负的联系大小的估量，每一个属性对的所述联系可以由所述对象对所述具体探针的反应而估计出来；

将根据所述对象的反应而获得的综合探测结果用于标定所述模型的具体探针；以及得到具有所述属性的预测，或者不具有所述属性的预测，或者针对每一个由对象和规定属性所构成的组合不做出任何预测。

42. 如权利要求 29 所述的方法，包括：

构建一个集合，其中包括若干二值记分的探针，当对象  $j$  对探针  $i$  做出负反应或正反应时，相应地分别有  $X_{ij} = 0$  或  $1$ ；选择一个潜在属性的集合  $\{\alpha_{jk}\}$ ，当对象  $j$  不具有或具有属性  $k$  时，相应地分别有  $\alpha_{jk} = 0$  或  $1$ ；以及

构建一个以数学方式表达的模型，包括参数  $\{\pi^*, r^*\}$ ，所述参数按照下述概率描述了所述探针如何依赖于所述选定的潜在属性集合：

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}}$$

其中， $S_{ij}$  是对探针  $i$  做出正反应的概率，由对象  $j$  具有及不具有所述规定属性所确定，并且采用了关于所述剩余状态的完全性；针对对探针  $i$  做出正反应所要求的  $m$  项属性对  $r^*$ 's 求积； $\pi_i^* = \prod(\pi_{ik})$ ，对  $k$  求积； $r_{ik}^* = r_{ik} / \pi_{ik}$ ，其中， $r_{ik} = \text{Prob}$  (在所述对象不具有属性  $k$  的情况下，对探针  $i$  做出正反应)， $\pi_{ik} = \text{Prob}$  (在所述对象具有属性  $k$  的情况下，对探针  $i$  做出正反应)；间接表达了对每一个属性对正联系大小的估量，正如二值属性  $\alpha_k$ ， $\alpha_{k'}$  的连续二变量正态属性前身  $\alpha'_k$ ， $\alpha'_{k'}$  之间的相关性  $\sigma_{kk'}$  所确定的那样，然后，所需的二值属性对  $\alpha_k$ ， $\alpha_{k'}$  可以通过将每一个  $\alpha'_k$  在某一具有水平分界点处进行分界而获得。

## 潜在属性的诊断方法

### 技术领域

本发明提供了一种方法，通常用于对对象的潜在属性进行认知的、医学的及精神病学的诊断，所述的对象通常是人，其中，采用了对对象进行二元计分检验(binary scored probing)的手段。

### 背景技术

#### 第1部分：描述现有技术所必需的背景知识

##### 目前正在使用的标准化测试；所定义的认知诊断

在描述与本发明有关的现有技术之前，有必要先讨论一下所需的背景材料。在大规模标准化测试和课堂测验中，通常利用考试分数在一个单一的尺度上(on a single scale)对受试者进行排名和/或找到受试者的位置。这一尺度通常被解释为在某一具体领域的能力或成绩，例如代数或运动物理学。事实上，下面的两种广泛用于给标准化考试“计分”的方法，也就是说，古典考试理论(classical test theory)(Lord, F.和 Novick, M., 1968, *Statistical Theories of Mental Test Score*(智力测试评分的统计理论), Reading, Massachusetts, Addison Wesley-虽然这是一本很古老的书，但仍然是古典考试理论里的权威)和“一维”(unidimensional)考题反应理论(item response theory, 以下简称为IRT)，都给每一个受试者赋以一个单一的考试分数。

“考题”仅仅是用来表示测试题目的一个术语。标准化考试中的得分，通常是考试中答对了的数目，但是，在确定得分的时候，也可能包括对某些考题不是全部记分(partial credit)，或者给某些考题赋以比其他考题更大的权重。在课堂测验中，教师通常也给每次考试赋以一个分数。

对于这一给考试赋以单个分数的方法，其结果是，考试或者只能用于对受试者进行排名，或者只能用于确定受试者对该次考试所涉及内容的整体掌握水平，后者的前提是掌握程度的标准(mastery standard)已确立。具体地说，它不能用于在一个单独的考试范围内，产生受试者“认知技能”(cognitive attributes)的精细图解(profile)。也就是说，一次代数考试可以用于评估 John 相对于其他人的整体代数技巧的水平，或者相对于代数掌握标准的整体代数技巧水平，但它不能用于确定认知技能的掌握水平，例如，John 是否能很好地分解多项式，理解指数的规则，理解二次方程(quadratic formula)等，虽然此类精细的分析很显然是为教师、学生、家长、学校和政府机构等所期望的。

此处，认知诊断(cognitive diagnosis)是指，提供受试者认知技能掌握/未掌握的精细图解。

#### 统计学方法或分析



形成本发明核心的认知诊断算法，是一种具体的统计学方法。统计学方法或分析，是将收集到的数据和一个产生这些数据的真实世界场景(setting)的适当概率模型结合在一起，来进行推断(下结论)。此类推断常常会导致实际的决策。例如，认知诊断表明，Tanya 对二次方程掌握的不好，那么接下来，可能就会给她进行补习，以改善她对二次方程的理解。

为了清楚地解释什么是统计学方法，可以举一个很简单的、非认知性的例子，以助于理解。作为背景，值得注意的是，统计学方法的一个很有价值的方面是，它们很明确地指出了其推断中的内在误差或不确定性。具体地说，站得住脚的统计学分析会很谨慎地避免做出那些超出了基于数据中所能获得的信息而能合理成立的范围的推断，与此相伴，会包括一个对于与该推断相关的不确定性的估计，例如提供标准误差，一个基本的统计学概念。因此，这使得用于认知诊断的任何统计学方法，优于任何基于确定性模型(deterministic model)的方法(分别被称为基于规则的方法，人工智能，数据挖掘等，取决于所采用的具体确定性方法)。

下面将通过一个简单的场景解释一下确定性推断和统计学推断的不同之处。有一枚硬币，据称在抛掷的时候，出现正面朝上的机会更多。将其抛掷 10 次，结果有 7 次是正面朝上。非统计学的确定性方法仅仅简单的报告说，推断出来的正面朝上的概率  $p$  为 0.7，因此下结论说上述推断是正确的，而所述的确定性方法有其内在的缺陷，不能指出可能的推断误差或不确定性。而统计学方法报告说，虽然正面朝上的概率  $p$  最有可能的值确实是 0.7，但是，由于所获得的数据非常有限，因此该推断具有不确定性，所以，能较为可信的预测出来的全部内容是  $0.348 \leq p \leq 0.933$ 。因此，从统计学推断的观点来看，没有很强的证据表明这枚硬币是不均一的。这种适当谨慎的统计学观点是一种具有优势的处理方式。

相似地，后面将要讨论的使用统一模型(Unified Model, UM)的认知诊断，只有当受试者的考试数据提供了充分的证据支持的时候，才会针对某一项具体技能给一个受试者下结论，确定技能已掌握或技能未掌握，例如，Jack 已掌握指数的运算规则。

下面，将给出一个统计学方法的非认知性实例，比上述解释更为详尽。

**实例 1** 将一种药物给 40 名患者服用，其中，这种药物的治愈概率  $p$ (值在 0 和 1 之间)未知。结果有 30 名患者被治愈。假定为标准二项概率模型(standard binomial probability model)(也就是说，假定患者之间均相互独立地做出反应，并且治愈每个患者的概率是相等的)。基于这一模型和上述数据，由二项概率模型的数学性质，可以得出如下统计学推断：实际治愈率  $p=0.75$ ，置信度是，这一估计的误差小于  $\pm 0.14$ 。因此，基于有限的的数据，可以得出的推断是， $p$  位于区间(0.60, 0.89)之间。与此形成对比的是，如果在这项药物测试中，有 400 名患者参与(也就是说，更多的数据)，并且其中 300 名被治愈，那么，可以像上面一样推断  $p=0.75$ ，但现在的置信度更高，估计误差小于  $\pm 0.04$ 。更多的数据提供了更高的置信度，即这一推断固有的不确定性很小。

教育测量(Educational Measurement)，考题反应理论(Item Response Theory) (IRT)，以及对教育测量/基于 IRT 的认知诊断模型的需求。目前，在教育考试数据的概率建模方面，占统治地位的范例是考题反应理论(Embretson, S. 和 Reise, S. (2000) *Item Response Theory*

for Psychologists(心理学者使用的考题反应理论). Mahwah, New Jersey, Lawrence Erlbaum)。这一理论中，将答对某一考题的概率设定为某一假定的潜在(观察不到的)能力变量的函数，所述的潜在能力变量总是被解释为相对较宽泛的、粗线条(coarse-grained)的能力，例如代数能力。假定不同受试者所具有的这种潜在能力的水平不同。由于水平越高，答对考题的概率越大，因此，称这种潜在变量为“能力”是合理的。图 1 示出了考题的标准逻辑考题反应函数(standard logistic item response function)(IRF)，是能力 $\theta$ 的函数。每一个此类函数给出  $P(\theta)$  = 具有能力 $\theta$ 的典型受试者答对一道题的概率。

典型地，如此处所述，受试者能力的尺度是，能力小于-2 表示能力很差的受试者(最低的 2.5%)，0 表示能力一般的受试者，大于 2 表示能力很高的受试者(最高的 2.5%)。目前，在目的为对受试者的潜在能力水平进行统计学评估的教育测量中，正在大量使用基于 IRT 的统计学方法。

教育测量是统计科学的应用，利用概率模型和统计学方法，分析教育数据(通常是考试数据)，以提供关于学习过程和不同教育方法(setting)的信息，并评估个体水平和群体水平(州，教学区，国家等)上的智力表现。

接纳了教育测量重点内容的一个现代发展是，努力发展考试设置(setting)的新测量模型，以便允许人们可以通过对考试数据的统计学分析，而对受试者进行认知诊断。正如前面已经指出的，认知诊断是指一种相对较为精细的分析，可以根据每一个受试者具备或缺乏一个宽泛的学科领域内的哪些具体技巧(通常称为技能)，而对受试者进行评价(参见 Frederiksen, N., Glaser, R., Lesgold, A.c,和 Schafto, M., 1990, *Diagnostic Monitoring of Skill and Knowledge Acquisition*(技巧和知识获取的诊断监控). Mahwah, New Jersey, Lawrence Erlbaum; 和 Nichols, P., Chipman, S., & Brennan, R., *Cognitively Diagnostic Assessment*(认知诊断评估), 1995, Erlbaum, Hillsdale, New Jersey, 可以找到与现代认知诊断有关的一系列编辑好的文章)。受试者的这两种状态被称为，掌握(具备这种技能)和未掌握(缺乏这种技能)。以代数为例，回忆一下上面所给出的有关代数技能的不完全列表：分解因式、二次方程等。认知诊断的重点在于，基于这些具体的代数技能对受试者进行评估，而不是仅通过受试者的考试表现而赋以一个代数考试的分数。例如，基于考试表现，一个受试者可以被判断为已经“掌握”二次方程，但尚未掌握分解因式。很显然，此类认知诊断能力对于标准化考试和教学环境中所进行的考试(例如教室中进行的考试，以及基于网络的远程教学课件中所进行的考试)而言，具有很大的实用价值。

**实例 2** 对认知诊断的需求。发明人之一是一名教师，负责讲授一门大学水平的基础统计学课程，他针对课本前三章的内容进行了一次考试。设计的考题代表了前三章中所讲授的不同概念。它应该不仅能通过考试分数评价学生，特别是还应该能评价他们对所讲授概念的理解程度。在考题设计出来之后，编辑出了一个包括 8 个概念或技能的列表：(1)直方图，(2)中数/四分位数，(3)平均值/均值，(4)标准差，(5)回归预测，(6)相关，(7)回归线，(8)回归拟合。正如预期的那样，在有些考题中，每道题里涉及了不只一项技能。在这项有 40 道题的

考试中，每一项技能平均出现在 6 道考题中。在技能的水平上，而不是利用总分，对考试进行评价，将会有助于确定学生还应该再复习哪些内容，而确定这些内容是很有必要的；而这也会有助于每个学生判断他/她应该学习些什么。在后面的具体实施方式部分中，这一实例被发展成了本发明的一个模拟实施例。

尽管认知诊断，一个困难的应用领域，对于社会具有很明显的潜在价值，它还是以很慢的速度正浮出水面。由认知科学家/心理学家以及计算机科学家所发展出来的用于学术目的的数学模型，其设计目的不同于认知诊断想要达到的目的，也就是说，详细理解智力认知过程如何发生，通常也包括它如何随着时间进行演化(学习)。因此，这类模型从本质上讲，并不适合于认知诊断的目的。它们是确定性的，而且在参数上很复杂，由于这两个原因，当它们用于对使用了简单计分的考题的典型考试设置(setting)进行认知诊断时，通常表现很差，其中，在典型考试设置中，数据量很有限，而且这些数据很显然会发生随机变化。仅仅由于一名受试者被判断为已掌握答对一道题所需的主要相关技能，并不意味着这名受试者就会实际上答对这道题。类似地，没有掌握某一项必需的主要相关技能，并不能确保一名受试者就会答错这道题。

**引入的正性** 与确定性认知模型所预测出来的东西不一致的，是被称为正性(positivity)的东西。而这仅仅是承认某种随机结构的测量模型的一个方面，其中，所述的随机结构将技能的掌握和在答题时正确运用所掌握的这项技能联系起来。例如，Towanda 可能被判断为已经掌握指数的运算规则，但她仍有可能没有正确地将其对指数的理解应用到一道考题上，因为 Towanda 试图解答的这道题，可能对指数运算规则掌握程度的要求特别高，而且事实上，这一要求已经超过了 Towanda 所具有的技能，即使她已经掌握了指数运算规则这一技能。

**数据的过度拟合：利用确定性模型进行推断的致命缺陷** 前面已经讨论过，由于忽略了数据的固有不确定性，确定性模型可能会走得太远，以致超出了数据中所能获得的信息，从而造成“过度预测”(over-predicting)。具体地说，正如最近的连环漫画 Dilbert 中所讽刺的，此类确定性“数据挖掘”模型，由于它们过度预测的倾向，可能会倾向于从数据中发现一些表面上看来很系统的、因此值得报告的模式(pattern)，而这些模式仅仅是随机噪音所致的偶然事件，因此不能代表任何真实的东西。具体地说，基于它们的预测，在用于新的类似数据集时常常不再成立，因此是不可靠的，也是危险的。统计学家将这种看到数据中的随机噪音并推断数据中的系统“信号”或模式的现象，称为数据的过度拟合。此类过度拟合是推断过程中没有将不确定性水平的信息包括进来的直接结果。

将前面所讨论的扔硬币的简单实例进行一点改变，可以帮助解释过度拟合问题。将一枚两面可能不均一的硬币扔 4 次，结果有 4 次是正面朝上。最简单的过度拟合的确定性方法可能会下结论说，这枚硬币总是会正面朝上，因此预测说，新的扔硬币试验的预期模式仍然是总会得到正面朝上。但是，随机统计学方法仅仅下结论说，所有能得到的推论是，正面朝上的未知概率在区间(0.4, 1)之间。从这种适当谨慎的观点来看，这枚硬币实际上很有可能是两

面均一的!

UM 是一种统计学方法, 因此, 可以避免数据的过度拟合, 只有当有充分的证据支持时, 才会对受试者是否掌握了某项技能做出预测, 这一点是很重要的。本发明正是部分地基于 UM。

广泛应用的随机“一维”IRT 模型, 在数学上和统计学上都较易处理, 因此能够通过它们的随机性质(具体地说, 不会对数据进行过度拟合)而恰当地处理随机的受试者变化, 但是很遗憾, 此类模型在参数上过于简单化, 因此不适合用作从理论上支持精细认知诊断的工具。也就是说, 此类模型是在粗线条的能力水平(例如, 在基础统计学上的能力)上对能力进行处理, 因此, 不能在精细的认知技能能力水平(例如, 是否掌握解释直方图、计算均值等)上进行处理。

有一种新的、很有前景的努力, 试图将确定性认知科学传统和随机测量/IRT 传统结合起来, 以获得易处理的、现实的随机认知诊断模型, 可以在认知技能的水平上工作。此类新模型远比标准 IRT 模型复杂得多。但是, 它们又比上面所讨论的经典确定性认知科学模型简单得多。特别地, 它们避免了对数据的过度拟合。UM 就是此类新的、较为复杂的概率模型中的一种。

## 第 2 部分 对现有技术的描述

第一个以认知为导向的、在 IRT 传统中发挥功能的测量模型, 可能是 Gerhardt Fischer 的线性逻辑模型(Fischer, G (1973) *Linear logistic test model as an instrument in educational research*(作为教育研究手段的线性逻辑测试模型), *Acta Psychologica*, 37, 359-374)。它只具有历史上的意义, 这是因为从本质上, 它实际上不能用于受试者考试数据的认知诊断。但是, 到现在为止, 已经有了几种重要的基于 IRT 的模型, 其重点在于对考试反应进行认知建模, 其中的每一种都构成了现有技术。具体地说, 下面将要详细描述的是 Kikumi Tatsuoka、Robert Mislevy、Susan Embretson 以及 Brian Junker 的统计学模型, 是从现有技术角度讲较为相关的实例。此外, UM 的一个较早的、原始的、不完整的、不可使用的版本, 出现在了 DiBello, L, Stout, W, 和 Roussos, L, 1995, *Unified Cognitive Psychometric Assessment Likelihood-Based Classification Techniques*(统一认知心理学测量评估的可能性为基础的分类技巧); 以及 Nichols, 等人的 *Cognitively Diagnostic Assessment*(认知诊断评估). Mahway, New Jersey, Lawrence Erlbaum 里面, 本发明也利用了这一版本。从现有技术的角度以及便于理解目前的 UM 的角度来讲, 这一版本都是很重要的。非随机性(确定性)认知模型有很多, 并且是高度专门化的。在方法上它们与 UM 有很大差别, 很不适合用于实际的认知诊断。

DiBello 等人所提出的构成现有技术的 UM 程序。1995 年版本的 UM 是现有技术中最相关的实例。

图 2 中的流程图解释了 DiBello 等人所提出的 UM 认知诊断(UMCD)程序。其中的一些

元素与本发明中的现行 UMCD 算法是相同的。本发明使用了 DiBello 等人所提出的 UM 方法的改进和变型。作为背景，它假定有  $i = 1, 2, \dots, n$  道考题， $j = 1, 2, \dots, N$  名受试者，以及  $k = 1, 2, \dots, K$  项技能。

执行考试的结果是受试者反应数据矩阵

$$X = \{X_{ij}\}.$$

此处， $X$  是随机的，反映了如下事实，即考试的执行被作为受试者的随机取样而进行建模，而受试者对考试中的一套考题是随机反应的。 $X = x$  是进行一次实际考试的结果，产生了观测数据  $x$  (框 207)。因此， $x$  是一个由若干 0 (0 代表考题/受试者组合的不正确反应) 和 1 (1 代表考题/受试者组合的正确反应) 构成的  $n$  乘  $N$  的矩阵。第  $j$  列代表一个具体的受试者  $j$  对这  $n$  道考题的反应。例如，如果有两个受试者参加了一项有三道题的考试，那么， $x$  可能是

$$\begin{array}{cc} 1 & 0 \\ 1 & 1 \\ 0 & 0 \end{array}$$

这表明，第一个受试者答对了前两道题，第二个受试者只答对了第二道题。

应当注意，一般的科学模型以及具体的概率模型中的参数，是该模型中的未知量，必须根据该模型在每一种具体应用环境下的数据通过统计学方法而加以确定，该参数的值也会随着具体应用环境的不同而变化。具有  $n$  道考题、 $N$  名受试者的 UM 的参数，通常表示为  $\omega$ ，可由下式给出

$$\omega = (\alpha, \theta; r, \pi, c)$$

其中， $(\alpha, \theta)$  是受试者参数， $(r, \pi, c)$  是考题参数，后者有时也称为考试结构。受试者参数通常标以下标  $j$ ，表明它们是第  $j$  名受试者的参数；考题参数通常标以下标  $i$ ，或同时标以下标  $i$  和  $k$ ，表明它们属于第  $i$  道考题，并且可能是特定于技能  $k$  的。下面将详细地解释  $\omega$  的每一个参数。图 2 中的流程图用图解的方式概略地介绍了一个人如何使用 DiBello 等人的 UM 进行认知诊断(此类诊断对于 DiBello 等人 1995 的 UM 来说是不可能的)。事实上，统计认知诊断程序通常与图 2 有很多共同之处，其中一个实质性的区别通常是概率模型  $f(X|\omega)$  如何构建。

DiBello 等人所提出的 UM 的基本概念，将通过频繁地参考图 2 和图 3 而加以解释。作为对认知诊断场景(setting)常见规模的解释，在我们对课堂上进行的统计学考试进行诊断实践时，共有  $N=500$  名受试者，可以将其视为在一所规模较大的大学里选修一门基础统计学课程的学生总数的近似值。在后面的具体实施方式部分中，这一实例被发展成了一个模拟实

例，它证明了本发明认知诊断的有效性。考试中有  $n=40$  道考题，测试了课程所用的教科书前三章中的统计学概念。假定不同的考题要求这  $K$  项技能的不同组合。在我们的实例中， $K=8$ ，即这次统计学考试中所测试的主要概念的数目。

回想一下，“技能”(attribute)是一个宽泛的术语，用于表示可以被判断为掌握或未掌握的任何一组知识。用于构建考题/技能关系矩阵(incidence matrix)(图 2 中的框 205)的选定的技能(图 2 中的框 201)，由该算法的使用者进行定义，并且可以是该使用者所希望的任何东西。实际上，使用者可以不受关于学习和/或精神过程的任何具体认知理论的限制而自由选择所述技能，正是 UM 的一个真正优势。也就是说，与用于认知诊断的许多其他方法不同，UM 允许使用者选择基于任何有关学习、精神机能(mental functioning)或认知的概念化内容的任何技能，甚至是一个讲授典型学校课程的教师所能理解的非常不正式的结构，而所述的“许多其他方法”，则包含并且因此依赖于对认知精神过程的一种具体理论的理解和接受。 $N$  名受试者中的每一个人都具有  $K$  项技能，因此， $\omega$  的  $\alpha$  部分是一个  $N$  乘  $K$  的矩阵。此处， $\alpha$  的每一行对应于一个单独的受试者，并且具有  $K$  个元素(0 和 1)。0 表示受试者没有掌握某项技能，而 1 则表示受试者已经掌握某项技能。

基于 UM 模型的认知诊断的目的是，利用从考试中获得的可以利用的考试数据  $x$ (图 2 中的框 207)，来对每一个受试者进行推断(图 2 中的框 213)，以得到下述信息：这  $K$  项技能中的哪些技能已有充分的证据表明她已掌握，以及这  $K$  项技能中的哪些技能已有充分的证据表明她没有掌握(应当注意，对于每一个受试者来说，有可能会缺乏足够的证据来证明某些技能是否已经被其掌握或尚未被掌握)。

启动提出的 UM 算法所需的输入数据由两个数据文件构成，这两个文件可以相对容易地被理解和产生，并不需要使用者具有对认知科学的深入理解，与其他现有技术相比，这是 UMCD 的一个优点。首先，针对每一道考题选定了一个技能列表，其中给出了为正确解答这道题所必须同时掌握的技能(图 2 中的框 201)。使用者/从业人员(practitioner)通常先确定要在具体的教育环境(setting)下对哪些技能进行认知诊断，然后构建所需要的考题(图 2 中的框 203)。有时候，使用者首先构建好考题，然后选择有待进行诊断的技能。

然后，使用者要确定每道考题都需要这些技能中的哪几项技能，从而生成  $n$  乘  $K$  的考题/技能关系矩阵(图 2 中的框 205)。在“具体实施方式”部分描述的图 18 中给出了用于统计学考试诊断实例的考题/技能关系矩阵的一个实例。

应该强调，基于 UM 的诊断算法的使用者，例如学区课程专家或大学教师，典型地进行图 2 的框 201、203 及 205 中的行为，也就是说，选择技能，设计考题，以及构建考题/技能关系矩阵。特别地，使用者通常选择相关的技能，并设计用于检测这些技能的考题(顺序也可以反过来)，然后针对每一道考题，确定答对这道题需要这些选定的技能中的哪几项。使用者的这一相对简单的活动，可以通过向具有 UMCD 知识的顾问咨询或参考 UMCD 的使用指南而获得帮助，UMCD 的使用指南中，给出了使用 UMCD 程序时如何更好地进行认知诊断考题的设计、技能的定义以及关系矩阵构建的基本原则。

作为考题/技能关系矩阵的一个例子，考虑三道考题和四项技能。那么关系矩阵

		技能			
		0	1	1	0
考题	1	0	0	0	0
	0	0	1	1	0

确定了考题 1 需要技能 2 和 3，考题 2 需要技能 1，考题 3 需要技能 3 和 4。

第二步，基于对受试者进行的考试，受试者反应数据由一个记录构成，该记录反映了每个受试者答对了哪几道题，答错了哪几道题。这可以通过下式用符号表达出来：

$$X_{ij} = \begin{cases} 0, & \text{如果受试者 } j \text{ 答错了考题 } i; \\ 1, & \text{如果受试者 } j \text{ 答对了考题 } i; \end{cases}$$

例如，考虑两名受试者对四道考题作答时的考试反应。

受试者 1 的反应：0 0 1 1

受试者 2 的反应：1 0 0 1

这表明，受试者 1 答对了考题 3 和 4，受试者 2 答对了考题 1 和 4。正如前面已经指出的，所有这些  $x_{ij}$  反应被集中到一起，以形成反应考试数据受试者反应  $x$  矩阵(the matrix of responses test data examinee responses  $x$ )。

回忆一下，对于每一个受试者， $\alpha_j$  表示长度为  $K$  的潜在(未知)向量，该向量针对这  $K$  项技能中的每一项，指出了受试者已掌握(由 1 表示)还是未掌握(由 0 表示)。例如：

$$\alpha_j = (1, 0, 1, 1, 0)$$

表示受试者  $j$  已经掌握技能 1、3、4，尚未掌握技能 2 或 5。推断每一个受试者的  $\alpha_j$  是什么，正是认知诊断的目标。

图 2 中的框 209，在关系矩阵构建好(图 2 中的框 205)之后进行，其任务为构建概率模型  $f(X|\omega)$ ，回忆一下， $\omega = (\alpha, \theta, r, \pi, \epsilon)$  表示具有  $n$  道考题、 $N$  名受试者的模型里的考题和受试者参数。这个框里的内容是 UM 的核心，为了便于理解这个框的内容，必须介绍几个技术概念。图 3 中给出了一个考题/受试者反应  $X_{ij}$  的 UM 概率模型的示意图，此处，参考一下图 3 是很有帮助的。

图 3 中已经部分指出的 DiBello 等人的 UM 的基本方程式 如同所有基于 IRT 的模型一样，UM 使用了考题反应函数(item response function, 以下简称 IRF)的概念。IRF 是一个 S

形的上升曲线，其下边界为 0，上边界为 1。在通常的 IRT 模型里，这给出了答对一道题的概率，而这一概率是一个连续的潜在能力的函数，例如统计学能力，通常由  $\theta$  表示。图 1 中以图形的方式表示出了这样的 IRF。符号  $P(\theta)$  是指，具有潜在能力  $\theta$  的受试者答对这道考题的概率。UM 的公式依赖于图 1 中 IRF 的使用。

UM 的基本构成单元(图 2 中的框 209)，是要生成受试者  $j$  答对考题  $i$  的概率的表达形式，其中，受试者  $j$  具有潜在剩余能力(residual ability) $\theta_j$  和潜在技能向量  $\underline{\alpha}_j=(\alpha_{j1}, \dots, \alpha_{jK})$ ，在该式中，对于  $K$  项技能中的每一项而言，每一个元素  $\alpha_{jk}$  等于 0 或 1，根据技能  $k$  未被掌握或已被掌握而定。一名受试者对一道考题做出反应的概率模型在下面给出。

$$Prob(X_{ij}=1|\omega) = S_{ij} \times P(\theta_j+c_i), \quad (1)$$

其中，IRF 在图 1 中给出， $S_{ij}$  将在下面进行解释。此处，“ $|\omega$ ” 仅仅表示  $X_{ij}=1$  的概率是在参数值等于  $\omega$  的时候计算出来的。图 3 中给出了一幅示意图，表示产生基本方程式(1)的参数影响。由于  $X_{ij}$  可能的取值只有 1 和 0，因此，根据基本的概率逻辑关系可以得到

$$Prob(X_{ij}=0|\omega) = 1 - Prob(X_{ij}=1|\omega)$$

而且，在 IRT 中，在建模的时候认为受试者之间均相互独立地做出反应。同样也是根据局部独立(local independence)这一基本 IRT 建模原则，具有同样一套受试者参数( $\underline{\alpha}, \theta$ )值的一组受试者对不同考题所作出的反应，在建模时也被认为彼此之间相互独立。在概率模型中，一系列独立事件同时发生的概率，可以通过将其中每一事件发生的概率相乘而得到。这样，方程式(1)中的单一考题和受试者模型，就可以用于包括所有  $N$  名受试者和  $n$  道考题在内的一套数据，

$$f(x|\omega) = Prob(X=x|\omega) = \prod \prod Prob(X_{ij} = x_{ij} | \omega) \quad (2)$$

其中，符号  $\prod \prod$  表示在  $i$  和  $j$  的取值范围内取乘积，也就是说，当  $j$  从 1 到  $N$  时的外部积(outer product)，以及  $i$  从 1 到  $n$  时的内部积(inner product)。应当强调，正是对不同受试者和不同考题而言  $X_{ij}$  反应的独立性，使得方程式 2 中所给出的基本 UM IRT 模型中的二重积(double product)得以成立。此外， $x_{ij}$  表示  $x$  的第  $i, j$  个成员，其值或者是 1，或者是 0，根据受试者  $j$  答对还是答错考题  $i$  而定。

**核心 UM 概念：正性和完全性** 方程式 1 和 2 中包括了 UM 的精髓，为了便于理解方程式 1 和 2，应当注意 UM 假定了两个很重要并且很有用的认知学概念，也就是说，正性(positivity)和完全性(completeness)。方程式 1 中的第一个因子  $S_{ij}$  建立了正性模型，而第二个



因子  $P(\theta_j+c_i)$ 建立了完全性模型。

事实上，完全性的引入使得 UM 在认知诊断模型中独树一帜，其中，完全性是通过第二个因子中的连续(或者简单地说，可以使用多个取值的离散变量  $\theta$ )潜在变量  $\theta$  进行建模的。此外，UM 之中完全性和正性这两个基本概念的结合，也使得 UM 与所有其他基于 IRT 的认知诊断模型相区别，而完全性和正性的结合，可以从方程式 1 中这两个因子的相乘看出来。现在将对方程式 1 和 2 进行解释。

**完全性** 首先考虑方程式 1 中的第二个因子  $P(\theta_j+c_i)$ ，它建模了考题  $i$  的完全度，以及 UM 中指定的技能。参数  $c_i$  是完全性参数，随着考题的不同而变化。在生成 UM 方程时，UM 的一个核心方面是：为了将每道考题的参数总数相对于可用数据集(data set)的大小而言，保持在一个合理的、因而在统计学上易于操作的数目，而有意地明确忽略了对那些为数众多、次要但很有影响的潜在技能所起到的作用进行建模。有影响的技能是指，该技能的掌握与否可以改变答对这道题的概率。当这些有影响但较为次要的技能被忽略时， $c_i$  量化了这些被忽略的技能的相对综合的影响，与此形成对照的是，在受试者对考题  $i$  作出反应时，被明确建模的技能  $\alpha$  的综合影响。

为了精确，假定前面所讨论的针对大学统计学考试的准确、完全(从包括了所有实际上影响受试者答题表现的技能这一角度而言)的认知诊断模型(例如，一个对基础科学的兴趣比对进行实际认知诊断的兴趣更大的认知科学家，在对选修这门大学基础统计学课程的许多学生进行了深入、详尽的认知心理学研究后，可能会得到的模型)中，包括 200 项技能。假定为了统计分析的方便，在可用的受试者数据较为有限以及这一考试中只有 40 道考题的情况下，该模型被限制，使得在 UM 的关系矩阵中明确具有 8 项技能。这样，就选定了 8 项技能，这 8 项技能被认为在确定受试者的考试表现上很重要，包括了教师想要进行认知诊断的所有技能。那么， $\theta_j+c_i$  的作用为，将对于受试者  $j$  和考题  $i$  而言较为次要的、被忽略的 192 项未选中的技能的影响保守地代码化。为了清楚起见，应当注意，在实践当中，人们几乎不了解这些被排除掉的次要技能到底有多少以及它们是什么。也就是说，为了构建一个 UM，使用者并不需要弄清楚在一次考试中所有的这些次要技能到底是什么，与传统的认知建模相比，这是一个很大的优点。

应当注意，剩余能力  $\theta_j$  体现了对这 192 项被排除的技能而言，受试者  $j$  所具有的基于技能的综合能力。允许  $\theta$  保守地“吸收”这 192 项次要技能的影响这一建模技术，是 UMCD 方法优于其他基于 IRT 的认知诊断方法的主要原因之一。

那么，对一道考题而言， $c_i$  的作用是，在确定受试者对考题的表现时，将考虑在内的主要技能  $\alpha_j$  按比例划分其相对于被排除在外的、次要但仍有影响的技能的相对重要性，这些次要技能通过  $\theta_j$  被纳入 UM。

按照 IRT 建模中的常规做法，假定  $\theta$  是一个标准正态随机变量(著名的“钟形”曲线)，如图 4 所示。

从图 4 中可以注意到，所有受试者中，大约 2/3 受试者的能力在 -1 到 +1 之间，而全部受

试者的能力都在-3 和+3 之间。因此, 举例来说, 合成值(composite) $\theta$  代表这 192 项被排除的技能, 一名  $\theta=0$  的受试者在这些被排除的技能上具有平均水平的总能力, 而一名  $\theta=2$  的受试者在这些被排除的技能上具有很高水平的总能力。

采用下面的方式, 通过  $c_i$  对一道考题(i)的完全度进行量化。对于某些考题而言,  $c_i$  会很大(例如  $c_i=2.5$ ), 表明对大多数受试者而言,  $P(\theta+c_i)\approx 1$ (这一点可以从图 1 中的 IRF 看出, 其中,  $P(\theta+c_i)\approx 1$  明显成立(hold), 除非一名受试者的  $\theta$  异常的小), 因此, 完全性成立, 受试者在这些考题上的表现基本取决于正性因子  $S_{ij}$ , 而正性因子  $S_{ij}$  将 UM 模型中所包括的技能  $\alpha$  的影响, 通过概率手段明确地进行了建模。也就是说, 受试者表现基本取决于构成  $\alpha$  的重要技能(那些由使用者明确选择的技能)。在这种情况下, 对于所讨论的考题而言, 这些主要的被明确建模的技能是相对完全的。

相似地, 对于另外一些考题而言,  $c_i$  会很小(例如  $c_i=0.5$ ), 表明对大多数受试者而言,  $P(\theta+c_i) < 1$ (基本上)。因此, 正如  $P(\theta+c_i)$  的值所表达的, 由剩余能力  $\theta$  所建模的被排除在外的技能, 在影响受试者的反应上, 其作用是很重要的, 同时, 被包括在内的主要技能也很重要。在这种情况下, 对于所讨论的考题而言, 这些被包括在内并被建模的技能是相对不完全的。

由于这一点较为抽象, 同时对于理解完全性概念来说又很重要, 因此, 下面将给出一个简单的例子。考虑一名具有平均水平能力  $\theta = 0$  的受试者。假定  $c_i = 3$ , 表明这是一道很完全的考题, 受试者对于这道考题的反应行为几乎完全受被包括在内的技能的控制。然后请注意, 参见图 1, 受试者将被排除在外的次要技能正确运用到这道题上的概率, 由  $P(\theta+c_i) = P(3) \approx 1$  给出。因此, 该模型可以恰当地将受试者掌握/未掌握这些主要技能这一情况, 有效地成为受试者能否答对这道题的唯一决定因素, 如方程式 2 中的  $S_{ij}$  所示。

**正性** UM 中第二个具有重要意义的概念是正性, 在下面用于表示  $S_{ij}$  的方程式 3 中对此给予清楚的描述。这给出了模型中所列出的技能被受试者  $j$  (已经掌握了某些技能) 正确地运用于解答考题  $i$  (这要求掌握某些技能) 的概率, 具体地说, 所述技能是按照关系矩阵(图 2 中的框 205)为解答考题  $i$  所需的技能。

$$S_{ij} = [ (\pi_{i1})^{\alpha_{i1}} \times (\pi_{i2})^{\alpha_{i2}} \times \dots \times (\pi_{im})^{\alpha_{im}} [ (r_{i1})^{1-\alpha_{i1}} \times (r_{i2})^{1-\alpha_{i2}} \times \dots \times (r_{im})^{1-\alpha_{im}} ] \quad (3)$$

注意, 当  $\alpha=1$  时, 只有与它相应的  $\pi$  在  $S_{ij}$  中是一个因子(而不是与它相应的  $r$ ); 当  $\alpha=0$  时, 只有与它相应的  $r$  在  $S_{ij}$  中是一个因子(而不是有它相应的  $\pi$ )。因此,  $S_{ij}$  是  $m$  个因子的乘积, 其中每个因子是一个  $\pi$  或  $r$ 。此处应当理解, 上述公式中的  $m$  项技能是考题/技能关系矩阵中考题  $i$  所要求的技能。同样地,  $\alpha_{i2}=1$  或  $0$  分别表示受试者  $j$  已掌握或未掌握技能 2 的状态, 等等。

回忆一下方程式 1、2 和 3, 可以看出, 在确定  $f(X|\omega)$  的时候, 需要将考题/技能关系矩阵输入, 如图 2 中从框 205 到 209 的箭头所示。这是因为, 考题/技能关系矩阵针对每一道考题

$i$ , 提供了正确解答这道考题所要求的  $m$  项技能。具体地说, 出现在方程式 3 中的  $\pi$  和  $r$ , 仅仅对应于考题  $i$  所要求的技能。

**方程式 3 中正性参数  $\pi$  和  $r$  的定义**  $\pi$  和  $r$  如下所示:

$r_{ik} = \text{Prob}$  (在受试者未掌握技能  $k$  的情况下, 技能  $k$  被正确运用到考题  $i$  中的概率)

类似地,

$\pi_{ik} = \text{Prob}$  (在受试者已掌握技能  $k$  的情况下, 技能  $k$  被正确运用到考题  $i$  中的概率)。

**高正性的解释** 考题表现出高正性(high positivity)是很理想的。对于一道考题而言, 当其  $r$  合理地接近于 0 并且其  $\pi$  合理地接近于 1 时, 高正性成立。也就是说, 在高正性的情况下, 当且仅当一名受试者掌握这些技能时, 受试者能以高概率正确地运用根据考题/技能关系矩阵中这道题所要求的技能。例如, 当高正性成立时, 如果一名受试者缺乏这道考题所要求的技能中的至少一项, 那么他很有可能答错这道题。考虑在统计学考试实例中一道要求技能 1 和 3 的考题。例如, 如果一名受试者不知道如何解释直方图(技能 1), 但这道题要求通过解释直方图而正确地计算平均值, 那么这名受试者很有可能答错这道题, 即使她已经掌握了平均值(技能 3)。相反地, 一名已经掌握这两项必需技能的受试者, 则很有可能答对这道题, 其前提是  $\theta+c$  较大, 表明该受试者要么能正确运用那些(可能有很多)为这道题所要求、但被排除在模型之外的技能(也就是说, 这名受试者的  $\theta$  较大), 要么那些被排除在外的技能只起到了很小的作用(也就是说, 这道考题的  $c$  较大)。因此, 如果根据直方图计算出均值是很简单的, 例如, 如果直方图很标准, 而均值的计算也不复杂, 那么一名已经掌握平均值计算(技能 3)和直方图(技能 1)的受试者, 将很有可能答对这道题, 因为那些被排除在模型之外的技能影响很小, 因此  $c$  较大。总之, 一道高度正性并且合理完全的考题, 在判断受试者是否具备全部所要求的技能或缺少其中至少一项技能上, 将会提供很重要的信息。

现在已经将 UM 概率模型部分中的基本参数介绍完毕, 图 2 中的框 209 以及图 3 中也涉及了这些参数。如果想要了解关于正性、完全性以及  $\theta$  的作用的更多细节, 请咨询 DiBello 等人的文献。

UM 最重要并且最有用的方面之一是, 完全性和正性提供了一种自然的、保守的手段, 以便将认知受试者反应的随机本质参数化, 这一点与其他基于 IRT 的模型是不同的。向 UM 程序的使用者解释这两个概念的含义是相对容易的, 而且, 对它们在 UM 中是如何被参数化的细节进行解释也是相对容易的。

**DiBello 等人的 UM 不能被标定** 框 211、213、215 不能在 DiBello 等人 1995 年的论文中进行, 其具体原因是框 211 不能进行。然后, 这又导致接下来的两个框 213 和 215 也不能进行。框 211 不能进行的原因是, 到 1995 年为止, 与典型考试数据集的大小相比, 为了

实现对 UM 模型参数的可以接受的标定，UM 的方程式里的参数太多了(回忆一下，标定仅仅意味着通过可用的数据，对模型的参数进行估计)。尤其是，不可能先估计出受试者技能掌握/未掌握(框 213)的情况，然后利用这些信息进行认知诊断(框 215)，例如告知一名受试者他还需要进一步学习哪些技能。

本发明的优选实施例 UMCD 中所使用的上述 UM (下面的讨论仅仅是为了认知诊断上的应用而提出的；其结果与医学或精神病学上的应用相同)1995 年的 UM 程序和本发明的 UMCD 在如下方面是相同的：设计一次考试，其中包括考题和选择一组技能(框 201 和 203)，以测量受试者的水平。1995 年的 UM 程序和本发明的 UMCD 在如下方面是相同的：构建考题/技能关系矩阵(框 205)。1995 年的 UM 程序和本发明的 UMCD 在如下方面是相同的：完全性要素  $P(\theta+c_j)$ 。也就是说，1995 年的 UM 程序和本发明的 UMCD 在如下方面是相同的：用于形成考题/技能关系矩阵的选定的技能，是一个更大的技能组的子集，其中，所述技能组中的技能可以影响受试者的考试表现，所述技能组中的其余部分在 UM 中是以剩余能力参数，即完全性，的形式来说明的。1995 年的 UM 程序和本发明的 UMCD 在如下方面是相同的：正性，也就是说，模型包括若干参数，这些参数通过一个概率值而描述了考题是如何依赖于选定的这一组技能的，所述概率是：对每一道单独的考题而言，每个受试者可能掌握了这一道考题所要求的全部技能，所述的全部技能是来自选定的这组技能的子集的，但未能在这道考题上正确地运用这些必需的、已经掌握的技能中的至少一项，因此答错了这道题。类似地，对每一道考题而言，每个受试者可能未掌握这道考题所要求的至少一项技能，但仍然在这道考题上正确地运用了这些必需、但没有掌握的技能，同时，也在这道考题上正确地运用了剩余的、必需、同时也掌握了、选自选定技能组中的技能，从而答错了这道题，而这也是正性的部分定义。但 1995 年的 UM 考题参数是无法确认的(not identifiable)，而本发明 UM 的参数则可以。此外，考试的进行也是相同的(框 207)。

#### 其他现有技术：基于概率模型的认知诊断程序；确定性程序

1. 基于概率模型的程序 大部分较为重要的基于 IRT 的(因此是基于概率模型的)认知诊断程序使用认知模型的贝叶斯公式(Bayesian formulation)，有时也使用一种被称为马尔可夫链蒙特卡洛(Markov Chain Monte Carlo，以下简称 MCMC)的计算工具，来作为标定它们的计算工具。前面所介绍的 UM 程序，形成了本发明的一个核心内容，它也具有贝叶斯概率模型公式，也使用 MCMC。因此，在进一步介绍现有技术的其他内容之前，我们先解释一下贝叶斯建模、MCMC 计算以及所需要的相关概念。

统计模型的标定(calibration) 考虑这个简单模型  $y = ax$ ，其中， $a$  是一个未知参数。模型标定是指，利用可用的数据对未知的模型参数进行统计学估计，而所述的数据被看作是由该模型所产生的。必需理解，如果不进行标定，那么概率模型对于进行真实世界的推断，例如认知诊断，来说是没有用的。特别地，模型标定是必须的，因为参数模型  $y = ax$  对

于进行所期望的统计学推断，即由  $x$  预测  $y$ ，来说是没有用的，除非参数  $a$  已通过真实数据加以标定(估计)。因此，如果  $a=2$  是由所述数据所获得的准确估计值，那么，经过标定的模型  $y = 2x$  对于由  $x$  预测  $y$  来说是有用的，其前提是，这一简单的直线模型在描述所关心的真实世界场景时，是相对准确的(不偏的)。

**在复杂的真实世界场景中对新型统计建模方法的需求** 在对复杂的真实世界场景进行适当的统计学建模时，一个经常遇到的主要实际问题是，建模的现实主义(realism)要求相应较复杂、从而有很多参数的模型，而可用的数据量通常是不足的，不足以对基于如此复杂的、带有很多参数的模型的统计学推断形成可靠的支持。(一个模型中的参数越多，那么存在于这些参数的估计(也就是标定)值里的不确定性就越大)因此，在实例 1 的单参数模型中，在对药物的治愈概率  $p$  进行估计时，400 个数据点所产生的不确定性是很小的。但是，如果在一个问题中有 30 个参数，而数据点为 400 个，那么，存在于参数估计值中的不确定性水平，将很可能使该模型在进行预期的统计学推断时，变得几乎没有用处，而所述的参数估计值是标定模型所需要的。

在需要合适的统计建模和分析的很多复杂场景中，存在一种不可接受的两难局面。一方面，有限的可用数据可以用来对一个有偏差的模型(一个过于简单的、扭曲了真实情况的模型)进行很好的标定，因为有充分的数据来精确地标定其相对较少的参数。例如，如果真实世界只能通过在参数上更为复杂的四参数模型  $y = c + ax + bx^2 + dx^3$  来准确描述的话，那么在模型  $y = ax$  中，对于进行预测而言，估计  $a = 2$  是没有用的。另一方面，假设只能对准确反映了真实世界的这一模型中的四个参数进行较差的标定，因为没有足够的数据来较好地估计这四个模型参数的值。出于解释的目的，如果实际情况是  $y = 5 + 9x + 12x^2 + 4x^3$ ，而该模型通过有限的的数据被较差地标定到了  $y = 3 + 4x + 20x^2 + 2x^3$ ，那么这一标定是如此之差，以至于标定后的三次多项式(cubic polynomial)对于准确地由  $x$  预测  $y$  来说是没有用的，即使真实世界可以由一个三次多项式模型很好地加以描述。

这一两难局面，即标定得好的坏模型和标定得差的好模型，是统计学家们有时称之为方差/偏差权衡(variance/bias tradeoff)的一个具体例子。在任何一种不可接受的模型建模妥协方案下，对所关心的真实世界场景进行正确的(也就是说，利用相对无偏差的和标定得相对较好的模型)简直是不可能的。

**作为主要统计学建模技术的贝叶斯概率建模(对需要很多参数的复杂真实世界场景进行建模的实用答案)**

幸运的是，统计学最近的发展提供了一种用于解决复杂场景的概率建模中的这一具有挑战性的两难局面的方法，所述的复杂场景要求在它们的模型中有相对较多的参数。尤其是，这些发展应用到了认知诊断场景的概率建模中，而所述认知诊断场景从本质上讲是很复杂的。一旦一名从业者将参数上很复杂的统计学模型重构(recast)为贝叶斯模型，由于它们新获得的贝叶斯性质，它们可以如同具有相对较少的参数那样被很好的标定，但仍然能准确地模拟复杂场景。更具体地说，这种贝叶斯建模方法通常可以克服如下问题，即通过可用的数

据, 对一个在参数上过于复杂的模型进行可靠的标定。实际上, 贝叶斯统计分析的主要来源之一, Gelman, Carlin, Stern, 以及 Rubin 引人注目地声称(Gelman, A, Carlin, J, Stern, H, 和 Rubin D., 1995, *Bayesian Data Analysis*(贝叶斯数据分析), 伦敦, Chapman 和 Hall), “正如我们在本章中所介绍的, 利用比数据点更多的参数来拟合分级(贝叶斯)模型, 通常是明智的”。特别地, 分级的贝叶斯建模可以用于对产生考试数据的复杂场景的 IRT 建模。一篇重要的论文(Richard Patz 和 Brian Junker, 1999, *A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models*(用于考题反应模型的马尔可夫链蒙特卡洛方法的一种简单途径), *Journal of Educational and Behavioral Statistics*, 24, 146-178)有效地举例说明, 在进行复杂的 IRT 建模时, 可以使用贝叶斯方法。更准确地说, 当需要复杂的 IRT 模型时(Patz 等人), 将与简单的 MCMC 计算相结合的贝叶斯模型框架用于进行必要的贝叶斯计算, 对于分析考试数据来说是非常有效的。尝试利用考试数据进行认知诊断时也正是本发明所遇到的情况。此外, 正如前面所暗示的, 引入了巧妙构建的贝叶斯方法的 UM, 具有将隐藏在考试数据里的全部信息释放出来用于认知诊断目的的潜力。

**贝叶斯建模实例** 虽然贝叶斯概率模型的概念复杂难懂, 但是, 一个简单的实例将使如下内容的基本思想变得很清楚, 即贝叶斯概率模型是什么以及其中的统计学分析如何进行。

实例 3: 实例 1(已改进) 考虑实例 1 的药物试验场景。假定除了数据以外, 还有在先的强有力的科学实验证据表明, 未知量  $p$  的真实值满足  $0.5 \leq p \leq 0.9$ , 并且, 离治愈率 0.7 的距离越远, 在此范围内  $p$  值的可能性越小。贝叶斯方法通过对参数  $p$  设定一个先验概率分布 (prior probability distribution), 而将研究者所具有的、关于模型中不同参数值的可能性的此类概率知识加以量化。也就是说, 贝叶斯模型给模型的参数设定一个概率分布, 其中, 该分布反映了使用者在多大程度上相信(基于现有的知识和/或原来的经验)未知参数的不同值可能的情形。假定  $p$  的先验分布在图 5 中以“密度”的形式给出。

例如, 可以由图 5 得到:

概率( $0.7 < p < 0.8$ )= 0.7 和 0.8 之间的面积 = 0.4

概率( $0.8 < p < 0.9$ )= 0.8 和 0.9 之间的面积 = 0.1

因此, 虽然区间(0.7, 0.8)和(0.8, 0.9)的长度是相同的, 但未知参数  $p$  落在区间(0.7, 0.8)内的概率要远远大于未知参数  $p$  落在区间(0.8, 0.9)内的概率, 这一事实将会影响我们在估计  $p$  值时对数据的使用。更一般地说, 在  $p$  远离 0.7 向 0.5 或 0.9 移动时,  $p$  值变得越来越不可能。很明显, 与估计值  $p = 0.75$  相比, 根据这一先验概率分布估计出来的  $p$  更接近于 0.7, 而估计值  $p = 0.75$  是在未采用贝叶斯方法的情况下得出的(因此  $p$  没有先验分布, 所述先验分布用于修正由数据单独得出的  $p$  的估计值)。当用于估计模型参数时, 贝叶斯方法不允许数据集本身完全自己说了算。

**将普通概率模型转化为贝叶斯概率模型** 必须强调, 将一个带有参数的普通概率模型转化为其参数具有先验分布的贝叶斯概率模型, 相当于发展一个新概率模型, 以便延续原来

的非贝叶斯的概率模型。实际上，将一个非贝叶斯概率模型转化为贝叶斯概率模型，并不是机械的或算法上的，而更像是“行会知识”，因为它要求关于贝叶斯建模的知识，特别是关于被建模的真实世界的知识。选择一个有效的贝叶斯模型，可以对统计推断的准确性产生很大影响。

**选择先验分布** 在贝叶斯建模的很多应用中，具体地说是本发明的贝叶斯 UM 方法中，对先验分布的选择是很谨慎的进行的，以便对参数提供较多信息，而同时不会提供过多信息，后者是从对现有信息赋以超过合理范围的更多权重的意义上讲的。例如，在前面所讨论的贝叶斯实例中，图 6 给出了一个先验分布，与图 5 中所给出的分布相比，所提供的信息稍微少一些，通常被称为模糊的先验(vague prior)，因为它对最终统计推断的影响很不明显。在这种模糊先验的条件下，在非贝叶斯情况下得到的  $p = 0.75$  的推断，仅仅向着 0.7 移动了很小的距离。

最后，图 7 中的先验分布对于  $p$  的可能值而言，是完全不能提供任何信息的。正如可能会受到怀疑的，当采用贝叶斯方法时，实例 1 中的非贝叶斯推断  $p = 0.7$  实际上并未被上述的完全不能提供任何信息的先验分布所改变。

**实例 3 的继续** 现在，将继续前面所描述的实例 3 中使用三角形先验分布的贝叶斯分析。在给定这一贝叶斯概率模型以及产生 75% 治愈率的数据的条件下，贝叶斯统计分析将以正规方式估计  $p = 0.72$ (而不是非贝叶斯分析所估计的 0.75)，下面将对此进行解释。这是因为，通过所提供的先验分布，如下事实已经被包括在了推断过程中，即：象 0.75 这样的  $p$  值与更接近 0.7 的  $p$  值相比，其可能性要小得多。也就是说，目前的贝叶斯估计值  $p = 0.72$ ，是通过将实例 1 中暗示  $p = 0.75$  的非贝叶斯数据分析以及现有知识结合在一起而得出的，其中，所述的现有知识表明，象 0.75 这么大的  $p$  值与更接近 0.7 的  $p$  值相比，其可能性相对小一些。在这两个信息来源(先验的和以数据为基础的)之间以数学方式达成的贝叶斯妥协生成了折衷的贝叶斯推断  $p = 0.72$ 。

**基本贝叶斯推断的范例：图形和公式** 图 8 的流程图示出了基本贝叶斯推断的范例。和所有的统计学程序一样，它以观测数据开始(图 8 中的框 801)。

在计算上，贝叶斯推断范例如下所述。令  $X$  表示观测数据(框 801)，令  $\omega$  表示贝叶斯模型参数  $\omega$ 。框 807 指出了贝叶斯概率模型，该模型将模型参数的先验分布  $f(\omega)$ (框 803)和似然概率分布(likelihood probability distribution) $f(X|\omega)$ (框 805)结合在了一起。注意，在实践中， $X$  和  $\omega$  的维数可能都很大。那么在给定所述数据的情况下，参数的后验分布(框 809 中指出)通过下式计算出来

$$f(\omega|X) = \frac{f(X|\omega)f(\omega)}{\int f(X|\omega)f(\omega)d\omega}$$

此处,  $f(\omega) \geq 0$  是专门为贝叶斯模型创建的参数的先验分布( $f(\omega)$ 被称为密度)。对先验的选择是由贝叶斯从业者而定的, 如框 803 所示。同样地, 在方程式 4 中,  $f(X|\omega)$ 是常见的似然概率分布(参见框 805; 下面将解释似然的概念), 这也是由观测数据  $X$  对  $\omega$  进行非贝叶斯统计推断时的核心内容。参数的先验分布和似然概率分布一起, 构成了贝叶斯概率模型(框 807)。似然概率分布  $f(X|\omega) \geq 0$  表明了每一个具体参数值  $\omega$  产生观测数据  $X$  的随机机理(random mechanism), 而先验分布  $f(\omega)$ 表明了从业者在多大程度上相信每一个不同参数值。

在方程式 4 中,  $f(\omega|X)$ 表示当数据  $X$  发生时,  $\omega$  的后验概率分布。之所以称为“后验”是因为它是通过观测数据  $X$  经先验修正后的  $\omega$  的概率分布(在  $X$  之后)。所有的贝叶斯统计推断都是基于通过方程式 4 获得后验分布  $f(\omega|X)$ , 如框 811 所示。例如, 实例 3 中的推断  $p = 0.72$ , 是找到了使得后验  $f(p|40 \text{ 次试验中有 } 30 \text{ 人治愈})$ 最大化的  $p$  值的结果。

在实际进行贝叶斯分析时, 一个关键点是: 当  $\omega$  的维数较高时(也就是说, 有很多模型参数), 计算方程式 4 分母中的积分通常是很困难的, 以至于几乎不可能, 在这种情况下, 进行贝叶斯推断也是很困难的, 以至于几乎不可能。当使用本发明贝叶斯版本的 UM 对考试数据  $X$  进行认知诊断时, 将会发现解决这一计算问题是很重要的。

**使用马尔可夫链蒙特卡洛的贝叶斯分析方法** 由于基于 MCMC 模拟的计算方法在迅速成熟, 因此, 带有很多参数的复杂贝叶斯模型的使用, 已经成为实用统计学推断的合理基础。假定数据集是由此类贝叶斯模型所产生的, 那么在对所述数据集进行统计分析方面, MCMC 是一种优秀的计算工具, 因为它允许不计算分析计算方法中所要求参数的复杂后验分布(方程式 4)。特别地, 本发明所使用的具体 MCMC 算法(参见具体实施方式部分), 也即, Gibbs 取样算法内的 Metropolis-Hastings 算法, 允许不计算经典贝叶斯方法(通过 Metropolis-Hastings 算法)分母中的复杂积分(参见方程式 4), 并且可以简化典型贝叶斯方法(通过 Gibbs 取样算法)中分子的计算(参见方程式 4)。在 MCMC 产生前, 复杂的贝叶斯模型通常只在理论上有用, 而不管实践者采用的是非贝叶斯方法还是贝叶斯方法。

目前, 利用受试者考试数据以及此类数据的复杂贝叶斯建模进行认知诊断最可行的手段是, 利用 MCMC(参见 Gelman 等人文献的第 11 章, 其中对贝叶斯统计推断中 MCMC 的值进行了很好的描述)计算模拟算法分析这些数据。针对有待建模的特定场景的贝叶斯统计模型一旦被发展出来, 那么在给定这些数据的情况下, 发展一种有效的 MCMC 计算程序以获得参数的后验分布, 就只是冗长乏味、但相对较为常规的一件事。贝叶斯推断的一个很有吸引力的方面是, 计算出的后验分布提供了未知参数的模型标定, 以及进行推断的骨架, 例如, 有关技能掌握与否的认知诊断。介绍贝叶斯模型 MCMC 计算的优秀通用参考文献是, Gelman 等人的文章以及 Gilks, W.; Richardson, S.; Spiegelhalter, D. (1996) *Markov Chain Monte Carlo in Practice*(实用马尔可夫链蒙特卡洛方法). Boca Raton. Chapman & Hall/CRC。使用贝叶斯 IRT 模型(贝叶斯 UM 属于 IRT 模型家族)的 MCMC 计算的参考文献是 Patz 等人的文章。实际上, 如同 Patz 等人所用的标题, “A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Theory Models(用于考题反应模型的马尔可夫链蒙特卡



洛方法的一种简单途径)”的含义是，贝叶斯 IRT 模型 MCMC 的发展和应用，对于 IRT 和教育测量来说是可以采用的(accessible)，如果已经构建了贝叶斯 IRT 模型的话。

**基于似然的统计推断** 在理解 MCMC 的计算任务之前，有必要先理解贝叶斯推断在计算上是如何进行的。这同样也要求理解基于似然的统计推断在计算上是如何进行的，下面将对此进行解释。统计学中的一个核心概念是，在给定具体数据集的情况下，通常采用最大似然(ML)方法对参数进行估计。基本地，这意味着一个模型参数的值事实上被推断为最有可能产生所述观测数据的那个值。在统计建模中，基本的假设是，给定的模型在其参数的某些具体值上产生了所述观测数据。这种思想很简单，如下面的说明中所示。如果在上述实例 1 的医学数据中，观测到 75% 的患者被治愈，那么， $p = 0.2$  的理论治愈率(概率)极不可能在数据中产生如此之高的治愈率，相似地， $p = 0.97$  也极不可能在数据中产生相对而言如此之低的治愈率。与这一不正规的推理形成对照的是，通过基本的微积分知识可以得到， $p = 0.75$  是未知参数在数据中产生 75% 治愈率的最有可能的取值。 $p = 0.75$  的这一统计学估计是基于最大似然的推断的一个实例。

基于似然的推断的核心是一个函数，该函数描述了有待估计的参数的每一个可能取值有多大可能产生了所述的数据。使这一似然函数或似然概率分布(在上述贝叶斯方程式 4 中为  $f(X|\omega)$ )最大化的参数值，就成为其最大似然估计值。 $f(X|\omega)$  最好被认为是给定参数  $\omega$  的概率分布。例如，图 9 中给出了关于 40 次试验中有 30 人治愈的试验的似然函数，该似然函数表明，实际上  $p = 0.75$  使得似然函数最大化，因此是  $p$  的最大似然估计值。

**基于贝叶斯似然的统计推断** 这只是被先验的、关于不同参数取值似然度的看法(belief)或信息(通过先验概率分布表示，此类先验的实例示于图 5、6、7 中)所修改的基于似然的推断，如图 10 所示。“先验”是指，在得到由收集到的数据本身所获知的信息之前(以及除了由收集到的数据本身所获知的信息之外)，已经得到的信息。特别地，后验概率分布是一个函数，示出了将实际观测到的数据的似然函数和贝叶斯先验“结合”在一起后所得到的参数的贝叶斯似然分布。例如，在前述的实例 3 中，图 5 示出了  $p$  的三角形先验分布，图 10 同时示出了  $p$  的似然函数， $p$  的三角形先验，以及  $p$  的贝叶斯后验分布(也成为贝叶斯似然分布)，其中，所述贝叶斯后验分布是由该先验和试验数据得到的，在所述试验数据里，40 次试验中观察到有 30 人被治愈。回忆一下，方程式 4 给出了一个公式，用于计算所需的给定先验的后验分布函数以及似然概率函数。从图 10 的后验分布中可以注意到，通过使后验分布最大化而获得的  $p$  的估计值接近于 0.72 而不是 0.75，后者是通过使似然函数最大化的最大似然估计而得到的。

**在 MCMC 所解决的复杂贝叶斯统计分析中计算后验概率的不易处理性** 正如前面所说的，在复杂贝叶斯统计分析中，在计算后验概率时，经常会碰到巨大的实际困难。对大多数复杂贝叶斯问题而言，计算过程涉及很难操作的多重积分，即使是采用目前已经有的高速计算手段，对于直接计算来说也太过复杂，其中，所述的计算过程对于得到所需的后验分布而言是需要的，而所述的后验分布表明了未知参数的不同可能取值的可能性有多大。

尤其是，MCMC 是一个工具，在很多贝叶斯问题中，可以用于模拟进行贝叶斯推断所需的后验分布，而如果没有 MCMC，那么上述的贝叶斯问题是很难处理的。在科学技术中，“模拟”是一种用来取代对真实事物的直接观察的东西；在我们所处的情境中，这一取代是指取代对贝叶斯后验的直接计算。然后，通过对模拟结果的观察，就有可能接近对真实事物进行直接观察所得到的结果。

为了解释起作用的蒙特卡洛模拟的思想，让我们先来考虑一个简单的模拟方法，用于估计一个很简单的积分，事实上，这一积分可以通过初级的微积分很容易地得到。

**实例 4** 估计  $\int x e^{-x} dx$  在范围  $0 < x < \infty$  内的积分。通过由指数概率密度  $f(x) = e^{-x}$  (图 11 所示) 模拟大量独立观测  $x$ ，而得到该积分。然后，计算观测数据的平均值。

由于大数字的基本统计学规律(例如，如果我们进行多次抛掷硬币的测试，那么对于一枚两面均一的硬币而言，出现正面的次数将为约  $1/2$ )，这一由数据产生的平均值将会接近理论上由积分所给出的指数密度平均值( $f(x)$  的第一个矩或重心)。例如，如果 5 个模拟数字为 0.5、1.4、2.2、0.9 以及 0.6，那么，我们将该积分估计为这些模拟数字的平均值，即 1.12，而该积分的计算值为 1。当然，如果我们要求较高的准确率，那么，进行 100 次、400 次、甚至 1000 次模拟而不是 5 次，将会是较为理想的。这样，这一蒙特卡洛方法允许对未知积分进行精确估计，而不需要任何理论计算。

但是，对于复杂的、有很多参数的贝叶斯模型而言，这种对蒙特卡洛模拟方法的单独复制通常不实用。作为一种可行的替代方法，可以使用 MCMC 模拟，从而避免复杂的、很难处理的积分，而该积分是贝叶斯统计分析中的后验分布所要求解出的。特别地，MCMC 模拟可以估计几种统计学认知诊断模型的后验分布。每一种此类的 MCMC，均使用模型的贝叶斯结构(UM 或其他)和观测数据作为输入，如同方程式 4 的基本贝叶斯公式一样。回忆一下，模型的贝叶斯结构涉及先验分布和似然概率分布。

**非 UM 的现有技术实例** 现在，已经解释了有关统计学概念和数据计算技术(特别是贝叶斯概率模型和 MCMC)的概念上的必要背景，将描述相关现有技术(作为 UM 的补充)，由某些其他的已提出或已实行的认知诊断程序构成。

描述四种基于非 UM 模型的统计学认知方法(也就是说，这些方法基于受试者对考题作出反应的概率模型)，这些方法可以使用经简单记分的考试数据进行认知诊断。这些方法看上去似乎是已经发展到了可以实际应用的水平的主要统计学方法。值得注意的是，看上去只有 Robert Mislevy 的方法已经进入商业应用的领域，但仅仅是用于复杂的、很专业化的应用领域(例如牙科卫生学者的培训)，这与对简单的、以对/错记分的考题进行认知诊断相比，是很不一样的。这四种方法是：

1. Robert Mislevy 的贝叶斯网络证据中心方法(Bayes net evidence-centered approach)
2. Kikumi Tatsuoka 的规则空间方法(Rule-space approach)
3. Susan Embretson 的一般性潜在特征模型(Generalized Latent Trait Model) (GLTM)

#### 4. Brian Junker 的离散 GLTM(Discretised GLTM)

**Robert Mislevy 的贝叶斯网络方法** 首先考虑贝叶斯网络方法。两篇较好的参考文献是 Mislevy, R, 1995, *Probability based inference in cognitive diagnosis*(认知诊断中以概率为基础的推断), 在 Nichols, 等人的 *Cognitively Diagnostic Assessment*(认知诊断评估). Mahway, New Jersey, Lawrence Erlbaum 和 Mislevy, Robert 和 Patz, Richard, 1998, *Bayes nets in educational assessment: where the numbers come from*(教育评估中的贝叶斯网络: 数字从哪里来), Educational Testing Company technical report(教育考试公司技术报告); 普林斯顿, 新泽西如同本发明的贝叶斯 UM 方法(参见具体实施方式部分)一样, 这也是一种基于贝叶斯模型的统计学方法。虽然其通常应用的场景与如下场景不同, 即受试者对普通考题反应的原始数据被简单记分(例如考题被记为对或错)的场景, 但它也可以用于此类场景, 如 Mislevy 等人文献的第 5 部分中所报道的研究所示。很重要的是, 虽然它没有假定潜在技能, 这一点和 UM 是一样的, 但它并未使用考题/技能正性或不完全性的概念(因此, 贝叶斯网络方法没有引入 $\theta$ 来处理不完全性), 而本发明的贝叶斯 UM 中则使用了上述概念。该模型简化了 UM 方法中 $\theta$ 和正性参数 $\pi$ 和  $r$ 所起的作用, 这些参数使得本发明所使用的 UM 模型易于处理, 与此不同, 在贝叶斯网络方法中, 该模型由图论(graph-theoretic)技术所替代, 以减少贝叶斯网络条件概率的概率树型结构(probability tree)的参数复杂度, 所述条件概率将潜在技能的掌握状态与受试者对考题的反应联系在一起。对于一个不是图论专家的使用者(大部分认知诊断的使用者都是这样)而言, 有效使用这些技术事实上是很困难的。

教育考试服务中心(Educational Testing Service, 以下简称 ETS)正在以 Portal 的名称, 将贝叶斯网络技术进行商品化上市, 而且实际上已经将 Portal 用于牙科卫生学者的培训中。但是这种方法对于从业者的独立使用而言并不容易, 其原因上面已经交代过了。尤其是, 将这种方法输出到 ETS 以外用于可靠的独立使用, 是一件很困难的事, 而且需要对使用者进行认真的培训, 这一点与本发明的贝叶斯 UM 方法不同。此外, 它可能不具有本发明的 UM 方法所具有的统计推断能力, 特别是由于正性、不完全性(伴随  $\theta$  的引入)以及本发明的贝叶斯 UM 给这些技能所赋予的正相关结构(其重要性将在下面的具体实施方式部分中加以解释)中的每一个所起的重要作用。图 12 示出了贝叶斯网络方法的一个示意图。应当注意, 图 12 贝叶斯网络方法中的框 201、203 和 207, 与 DiBello 等人 1995 年的方法(回忆一下图 2)是相同的。框 1201 就是图 8 中的框 807, 是专用于贝叶斯网络模型的普通贝叶斯推断方法。相似地, 框 1203 是计算贝叶斯后验(图 8 中的框 809)的一种特殊情况, 实际上使用的是 MCMC。最后, 认知诊断步骤(框 1205)就是贝叶斯推断步骤(框 811)的一种特殊情况。

**Kikumi Tatsuoka 的规则空间方法** 两篇较好的参考文献是, Tatsuoka, K., 1983, *Rule space; an approach for dealing with misconceptions based upon item response theory*(规则空间; 基于考题反应理论处理误解的方法); *Psychometrika* 20, 34-38, 和 Tatsuoka, Kikumi, 1990, *Toward an integration of item response theory and cognitive error diagnosis*(考题反应理论和认识

错误诊断的整合); 在 *Diagnostic Monitoring of Skill and Knowledge Acquisition*(技能和知识获得的诊断监控)的第 18 章, Mahwah, New Jersey, Lawrence Erlbaum。图 13 示出了规则空间方法的一个示意图。针对受试者对每种可能的技能向量结构作出反应的随机性的规则空间方法,在某些方面是更为原始的,与本发明的贝叶斯 UM 很不相同。它完全基于随机受试者错误的概率模型,所述的错误被 Tatusoka 称为“犯小错误”(slips)。因此,完全性的概念没有了,正性的概念被完全表达为犯小错误(精神上的失灵)的概率。所采用的计算方法通常是贝叶斯方法。其基本思想是,对考题的实际反应应该像基于“理想”生产规则(production rule)的确定性反应(称为理想反应模式),所述确定性反应由考题/技能关系矩阵以及受试者的、可以由他/她的技能向量所刻画真正认知状态(随机的小错误不包括在内)所决定。通过给实际受试者反应模式以简单的贝叶斯方法赋以“最接近的”理想反应模式,而实现认知诊断。因此,规则空间方法基本是一种模式识别方法。

通过一个复杂的减少维数的操作,将  $n$  维反应空间(因为有  $n$  道考题)减小到二维“规则空间”(具体细节请参考框 1303 以及 Tatsuoka 的两篇参考文献),从而实现对规则空间认知诊断的计算。这产生了二维贝叶斯模型(框 1301,这与图 8 中普通贝叶斯模型的组成单元 807 相类似)。减小到低维度的“二空间”(two space)后,允许一个人直接进行所需要的贝叶斯计算(参见框 1305),而不必求助于 MCMC。然后,最能预测所赋予的理想反应模式的技能状态 $\alpha$ ,被推断为受试者的认知状态,从而作出了认知诊断。这种方法中没有完全性,没有正性,没有赋予所述技能的正相关结构,而且其犯小错误的概率分布(probability of slips distribution)是建立在一些看上去有点不切实际的假设之上的。特别地,由于上述原因,本发明的贝叶斯 UM 方法应该比规则空间方法做得更好。在所使用的受试者反应行为的概率模型以及贝叶斯标定和诊断算法上,这两种方法是很不一样的。应该注意,在 DiBello 等人 1995 年的 UM 方法以及规则空间方法中,框 201、203、205 和 207 是相同的。与所有认知诊断方法一样,最后一个框,此处也就是框 1307,是进行实际的认知诊断。

**Susan Embretson 的一般性潜在特征模型 (GLTM)** 两篇较好的参考文献是, Susan Embretson 的书题为 *Item Response Theory for Psychologists*(心理学家使用的考题反应理论)的第 11 章, 2000, Erlbaum, 新泽西, 和 Embretson Susan, 1997, 在 *Handbook of Modern Item Response Theory* (现代考题反应理论手册)(van der Linden 和 Hambleton 编, 纽约, Springer 出版)中的第 18 章 Multicomponential response models (多组分反应模型)。该方法与本发明的贝叶斯 UM 方法不同。它假定有待推断的技能是连续的,而不是贝叶斯 UM 中所假定的二元性(binary)(0/1),它没有不完全性部分,也没有正相关的技能结构。由于它认为技能是连续的,因此,它倾向于应用于连续的潜在技能,例如“工作记忆”能力以及任务完成前的时间。它使用一种被称为 EM 算法的计算方法,至少在其公开发表的描述中是这么说的,因此,GLTM 模型并未在贝叶斯框架内被重构。虽然从原则上讲,它适用于普通的、经简单记分的考试数据,但那似乎并未成为其应用的重点。图 14 示出了 GLTM 的一个示意图。框 1401 与图 2 中的框 201 相似,但是此处的技能是连续的。框 203 和 207 与其他现有技术的程

序是相同的。框 1405 与图 2 UM 中的框 209 类似，框 1405 与图 2 UM 中的框 213 类似，最后，与所有程序一样，最后一个框 1407 是进行认知诊断。

**Brian Junker 离散(0/1)版本的 GTLM** 其思想是，将 Embretson 的 GTLM 模型中的连续潜在技能替换为二元的(binary)潜在技能，并保持模型的主要结构不变。一篇较好的参考文献是，Junker, Brian, 2001, 在 *Essays on Item Response Theory*(考题反应理论集)(A. Boomsma 等人编辑，纽约，Springer 出版)的第 14 章 On the interplay between nonparametric and parametric IRT, with some thoughts about the future(非参数化和参数化 IRT 的相互影响以及对未来发展的想法)。在这种新方法和本发明的贝叶斯 UM 方法之间的主要区别可能是，离散 GTLM 没有不完全性部分。此外，它也没有正相关技能结构。最后，它的正性结构比本发明的贝叶斯 UM 要简单的多，因为对于离散 GTLM 而言，不允许一项技能的正性度依赖于哪道考题正在被解答。离散 GTLM 的计算方法是 MCMC。

只对上述的前三种统计学程序提供了可供对照的流程图(Junker 的离散 GTLM 与 Embretson GTLM 的示意图几乎完全相同)。

在各种现有技术方法和本发明之间，最主要的区别始终是模型不同，虽然也还有其他的区别特征。

**2. 基于确定性认知模型的程序** 有多种方法使用确定性认知诊断方法。由于其自身的统计性质，统计学方法要优于任何确定性方法(也就是说，基于规则的(rule-based)数据挖掘，人工智能，专家系统，AI，基于神经网络的方法，等等)。所有的确定性方法都没有深入的、有效的方法来避免对数据的过度拟合，因此，错误地对技能的掌握与否做出结论，而事实上，支持这些结论的证据是非常弱的。

此外，这些确定性方法的模型在参数上都过于复杂，不能支持利用普通的、简单记分的考试数据所进行的模型标定。这些模型的数量很多，而且，对于在简单考试数据的环境下进行认知诊断而言，这些模型几乎无法使用。

**第 3 部分 在医学和精神病学领域的现有技术** 前面只考虑了以教育为导向的认知诊断场景。基于在所观测的考题上的表现而对受试者进行的认知诊断，与对患者所进行的医学诊断之间，有某种相似的结构。在这两种情况下，都试图基于观测信息而测量一种潜在的状态(技能或医学/精神病学上的疾病，以下简称为“疾病”(disorder))，所述的观测信息与这种潜在的状态之间是有关联的。为了对某种具体的技能或疾病做出推断，根据其他的技能或疾病来理解这个人的状态也是很重要的。尤其是，在医学和精神病学上，诊断工具的目标是为从业者提供一个简短的列表，其中列出了某些根据观测到的症状和患者的个人特征(例如性别、种族、年龄等)来判断似乎是很有可能的疾病。具体地说，赋予疾病集合的贝叶斯后验概率与赋予认知技能集合的一套后验概率是类似的。虽然概率建模方法已经在医学和精神病学中尝试过了，但基于概率的 IRT 模型还没有被尝试过。

下面，我们将列举医学和精神病学诊断现有技术中的几个具有概率意味的例子。

**基于贝叶斯网络的系统** 用于医学诊断的贝叶斯网络通过一幅图代表了疾病和症状/特征之间的随机关系，这幅图将相互之间通过连接线而随机地相互依赖的节点连在一起。一篇很好的一般性参考文献是，Herskovits, E.和 Cooper, G., 1991, *Algorithms for Bayesian belief-network precomputation*(贝叶斯信任网络预先校正的算法), *Meth. Inf. Med*, 30, 81-89。贝叶斯网络建模专家创建了一幅定向图(a directed graph)，从代表疾病集合的初始节点集合开始，经过可选的中间节点集合，至最终得到的症状/特征观测数据集合为止。在给定患者的具体症状/特征观测数据集合的情况下，具有某种疾病的后验概率可以通过方程式 4 的贝叶斯方法计算出来，MCMC 可能也可以用。此处，已经给提出的可能疾病的集合赋以一个先验分布，而且，在给定图中前面一个节点的情况下，给出了每一个节点的条件概率，从而给出了方程式 4 中所需要的似然函数。以这种方式，图中的每一条线都有与之相关的条件概率。贝叶斯网络在医学上的应用，最初是通过参考适当的医学文献、参考可以获得的大型数据集或咨询专家意见的方式，而获得所需的条件概率的数值。目前，获得这些条件概率的估计技术已经发展起来了。对于贝叶斯网络的工作而言，虽然估计条件概率的能力很重要，但是，主要的障碍仍然是需要做出很多简化模型的假设，以便使得网络在统计学上更易处理，这一点在上面的用于认知诊断的贝叶斯网络现有技术方法的讨论中，已经进行了解释。

**基于神经网络和模糊集合理论的系统** 基于神经网络和模糊集合理论(Fuzzy Set Theory)的方法都是图形网络，通过使用网络，设计症状/特征和疾病之间的概率关系，然后使用较大的数据集进行大量的训练。与贝叶斯网络相比，在神经网络和基于模糊集合理论的网络中，网络规定得不如贝叶斯网络严格。网络的训练主要是比较通过训练过程所标定的很多模型，以便找到一个符合得合理好的模型。模糊集合理论技术允许将随机误差构建入系统中。神经网络可能也允许将随机误差构建进来，但其方式不像模糊集合理论那么正式。这两种系统都有某些问题，这些问题产生于训练阶段较大的自由度，所述训练阶段包括：训练过度/训练不足；确定用于训练的案例(cases)(数据)，因为模型越复杂，需要的案例就越多；确定节点数目；以及概括较好的适当数据集的可行性。该方法与 UM 规定的模型参数方法很不一样。较好的参考文献是，针对神经网络的 Berman, I.和 Miller, R., 1991, *Problem Area Formation as an Element of Computer Aided Diagnosis: A Comparison of Two Strategies within Quick Medical Reference* (作为计算机辅助诊断的问题域的形成：在快速医学查询领域的两个策略的比较)*Meth. Inf. Med.*, 30, 90-95, 和 Adlassnig, Kl, 1986, *Fuzzy Set Theory in Medical Diagnosis*(医学诊断中的模糊集合理论), *IEEE Trans Syst Man Cybernet*, SMC- 16: 260-265。

**确定性系统** 所使用的两种确定性方法是分支逻辑系统(Branching Logic Systems)以及启发式推理系统(Heuristic Reasoning Systems)。正如前面在认知诊断现有技术部分中所讨论的，与基于概率模型的方法例如 UM 相比，所有的确定性系统都有缺点。

## 发明内容

本发明基于对象所产生的二值性数据对状态未知的对象(通常是人)进行诊断。本发明的

应用包括但不限于：(1)在课堂授课场景中，对学生的考试数据进行认知诊断，其目的可以是，例如，评估学生个体的认知进步以及课程范围(course-wide)的学生认知进步，以用于——例如——指导基于授课的补习/干预，以便解决所发现的认知缺陷；(2)在计算机化授课场景(例如基于网络的授课系统)中，对学生的考试数据进行认知诊断，其目的可以是，例如，评估个体的认知进步以及课程范围(course-wide)的认知进步，以用于——例如——指导计算机交互式补习/干预，以便解决所发现的认知缺陷；(3)大规模标准化考试的认知诊断，从而评估认知上定义的基于群体的认知图谱(cognitive profiles)，其目的可以是，例如，评价教学区的授课效率，为受试者个体提供认知图谱作为反馈；以及(4)对医学和精神疾病进行医学和精神病学诊断，其目的可以是，例如，对个体患者/客户进行诊断、治疗干预以及研究。

除了在上面所列出的场景中进行认知诊断或其他诊断外，本发明的应用范围还包括：利用任何类似于考试的观测数据(也就是说，来自每一个对象的、可以二值性记分的大量数据，例如针对每一名参加考试的考生所观测到的、按照对/错记分的多道考题的记录)，对任何潜在(不能直接观测到的)结构(一群个体对象(通常是人)所具有的)进行认知诊断，其中，所述的观测数据在概率上受该潜在结构的控制，正如本发明所建模的那样。为了便于解释，可以利用本发明对关于态度的调查问卷数据进行诊断，以便针对每一个答题人对其某些属性(attributes)进行推断，例如社会自由主义 vs. 保守主义，财政自由主义 vs. 保守主义等。

### 定义的术语

技能——任何影响可观察的精神(mental)技能的潜在精神能力。

考题——考试中的问题，受试者对其所做出的反应可以被编码为对或错。

剩余能力参数——一个维数较低(当然不大于 6，通常是一维的)的量的集合，所述的量合在一起概括了受试者在更大的技能集合的剩余部分上的水平，所述的技能可以影响受试者的答题表现。

二值性记分的探针——与认知诊断场景中的考题相类似。可以从被评价的对象处得到具有两个值的反应的任何事物。

对象——与认知诊断场景中的受试者相类似。被诊断的任何实体的集合。

联系(association)——两个变量，例如技能，之间的任何关系，其中，值较大的变量会使得另一个变量的概率倾向于更大(正的联系)或更小(负的联系)。相关(correlation)是量化联系的一种常见方法。

观察不到的二值性——与认知诊断场景中的技能相类似。对象的、观察不到的任何性质，但该属性或者具有两种状态，或者可以被编码为具有两种状态，其中，一种被称为具有该属性，另一种被称为不具有该属性。适当应用属性的意思是，提高依赖于该属性的探针的正反应的机率(chance)。

症状/特征——与认知诊断场景中的考题相类似。在医学或精神病学场景中，患者的可

以观察的方面。可以是很明显的，例如性别或喉咙痛的症状，或者，也可以是医学测试结果或发给患者的问题的结果。在当前 UM 的应用中，要求可以二值化。

健康或生活质量参数——与认知诊断场景中由 $\theta$ 所给出的剩余技能的总和相类似。与 UM 医学诊断应用中所列出的、规定的疾病不同，是对患者处于医学上健康状态的一个一般性的、宽泛的指标。

疾病——任何医学或精神病学的状态，并且是潜在的，因此需要被诊断，在某方面构成了患者的不舒适。

探针——与认知诊断场景中的考题相类似。可以从被诊断的对象处得到具有两个值的反应的事物。

对探针的正反应或负反应——与认知诊断场景中答对或答错一道考题相类似。正和负仅仅是给予对探针的两种可能反应的标记，注意，“正”反应有时候在上下文上是有意义的，有时候则是没有意义的。

#### 附图说明

图 1 示出了标准逻辑考题反应函数  $P(\theta)$ ，一般来说，该函数被用作 IRT 模型的基本构成单元，尤其是，该函数被用作 UM 的基本构成单元。

图 2 示出了 1995 年的现有技术所提出的 UM 认知诊断程序的流程图。

图 3 示出了用于一名受试者对一道考题的随机反应  $X_{ij}$  的 1995 UM 概率模型的示意图，指出了影响受试者反应  $X_{ij}$  的受试者参数和考题参数。

图 4 示出了 UM 中所采用的受试者剩余能力  $\theta$  分布的标准正态概率密度函数。

图 5 示出了在统计学药物试验研究中，可以提供较多信息的、参数  $p = \text{Prob}(\text{治愈})$  的三角形先验密度  $f(p)$ 。

图 6 示出了在统计学药物试验研究中，参数  $p = \text{Prob}(\text{治愈})$  的、模糊的(相对提供信息较少)贝叶斯先验密度  $f(p)$ 。

图 7 示出了在统计学药物试验研究中，完全不能提供任何信息的贝叶斯先验密度  $f(p)$ 。

图 8 示出了基本贝叶斯概率模型统计推断范例的组成部分。

图 9 示出了在统计学药物试验研究中，当数据为 40 次试验中有 30 人治愈时， $p = \text{Prob}(\text{治愈})$  的似然函数  $f(X|p)$ ，表明  $p = 0.75$  使得似然函数最大化。

图 10 同时示出了在统计学药物试验研究中，当 40 次试验中有 30 人治愈时， $p: f(X|p) = f(40 \text{ 人中有 } 30 \text{ 人治愈} | p)$  的先验密度、似然函数以及后验分布，其中， $p = \text{Prob}(\text{治愈})$ 。该图解释了贝叶斯先验分布在标准统计学最大似然估计  $p = 0.75$  上的效果，得到了贝叶斯后验估计  $p = 0.72$ 。

图 11 示出了函数  $e^{-x}$ ，该函数将通过模拟而被计算积分。

图 12 示出了 Robert Mislavy 的贝叶斯概率推断网络方法用于认知诊断的流程图。

图 13 示出了 Kikumi Tatsuoka 的贝叶斯规则空间方法用于认知诊断的流程图。



图 14 示出了 Susan Embretson 的 GLTM 方法用于认知诊断的流程图。

图 15 示出了用于一名受试者对一道考题的随机反应的 UM 似然的示意图，指出了影响受试者反应  $X_{ij}$  的受试者参数和考题参数，用于本发明所使用的再参数化的统一模型(Unified Model)中。

图 16 示出了本发明所使用的再参数化 UM 的可识别贝叶斯版本里的相互依赖关系，包括先验分布和超级参数。

图 17a 示出了本发明所使用的 UM 认知诊断程序的流程图。

图 17b 示出了本发明所使用的 UM 医学/精神病学诊断程序的流程图。

图 17c 示出了本发明所使用的一般性 UM 程序的流程图。

图 18 示出了基础统计学考试中的一页，以解释在 UMCD 示例性实例中所模拟的考题。

图 19 示出了在 UMCD 示例性实例中所模拟的基础统计学考试的考题/技能关系矩阵。

## 具体实施方式

本发明部分地基于发现在 DiBello 等人 1995 年所提出的 UM 方法中的不足。这些不足是：导致参数不可识别性(nonidentifiability)的过度参数化(overparameterization)；未能设定技能水平，而这也是导致参数不可识别的另一个原因，而且对于使用者的解释而言，也产生了实质性的问题；缺乏实用的、有效的标定程序；以及未能对存在于技能之间的自然的正相关结构进行建模，而对该结构的建模将会提高认知诊断的准确性。下面将首先讨论这些不足。在讨论之前，必须先对参数化和可识别性(identifiability)有一点更多的理解。

**统计建模中的不可识别性以及模型的再参数化(reparameterization)** 在统计建模中，相对于具有更多参数的、对真实世界的描述最多稍微好一点的模型相比，对真实世界描述得相当好的、具有更少参数的模式是优选的。如果具有更多参数的模型中带有不可识别的参数，那么前面所说的这一点是特别重要的，其中，不可识别的参数，也就是在统计上不能相互分开的参数，此类参数根本不能根据数据而估计出来。一个简单的实例解释了不可识别性的重要思想以及对再参数化的需求。考虑模型  $y = a + bx + cx$ 。这个模型有三个参数  $a$ 、 $b$ 、 $c$ 。但该模型是过度参数化的，因为  $b$  和  $c$  扮演了完全相同的角色(与变量  $x$  相乘的一个参数)，因此在统计学上相互之间是无法区分的。因此，模型参数  $b$  和  $c$  是不可识别的，无法由可以获得的数据中估计出来。两个参数的模型  $y = a + bx$  是更好的，因为它的参数少一个，它所有的参数都是可以识别的，而且它对真实世界的描述也很好。通过本发明，通过减少参数的数目，没有用的、不可识别的 1995 UM 被再参数化，其中，引入了一套数目更少、但实质上更有意义的参数，而且规定了技能掌握水平，从而使所得到的参数都能识别，因此可以对其进行估计。

**再参数化的一般方法** 假定一个模型具有一套有意义的参数，其中包括  $K$  个参数，也就是说，这些参数具有有用的、真实世界的实质性解释(例如物理模型中的速度、质量、加速度等)。一般方法是，对于  $k < K$ ，定义新的、有意义的参数  $a_1, a_2, \dots, a_k$ ，其中，每一个

参数是包括  $K$  个参数的原始参数集的不同函数。最好对函数进行选择，从而使得这套新参数可以识别而且实质上有意义。正确的再参数化不是唯一的，因此可以存在多种有用的、正确的再参数化。

现在考虑在 1995 UM 中的不可识别性。

**在现有技术 DiBello 等人的 1995 UM 中不可识别性的来源：未能极度节省地进行参数化以及未能规定技能水平。** 已经发现，不可识别性的来源有两部分。首先，参数的数目必须通过实质上有意义的再参数化、通过上面所介绍的一般方法而加以减少。

第二，已经发现，在模型中针对每一项技能规定掌握水平，是该模型的一个必要组成部分。实质上，规定掌握水平时，定义了一名受试者为了将一项技能应用到考题上，需要将该技能掌握到什么程度，才能被归为已掌握该技能。需要对掌握程度进行规定，不只是为了获得可识别性，而且也是使用者能够从 UM 认知诊断中获得实质上有意义的结论所要求的。实际上，宣称一名受试者已经掌握一项技能是没有意义的，除非使用者知道在构成考试的考题的上下文环境中掌握技能实际意味着什么。因此，任何未能在某种程度上设定掌握水平的认知诊断模型，都有重要的缺陷，可能会引发严重的失误(malfunctioning)。

**在 1995 UM 中未能使用技能的正相关结构** 在 1995 UM 中所发现的另一个问题是，从受试者数据中可以获得的、关于技能之间的联系的信息，未被 1995 UM 所利用，而这一缺陷可以通过将该模型仔细地重构为贝叶斯模型而加以克服。当然，将来也可能发现其他方式，也可以获取很多可以利用的信息，从而使得贝叶斯建模不再是唯一的选择。

有效地处理这些发现(过度参数化，没有规定技能掌握水平，没有使用技能正关联结构)的结果是一种实用的、有力的认知诊断程序，可以应用到实际的考试数据上，以得到参加考试的受试者的实际认知诊断，也即，是本发明的 UMCD。

**1995 UM 未能实现标定** 与发展一种有用的、基于 UM 的认知诊断程序同样基础的是，找到一种有效的标定程序。实际上，在 DiBello 等人的方法中，没有实现对模型的标定。模型的不可识别性和非贝叶斯特性都是标定的障碍。没有实现此类标定已经阻碍了有效认知诊断的进行。最近，新型数据计算方法 MCMC 的流行，使得对贝叶斯模型的标定成为可能，即使在模型参数很复杂的情况下。这意味着，为了对模型进行有效的标定，将 1995 UM 重构为贝叶斯模型是一种可行的策略。必须再次明确，如果不经标定，那么认知诊断是不可能进行的，不管模型是多么的有现实意义。例如，DiBello 等人的文章中提出了基于 UM 的认知诊断，对于一项模拟的、基于 UM 的认知诊断所进行的解释，只能通过如下方式才能获得，即：假装该 UM 已经被标定过，而在这篇论文发表的时候，对 UM 进行标定在统计学上是不可能。因此，在 1995 UM 发表的时候，利用 1995 UM 进行认知诊断是不可能的，而且实际上，在本发明的贝叶斯 UM 及其基于 MCMC 计算的模型标定被发展出来以前，这都是不可能的，其中，本发明的贝叶斯 UM 具有可以识别的参数，而且对掌握程度做出了规定。

现在来讨论本发明的 UMCD 中所使用经过发展的再参数化。

**用于取代 1995 UM 过度参数化的再参数化** 具体地说，对 DiBello 等人的文章中所发表的非贝叶斯 UM 进行再参数化是必要的，其目的是使得参数可以识别(参见下面的方程式 5)。特别地，已经意识到，对于进行适当的认知诊断而言，需要对 1995 UM 进行再参数化。也就是说，UM 中冗余的原始参数必须被取代，即使它们具有实质上有意义的解释。(如前所述，对非贝叶斯 UM 的再参数化，在概念上与下面的处理类似：利用更简单的、未过度参数化的、可识别的模型  $y = a + bx$ ，替换不可识别的、过度参数化的模型  $y = a + bx + cx$ 。)

而且，再参数化必须产生可识别的参数，这些参数的意义对于实际的从业者来说很容易理解。如下面所解释的，对再参数化的具体选择似乎是何为 UM 程序在实际应用中工作的很好以及该程序对使用者来说很容易理解和解释的一个主要原因。

下面将对本发明所使用的重构 UM 的基本概念进行解释。频繁参考图 15，将图 15 和图 3 进行比较，以及仔细查看方程式 5 和 6 是很重要的。理解本发明所建模的 UM 的独特之处，对于理解本发明认知诊断算法的独特之处而言是很重要的。在前面对现有技术 1995 版本的 UM 进行描述时，已经解释了其中的一部分。使得 UMCD 可以很有效地进行认知诊断的内容，是下面将要描述的图 15 以及方程式 5 和 6 所特有的。

如前所述，在 UM 里一个具有基本重要性的认知概念(construct)是正性，这一点可以从表示  $S_{ij}$  的方程式 5 中清楚地看出，其中使用了方程式 6 中的再参数化的  $\pi^*$  和  $r^*$ ，下面将对此进行解释。方程式 5 与用于表示  $S_{ij}$  的方程式 3 类似，方程式 3 使用了根据  $r$  和  $\pi$  的原始参数化。用于表示  $S_{ij}$  的这两方程式都给出了被包括在内的技能被受试者  $j$  正确地用于解答考题  $i$  的概率。为了获得可以识别的、实质上有意义的参数，方程式 5 提供了  $\pi$  和  $r$  的再参数化。 $S_{ij}$  的方程式 3 版本被方程式 5 版本所替代，注意，这两个公式都得到了同样的  $S_{ij}$  值。

$$S_{ij} = (\pi_i^*) \times (r_{i1}^*)^{1-\alpha_{j1}} \times (r_{i2}^*)^{1-\alpha_{j2}} \times \dots \times (r_{im}^*)^{1-\alpha_{jm}} \quad (5)$$

如前所述，再参数化的一般方法要求将新的、可以识别的参数( $\pi^*$ 's,  $r^*$ 's)定义为老的、不可识别的参数( $\pi$ 's,  $r$ 's)的函数。通过下面的方式进行。考虑考题  $i$  要求  $k = 1, \dots, m$  项技能。那么，定义

$$\begin{aligned} \pi_i^* &= \prod \pi_{ik} \quad (\text{对 } k \text{ 求积}) \text{ 以及} \\ r_{ik}^* &= r_{ik} / \pi_{ik} \end{aligned} \quad (6)$$

可以实现再参数化。注意，有  $2m$  个  $\pi_k$  和  $r_{ik}$ ，只有  $m+1$  个  $\pi_i^*$  和  $r_{ik}^*$ 。

如前所述，第  $i$  道题要求  $m$  项技能，分别记为  $1, 2, \dots, m$ ； $\alpha_{jk} = 1$  或  $0$  表示受试者  $j$  已经掌握或尚未掌握技能  $k$ 。 $\pi_i^*$  被解释为一名已经掌握考题  $i$  所要求的全部技能的受试者，实际上能正确应用这些技能的概率。也就是说， $\pi_i^*$  量度了这道考题对一名已经掌握全部必需技能的受试者来说的难度。

接下来，按照前面的定义，技能 1 的  $r_{i1}^*$  是：在不掌握技能 1 的情况下将该技能正确应用于考题  $i$  的概率，除以在掌握技能 1 的情况下将该技能正确应用于考题  $i$  的概率。其他技能的  $r^*$ 's 也采用类似方式定义。技能  $k$  的  $r_{ik}^*$  值  $\approx 0$  仅仅表示：当试图正确解答考题  $i$  时，如果掌握该技能，将会有很大的优势。相对更接近 1 的  $r_{ik}^*$  仅仅意味着：当试图解答考题  $i$  时，掌握技能  $k$  与未掌握技能  $k$  相比，优势很小。

对于考题  $i$  而言，如果  $\pi_i^*$  接近于 1 而所有的  $r_{ik}^*$ 's 接近于 0，那么必需技能被称为对考题  $i$  而言高度正性。如前所述，“高度正性”仅仅意味着当且仅当一名受试者掌握该模型所指出的、该考题所需的全部技能时，这名受试者能够正确运用这些必需技能的概率较高。

应该注意， $r^*$ 's、 $\pi^*$ 's 以及图 16 的掌握程度设置  $p_k$ 's(下面也会对掌握程度设置进行解释)一起，对于得到所需要的、DiBello 等人的方法中所缺少的可识别性而言，是充分的。在规定了技能掌握水平的情况下，这么多的参数对于获得可识别性而言是充分的。

**分级贝叶斯 UM 模型，包括掌握水平的设置以及技能正相关结构的引入** 贝叶斯 UM 中贝叶斯部分的作用的重要性，与为了获得有效的、强有力的认知诊断，再参数化的 UM 公式所起的作用相当。这是通过在贝叶斯模型中引入超级参数而实现的，也就是一个分级贝叶斯模型。正如背景技术部分中所述，贝叶斯模型是一个概率模型，其中的模型参数也被赋予了一个概率分布。具有超级参数的贝叶斯模型是一个贝叶斯模型，其中，模型基本参数的先验分布也被给定了参数(每个参数都有先验分布)。这些控制常规模型参数先验分布的附加参数被称为超级参数。一篇较好的、一般地涉及贝叶斯建模的参考文献是 Gelman 等人的文章，这篇文章也具体地涉及分级贝叶斯建模。

图 16 图示了一名受试者对考题做出反应的分级贝叶斯模型，如同我们的分级贝叶斯 UM 建模的那样。因此，它是图 15 的再参数化似然示意图的提高。

在图 16 中，模型参数  $\pi^*$ 、 $r^*$  以及  $c/3$  具有先验 beta 分布，对每道考题  $i$ ，记为  $\beta(a, b)$ ，每一个这样的分布都由两个参数  $(a, b)$  所决定。当参数被限制在区间  $(0, 1)$  上时，作为先验分布的 Beta 分布往往工作得较好，这一点在 Gelman 等人的著作的第二章中已经指出并做了解释。具体地说，beta 分布参数  $(a, b)$  提供了丰富的密度族(density family)，可以从中选出几乎任何一种形状的先验分布，从建模的角度讲，这是一个很吸引人的性质。每一个超级参数  $(a, b)$  都被赋予了区间  $(0.5, 2)$  上的均匀分布。这意味着，每一个参数值，比如  $a_r$ ，在区间  $(0.5, 2)$  上的可能性都是相等的。那种合适的、相对而言不能提供什么信息的(模糊的)先验，在分级贝叶斯模型中很有效，而在较大区间上的这一均匀先验分布就是这样一种先验分布，这是由于它允许模型在没有先验的情况下很好地对数据进行拟合，而所述先验会对统计推断产生不适当的、很强的影响。应该注意，关于分布的这些选择(beta 分布，均匀分布)是相当标准的选择，虽然为了构建有关变量的先验分布，还需要进行一定数量的判断。

现在来解释与受试者潜在能力参数(也就是说，的不完全性剩余能力  $\theta$  和  $\alpha$  技能掌握/未掌握成分)相关的贝叶斯结构。进行解释的目的是，强调目前 UM 程序的两个重要部分，也就是说，规定技能掌握水平和假定一个正相关技能结构，并将其作为贝叶斯模型的一部分。假

定受试者技能和 $\theta$ 是由正相关多变量正态分布得到的。对于统计学家来说，多变量正态分布是一种常见分布，很容易理解。例如，如果测量一个人的体重和身高，那么标准模型是一个二变量正态分布，其中体重和身高为正相关。如果要了解更多的信息，请参考普通的统计学教科书。

规定技能 $\alpha$ 和 $\theta$ 的先验分布是通过两个阶段完成的。在第一阶段， $(\theta, \alpha')$ 被赋予多变量正态先验，其中， $\alpha'$ 是 $\alpha$ 二值(取值为 0/1)部分的连续前身(precursor)， $\alpha$ 的二值部分规定了每一名受试者对每一项技能掌握/未掌握的情况。 $\alpha'$ 的技能对相关(attribute pair correlation) $\sigma_{kk}$ (超级参数)被赋以区间(0, 1)上的均匀分布，因为仅仅知道它们是正相关的。然后，根据 $\alpha'$ 每一个部分的值大于还是小于使用者所规定的掌握水平而赋以 0 或 1，从而得到技能掌握/未掌握向量 $\alpha$ ，而掌握水平最简单的是由使用者所规定的、针对每一项技能的受试者掌握比例(概率) $p_k$ 所确定。也就是说，使用者通过规定每项技能掌握者的比例，而规定了技能已掌握的含义(可能还会发现其他方法用于规定技能的掌握情况，而且可能实际效果更好，但这种方法是最直接的方法)。例如，如果使用者规定  $p_k = 0.7$ ，那么可以认为技能 k 被 70%的受试者所掌握。那么，当实际上它相应的 $\alpha'_k$ 足够大时，在 70%的时候 $\alpha_k = 1$ 。在剩下的 30%的时候 $\alpha_k = 0$ 。

为了便于解释规定掌握水平的必要性，考虑如下的试验。说一个人已经掌握分解多项式(技能 1)意味着什么呢？很明显，对于所需的适当能力水平会有很多不同看法。因此，规定 60%的受试者( $p_1 = 0.6$ )已经掌握，具有精确定义掌握水平的效果。而如果选择 80%，那么在宣称一个人已经掌握该技能之前，需要更高水平的认知机能(cognitive functioning)。

除了规定技能水平很重要以外，必须再次强调，UM 的贝叶斯部分中所采用的组成技能对(component attribute pairs) $\alpha$ 的正相关结构，提高了认知诊断的准确性。因为这一正相关结构使得模型可以捕捉住这一至关重要的事实，即：已经掌握某一项技能的受试者更有可能掌握另一项技能，也就是说，技能是正相关的，或者更简单地说，是正性联系的(positively associated)。而且，将技能的这一很重要的正相关结构整合进来，是通过在贝叶斯框架内重构(cast)UM 而实现的。但是，本发明并不限于贝叶斯框架。因此，将有效的正相关技能结构和再参数化的、从而可识别的、技能水平已规定的 UM 结合在一起(目前是通过贝叶斯方法实现的)，是对于获得有效的 UMCD 而言有用的所有部分。也就是说，这其中的每一个，与其他部分相组合，以及与 UM 相组合，对本发明的表现做出了贡献，其中，UM 被定义为任何基于技能的、使用正性和完全性来发展其方程的诊断模型。

图 16 示出了 UMCD 的分级贝叶斯 UM 的一个实施例。因此，本发明不限于具有贝叶斯模型和认知诊断 MCMC 算法的 UMCD 的这一实施例。

意识到如下事实是很重要的，即：将非贝叶斯概率模型转化为贝叶斯概率模型，在其细节上随应用环境的不同而有很大不同。此类转化操作很少是相同的。因此，对于需要对数据进行贝叶斯建模的每一个不同新环境而言，此类转化操作都要重新开始。特别地，发展适当贝叶斯模型的合适(right)方法不只是一种。而且，一个经适当选择的贝叶斯模型，如同对

UM所做的那样，可以有效地利用数据中的所有信息，因此获得准确得多的推断(此处，是准确得多的认知诊断)。

图 17a 提供了本发明方法的流程图。首先注意到框 201、203、205 和 207 与图 2 中基于 UM 的框是相同的。这反映出，除了 UM 模型所使用的细节以外，两者采用了同样的途径。因此，图 2 的非贝叶斯方法和图 17a 的贝叶斯方法从框 205 以下开始分化。首先，虽然如前所述，这两者都要求似然模型，但是，与 1995 UM 的不可识别性相关的再参数化问题导致了方程式 5 中的再参数化的发现，此发现的目的是要替代方程式 3 中的、老的参数化。此外，似然模型(分别为框 209 和 1701)的构建现在也要求有“构建 UM 贝叶斯先验  $f(\omega)$ ”框(框 1703)，从而产生了贝叶斯模型框 1705。图 17 的框 1701、1703 和 1705 反映出了方程式 5 和 6 以及图 16 的示意图。框 1707、1709 和 1711 按如下方式理解。如前所述，所需的后验分布  $f(\omega|X)$  通过 MCMC(框 1707)获得。然后，利用常规技术，将一维  $\alpha_{ik}$ 's 的后验概率(其目的是进行单个的技能/受试者认知诊断)从后验概率  $f(\omega|x)$  中提取出来，得到框 1709，框 1709 针对每一个受试者/技能组合产生  $\text{Prob}(\alpha=1|X)$ 。然后，根据证据规则(如下面的实例所述)，可以获得针对每一个受试者/技能组合的认知诊断(框 1711)。

**对本发明贝叶斯 UM 中所用的 MCMC 算法的简要描述** 对贝叶斯 UM 中所用的 MCMC 算法的一般性描述，可以参考 Patz 等人的文章，其中给出了足够的细节，足以使本领域的普通技术人员创建和使用它。如前所述，该方法被称为嵌于 Gibbs 取样器(Gibbs Sampler)中的 Metropolis-Hastings 算法，或简称为 Gibbs 中的 M-H 算法。Metropolis-Hastings 算法可以通过避免计算分母(见方程式 4)，而简化后验分布的计算，而在后验分布计算中通常需要计算分母。Gibbs 取样器可以将计算的剩余部分(方程式 4 的分子)分解为几个部分，单独计算这几个部分比把它们合在一起计算要容易(因为它们合在一起之后在计算上相互依赖)。Gibbs 中的 M-H 算法是基本 MCMC 方法的多种变形中的一种。

就 MCMC 而言，马尔可夫链的模拟随机数字在概率上是不独立的(如同在连续两天内的高温)。而且，Patz 等人的文章(以及其他的、关于利用 MCMC 进行贝叶斯分析的、较好的一般性参考文献，例如 Gelman 等人文章或 Gilks 等人的文章)中很仔细地解释过，MCMC 模拟完全避免了计算(甚至是模拟)分母中的积分，而是产生了一条随机数字的“链”，其稳态概率分布正是我们想要的后验分布。换用简单实用的术语来讲，这意味着如果该链可以运行较长时间，那么其模拟随机数字的观测到的概率就可以接近于所要求的后验概率，因此避免了对其进行直接的或模拟的计算。

作为一个实际问题，在贝叶斯 UM 的场景中，MCMC 估计所需后验分布时的准确度出人意料得高，因为产生了该链的大量随机数字。尤其是，本发明的程序通常运行长度为 15000 的链，其中，所产生的该链的前 5000 个模拟被忽略掉，因为它们还没有处于所要求的稳态之下。MCMC 模拟方法是目前可以用于对参数上很复杂的贝叶斯模型进行统计学分析的唯一可行的方法。

回忆一下，统计学分析的精髓是：有时候，在支持所做出的结论的证据很有限时，统计

学分析会很谨慎地避免做出那些超出证据所能支持的范围的结论。在本发明里，这与图 17a 的框 1711 有关，其中，针对某些受试者/技能组合的、技能掌握/未掌握的推断，有时会由于缺乏较强的统计学证据而被禁止做出。

**要求较强的统计学证据以做出掌握或未掌握的推断(图 17a 的框 1711)** 回过头去参考统计学考试的认知实例，Susan 可能被推断为：掌握直方图(技能 1)的后验概率为 0.1，掌握中数/四分位数(技能 2)的概率为 0.53，掌握平均值/均值(技能 3)的概率为 0.81，等等。目前的贝叶斯 UM 认知诊断掌握赋值规则，将后验概率大于 0.65 的赋值为掌握，将后验概率小于 0.35 的赋值为未掌握，其他情况下不对掌握与否进行赋值(参见框 1711；这显然是一个可以改变的惯例)。在赋以掌握或未掌握之前，如果要求很强的证据，那么有时候也使用 0.8 和 0.2 的分界值(cutoff value)。

假定使用了 0.35 和 0.65 的分界值。那么，由于 Susan 的后验概率 0.81 大于 0.65，因此 Susan 被判断为已掌握直方图；由于 0.1 小于 0.35，因此 Susan 被判断为尚未掌握平均值/均值；由于 0.53 在未掌握的分界值之上，在已掌握的分界值之下，因此，对于中数/四分位数掌握情况的判断不能做出。当数据中的信息量不足以为技能的掌握或未掌握情况提供强有力的证据时能够不进行赋值的能力，是 UM 统计学方法的一个真正的优势。

**利用实例 2 中基础统计学考试的认知结构进行的、对用于考试数据的 UMCD 的计算机模拟研究**

此处的目的是两方面的。首先，希望能进一步解释清楚使用目前的 UMCD 的主要步骤，以便搞清楚该程序是如何实现的。第二，给出了本发明在获得认知诊断上的有效性的证据。

构建了一项计算机模拟研究，来证明使用目前的 UMCD 对学生技能的掌握情况进行认知诊断的能力，所述认知诊断是基于实例 2 中(具体的考题/技能结构也可以参考图 19)所提到的基础统计学考试而进行的。下面将顺着图 17a 的流程图介绍该模拟研究。

利用该项考试的认知结构对计算机进行编程，以产生数据。图 18 给出了这次有 40 道考题(图 17a 的框 203)的考试的一组样题，为问题(考题)9—18。选定了前面所述的八项技能(框 201)。图 19 中所给出的考题/技能关系矩阵(框 205)的列表中，给出了技能/考题结构。使用者，在此处是本专利申请人，发展了这一矩阵。

应该回忆一下实例 2 中的八项统计学技能：(1)直方图，(2)中数/四分位数，(3)平均值/均值，(4)标准差，(5)回归预测，(6)相关，(7)回归线，(8)回归拟合。例如，上面的考题 17 要求技能(1)、(3)和(4)。已经注意到，正如在这一模拟实例中所看到的，在使用 UMCD 的时候，使用者通常会构建考试的题目，并确定有待进行诊断的主要技能(可能是先选定技能，然后设计考题以便诊断这些技能)，从而成为 $\alpha$ 的一部分。参考图 19 中的这一考题/技能表格，为了模拟数据的正性和完全性，针对这 40 道考题产生了参数，这些考题考虑到了很低程度至中等程度的不完全性以及很低程度至中等程度的非正性，但总的来讲反映了一项具有高度认知结构的考试，并生成了模拟的受试者反应数据(也就是说，针对这 500 名模拟受试

者中的每一个人，都模拟了一个由 40 个 0 和 1 构成的字符串，表明哪些考题答对了哪些考题答错了)。“很低程度至中等程度的不完全性”是指，受试者能否答对一道考题的概率，主要是基于受试者具有或缺乏与这道题有关的、这八项规定的技能之中的哪几项。模拟数据中的很低程度至中等程度的不完全性是通过在 1.5 和 2.5 之间相当均匀地(fairly uniformly)分布  $c$  值而实现的。假定能够影响考题表现的其他技能(可能有很多)只有很小的影响。

“很低程度至中等程度的非正性”是指，缺乏一道考题所要求的必需技能(在所列出八项技能之中)中任何一项的受试者，很有可能会答错这道题。“很低程度至中等程度的非正性”是通过下述手段实现的：令  $r^*$ 's 在 0 和 0.4 之间是相当均匀的，令  $\pi^*$ 's 在 0.7 和 1 之间是相当均匀的。注意，如前所述，不完全性也是很低程度至中等程度，可以看出，一名已经掌握这道考题所要求的全部技能的受试者将很有可能答对这道题。同样地，一名缺乏至少一项必需技能的受试者将很有可能答错这道题。

产生了针对 500 名模拟受试者的能力  $\theta$  和技能  $\alpha$ ，其中，每一项技能的掌握率为 50%，剩余能力  $\theta$  为标准正态分布。此外，假定技能对(attribute pairs)之间的相关性以及  $(\alpha, \theta)$  对之间的相关性在 0.3 左右，这被判断为是符合实际的。例如，受试者 1 可能被模拟为具有  $\alpha = (01110111)$ ，总计掌握了八项主要技能中的六项。

接下来，对于每一名受试者和每一道题，该模拟在效果上相当于抛掷了一枚硬币，并通过他/她按照方程式 1、2、5、6 对这道题做出正确反应的预测概率进行加权。模拟的样本大小为 500 名受试者(框 207)，因为这是在一个较大的大学里一个学期内选修一门普通大型基础统计学课程的学生总数的近似值(或者比学生总数少一些)。同时，这也是在一个相当大的学区内选修一门核心课程(例如代数 II)的学生总数的适当大小。

这一研究的目的是，观察 UMCD 在重新获取受试者的已知认知能力方面的有效性大小(回忆一下，认知能力是已知的，因为它们是通过输入计算机的已知模拟模型而获得)。为了确定统计学方法例如 UMCD 的有效性大小，在实际的计算机模拟中评估该方法的有效性，是统计学家所采用的几种重要方式之一。实际上，模拟模型以及产生数据的模型参数是已知的这一事实，在利用模拟研究评估统计学程序的有效性方面是很有用的。

图 17a 的框 205、1701、1703 和 1705 构成了所设想的贝叶斯模型，正如公式 1、2、5、6 中所给出的那样。按照图 16 的示意图中所给出的可识别贝叶斯 UM，利用 MCMC(框 1707)对模拟的受试者反应数据(一个由 0 和 1 构成的 500 乘 40 的矩阵)(框 207)进行分析。对于每一个受试者技能组合，产生了一条长度为 15,000 的链，其中前 5000 个值舍去，以避免链的起始值可能带来的任何潜在影响(框 1707)。按照 MCMC 理论，具有 10000 个值的这条链估计了每一个受试者技能掌握的后验分布，而这正是我们想要得到的。例如，如果对于技能 4 受试者 23 得到了 8500 个 1 和 1500 个 0，那么基于模拟数据而得到的、受试者 23 掌握技能 4 的后验概率为  $8500/10000 = 0.85$ (框 1709)。按照该程序，如果后验概率大于 0.65，那么受试者将被认为已掌握某项技能，如果后验概率小于 0.35，那么受试者将被认为尚未掌握该项技能(框 1711)。在实施本发明的时候，对于掌握/未掌握的这些设定可以进行修改。



该程序极其有效，针对 96.1% 的受试者/技能组合，正确地诊断出了技能掌握或未掌握的情况(8 项技能乘以 500 名受试者等于 4000 个受试者/技能组合，然后减去 176 个受试者/技能组合，由于证据较弱，后验概率在 0.35 和 0.65 之间，因此针对这 176 个受试者/技能组合的诊断没有做出)。考虑到所采用的考试是一项包括 40 道选择题涉及 8 项技能的中等长度的考试，因此，认知诊断如此准确是令人印象很深刻的。事实上，如果要求更强的证据，并采用 0.8 和 0.2 的分界值，那么诊断的正确率可以提高到 97.6%，但是，针对 456 个技能/受试者组合的诊断没有做出。这提供了很强的科学证据，以证明该程序作为一种认知诊断工具是有效的。

考题参数也进行了很好的估计(标定)。 $\pi^*$  的估计值和真实值之间的平均差异以及  $r^*$  的估计值和真实值之间的平均差异为 0.03(这两种参数的范围均为 0 到 1)， $c$  的估计值和真实值之间的平均差异为 0.3(其范围为 0 到 3)。正如我们所预期的，对  $c$  值的估计不如对  $\pi^*$  值和  $r^*$  值的估计准确，因为这次考试被设计为具有高度认知的结构(也就是说，正性和完全性相对较高)，而且设计这次考试的目的是为了测试一组被模建为对技能理解得较好的受试者(也就是说，其中的很多人都是技能的掌握者，因此可以期望他们具有相对较高的  $\theta$  值)。虽然该模型在参数上很复杂，但是，很好地估计关键参数并因此很好地标定模型仍然是可能的。正因为如此，不会受到方差/偏差权衡的伤害，这一点在前面的实际上遵循四参数三次多项式模型的数据实例中已经描述过了。在那种情况下，该场景可能会由于通过对有偏差的线型模型中的这一个参数计算出可靠的估计值而被错误地表征，该场景也可能会通过对无偏差的三次多项式模型中的四个参数计算出不可靠的估计值而错误地描述。与此不同，在此处的 UMCD 模拟中，复杂的、拟合很好的 UM 的参数可以被很好地加以估计。

通过可识别并且易于解释的参数所表达的正性和完全性的构想，对教育上的从业者来说在直观上是很容易掌握的。而且，这些构想给实践者提供了现实而易行的方法，用来对基于技能的受试者反应的内在随机性进行建模。此外，特别地，用来处理不完全性的潜在变量  $\theta$  的引入，为教育上的从业者在选择基于 UM 的认知模型中明确包括哪些以及多少技能方面，提供了巨大的自由。最后，允许使用者对技能掌握水平进行明确的控制是很重要的，UM 的贝叶斯部分中所采用的技能的正相关结构也是很重要的。事实上，认识到一个人应该选择贝叶斯模型——特别地，是假定了正性联系的技能并为技能选择了适当先验的贝叶斯模型——的这一事实，解决了存在于 1995 UM 实施过程中的一个主要的实际问题，也就是说，它不能利用技能总是正相关的这一事实，而在进行认知诊断时，这一事实对获得较高的准确度而言是很有用的(当使用的时候!)。实际上，模拟研究表明，引入了技能间的正相关结构的贝叶斯 UM 比未引入这一正相关结构的贝叶斯 UM 表现得好得多。应该明确，目前版本的 UM 诊断方法所引入的一个主要贡献是，认识到应该构建可以获得正相关技能的概率建模结构，以及意识到采用贝叶斯概率建模方法是实现这一目的的一种优秀的方式。

在教育考试服务中心资助下，研究了真实数据考试/再考试 PSAT 场景，其中，UMCD 方法按照两次考试中技能掌握/未掌握的情况，成功地将 2/3 以上的受试者进行了一致的分类

(两次考试都赋以技能已掌握, 或两次考试都赋以技能未掌握)。这是令人印象特别深刻的, 因为从设计上看, PSAT 这项考试在提供具体技能的认知信息方面是很弱的。

UMCD 之所以在认知诊断的表现上不同于并且超过这些其他方法的原因有好几个。如前所述, 这些其他的方法所使用的模型与贝叶斯 UM 方法所使用的模型不同。此外, UMCD 是唯一一种同时具有如下特点的模型: 在统计学上易于处理, 含有可识别的模型参数, 规定了技能掌握水平, 在其认知诊断中引入了数据中技能的正联系, 在允许不同的认知科学观点以及引入预测的受试者错误以便得到合适的认知推断警告这两方面都很灵活, 其中, 所述的模型参数能够使数据进行较好的模型拟合, 并且具有有意义的认知解释, 易于被使用者解释。当方法不能很好地描述真实的认知现实时, 其他的模型可能会是不切实际的(因为它们依附于具体的认知建模方法)。它们通常难于解释, 因为它们所具有的参数很难被使用者解释清楚, 而且不易理解, 特别是对于普通的教育从业者。而且, 很多此类模型看上去与数据拟合得不是特别好, 而这对于一个统计程序的有效工作来说是绝对必需的。而且, 它们之中没有一种方法能解释清楚规定技能掌握程度这一重要概念。

**将本发明的 UM 方法应用到医学/精神病学诊断之中** 医学诊断模型是很有用的, 可以协助从业者提出由可能疾病的列表所构成的诊断结果, 所述的列表是由从业者根据患者所具有的症状而得到的, 但这些模型不能替代从业者。因此, 一个好的系统会合理地给出完整的可能疾病列表, 尽管在患者信息较为充分的情况下, 疾病的数目应当是可以把握的。

图 17b 是本发明所使用的 UM 医学/精神病学诊断程序的流程图。应该将其与图 17a 的流程图相比较, 图 17a 中给出了用于认知诊断的类似 UM 程序。潜在疾病的集合取代了技能的集合(框 201'), 症状和其他患者特征的集合取代了考题(框 203'), 其中, 所述的患者特征包括下列特征: 例如二值性的实验室测试的数值, 年龄, 种族, 性别等。 $\theta$  是潜在健康或潜在生活质量变量(latent quality of life variable), 它整合了所有不属于明确列在模型中的潜在疾病的潜在健康变量及生活质量变量。然后, 应用 UM 的方式与教育诊断条件下应用 UM 的方式完全相同(图 17a)。具体地说, 定义了症状/特征和疾病(框 201'和 203'), 然后构建关系矩阵, 以指出哪些疾病可能与某一具体症状/特征的存在有关联(框 205')。考题参数  $\omega$ (如框 1701、1703、1705、1707'中所使用的)现在是症状/特征参数, 如果用于标定模型的数据集(框 207')中包括患有已知疾病的患者的话, 这些参数实际上可以被准确地估计。这将提高症状/特征参数标定的准确率(框 1707')。然后, 基于由 UM 估计程序所计算出来的后验概率, 某一个具体的患者可以被赋以一个疾病列表, 其中, 他/她患有所述疾病的概率足够高(框 1711')。给潜在诊断从业者的报告中可以包括给每种疾病所赋予的后验概率(框 1709')。在这两种情况下, 统计分析以相似的方式进行(框 1701、1703、1705、1707'、1709'、1711')。然后, 将该诊断用于支持从业者的诊断工作(框 1713')。

这种情形与教育测量情形的一个区别(除了在精神病学中)是: 对于大部分疾病而言, 存在一个“黄金标准”的诊断。因此, 可以利用患有已知(因此不是潜在)疾病的患者来进行“症状/特征标定”。

将本发明的 UM 用于除教育或医学/神经病学以外的新环境中 图 17c 示出了将本发明用于一般场景下的流程图。应该将图 17c 和图 17a 中所示的用于教育环境的本发明 UMCD 认知诊断的流程图相比较。需要如下对应关系：

技能	性质	框 201'', 205'', 1709'', 1711''
考题	探针	框 203'', 205'', 207'', 1707''
考题/技能关系矩阵	探针/性质关系矩阵	框 205''
认知诊断	潜在诊断	框 1711''

在这两种情况下，统计分析以相似的方式进行(框 1701、1703、1705、1707''、1709''、1711'')。因为场景是一般性的，因此，关于其应用，只能说潜在诊断的结果将会用于做出推断或可能的决定，而所述的推断或决定与本发明所应用的真实世界场景有关。

对于支持本发明的方程式和关系式的一般性结构的半定性描述 方程式 1、2、5、6 以及  $\pi^*$ 、 $r^*$ 、 $c$ 、 $\underline{c}$ 、 $\theta$  的定义被用于解释本发明具体实施例的某些部分。图 17a、图 17b 和图 17c 示出了本发明的流程图，每一个流程图的用途都不一样。为了方便起见，此处将使用术语“认知诊断”(图 17a)，注意，术语“医学和精神病学诊断”(图 17b)或者术语“一般性诊断”(generic diagnosis)(图 17c)在功能上将是一样的。

通过中间的、非方程形式的表达方式来描述本发明的必要组成部分是很有用的。方程式 1、5、6 与它们的可识别的且能被标定的参数  $r^*$ 's 和  $\pi^*$ 's 一起，说明了如下事实：(i)如果一名受试者掌握了一道考题所需的全部技能，那么与缺乏这些必需技能中的一项或多项相比，答对这道题的概率增加；以及(ii)未掌握的必需技能越多，答对这道考题的概率越小。上面的(i)和(ii)款定性地描述了考题正性的概念，这一概念在本发明的实施例中以一种具体的方式表达了出来。一般来讲，任何一组模型方程式都可以用于在本发明所使用的 UM 中被用来体现(capture)正性的概念，其前提是：方程式的参数可以识别、对从业者来说有实质性意义并且表达了(i)和(ii)或者单独表达了(i)。

UM 建模的完全性的特征在于，使用一个或少数几个潜在变量以发现对答对考题概率的影响，所述的影响是由模型中未通过关系矩阵(框 205、205'、205'')明确列出、但很有影响的全部技能所产生的。除了本发明的  $P(\theta_j + c_i)$  以外，任何满足下列条件的表达式，都是用于建模 UM 完全性的可以接受的方式，所述条件为：能表达出这一事实，即除了 UM 关系矩阵中明确列出的技能以外，还有些技能可能会影响答对一道考题的概率；并且可以通过一个或少数几个潜在变量保守地(parsimoniously)体现这种影响。如图 16 的示意图所示，目前的实施例通过设定参数  $p_k$  的值而规定了技能掌握水平，注意，目前用于设定掌握水平的方法是与本发明的贝叶斯建模方法联系在一起。但是，任何一种能够使基于技能的认知诊断程序的使用者定量设定技能掌握水平的方法，都是可以使用的。

此外，任何能对技能之间的联系进行建模的方法都是可以使用的；这并不一定非要通过

在贝叶斯框架内使用图 16 的  $\sigma_{kk}$  而完成。

另外，除了 0/1 关系矩阵(正如目前所进行的：参见图 19)这种方式以外，人们还可以以其他方式来表达这个事实，即：为了能成功地解答出来，每一道题都要求某些特定的技能。

因此，总的来说，任何满足下列条件的方法，都能表达本发明所要求保护的 UMCD 的几个方面，所述条件如下：能阐明对可识别的、表达正性和完全性的参数的需求；规定技能掌握水平；将技能构建到模型中；以及能够表达规定技能的一个子集(a subset of the specified attributes)对每道考题的依赖性，其中，第三项中所述的技能往往在教育场景(educational setting)中呈正联系，或者，也可能在其他场景中呈正联系或负联系。

虽然本发明的一种优选应用方式是使用 UM，但是应当理解，本发明的特征也可以不基于 UM 的方式应用到认知建模和诊断程序中。具体地说，任何与对象有关的模型——所述的对象通常是人，并且与具有两个值的潜在属性有关——都可以规定每种属性的水平(例如规定掌握的水平或规定疾病的水平，所述疾病被判断为患有该疾病的病人所具有的)；也可以对属性(例如技能或疾病)之间正或负的联系进行建模，从而在进行诊断的时候，可以进行标定并进一步利用估计出来的所述联系的大小，以便提高准确率。

上面所引用的全部出版物，在此均以引用的方式将其全部内容包括在本文中。

图 1  
(现有技术)

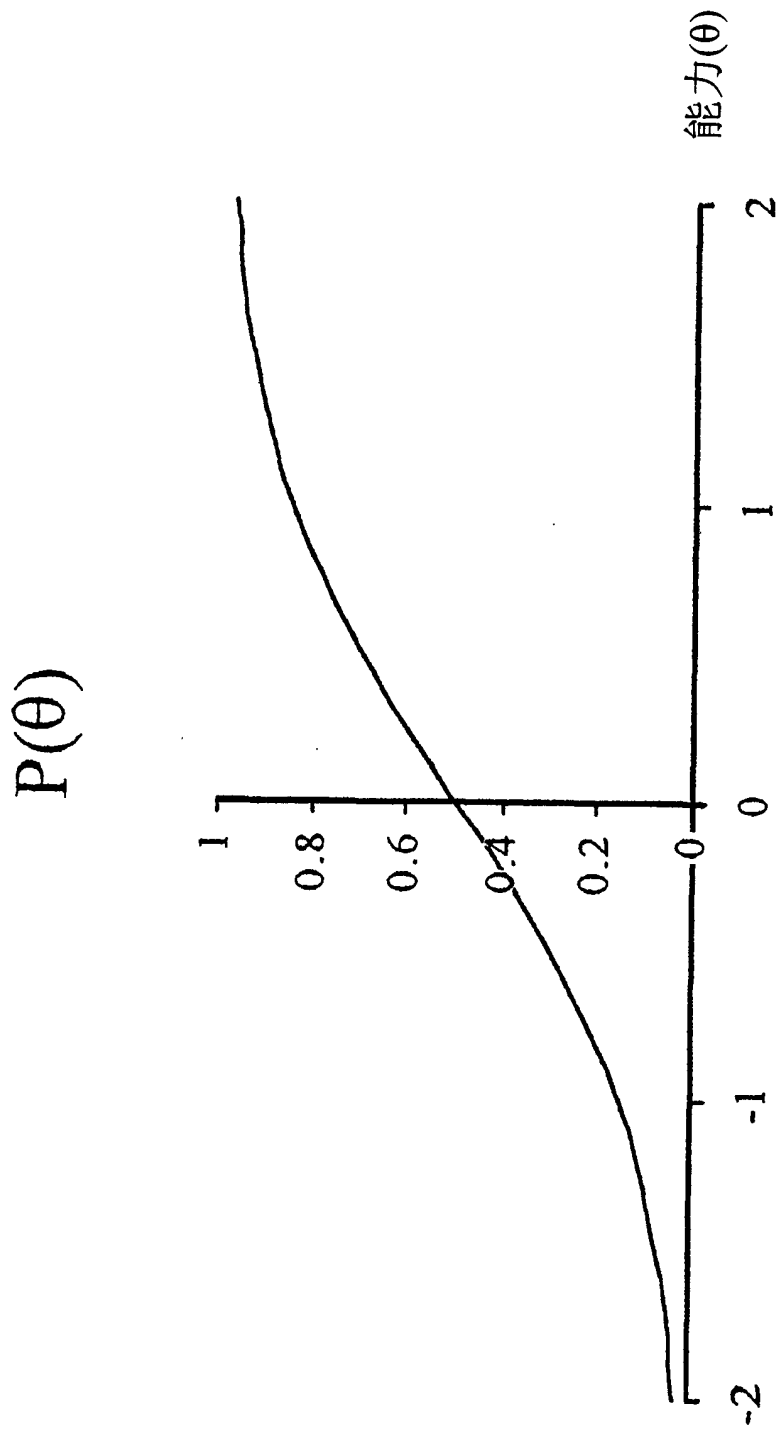


图 2  
(现有技术)

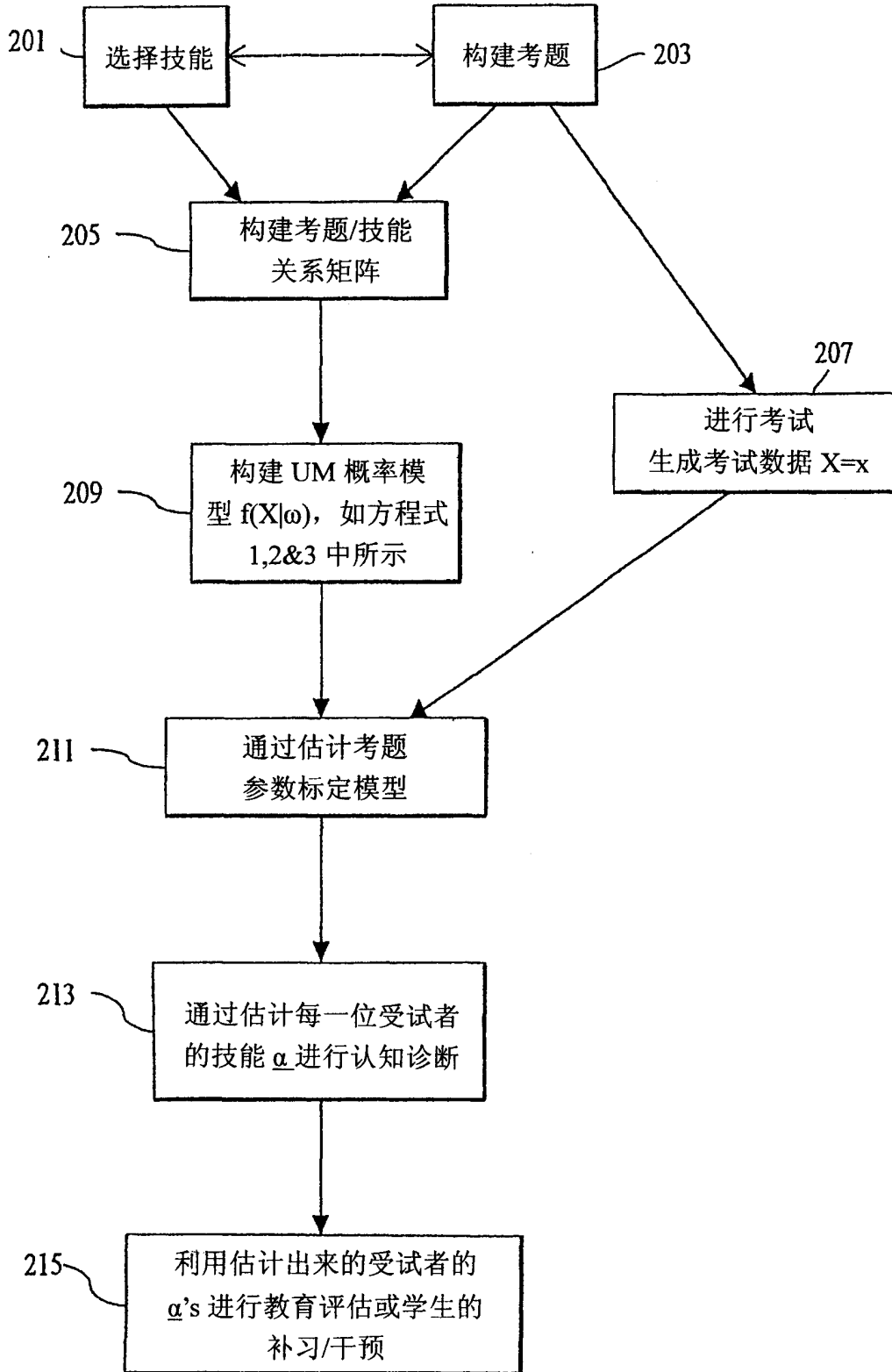


图 3  
(现有技术)

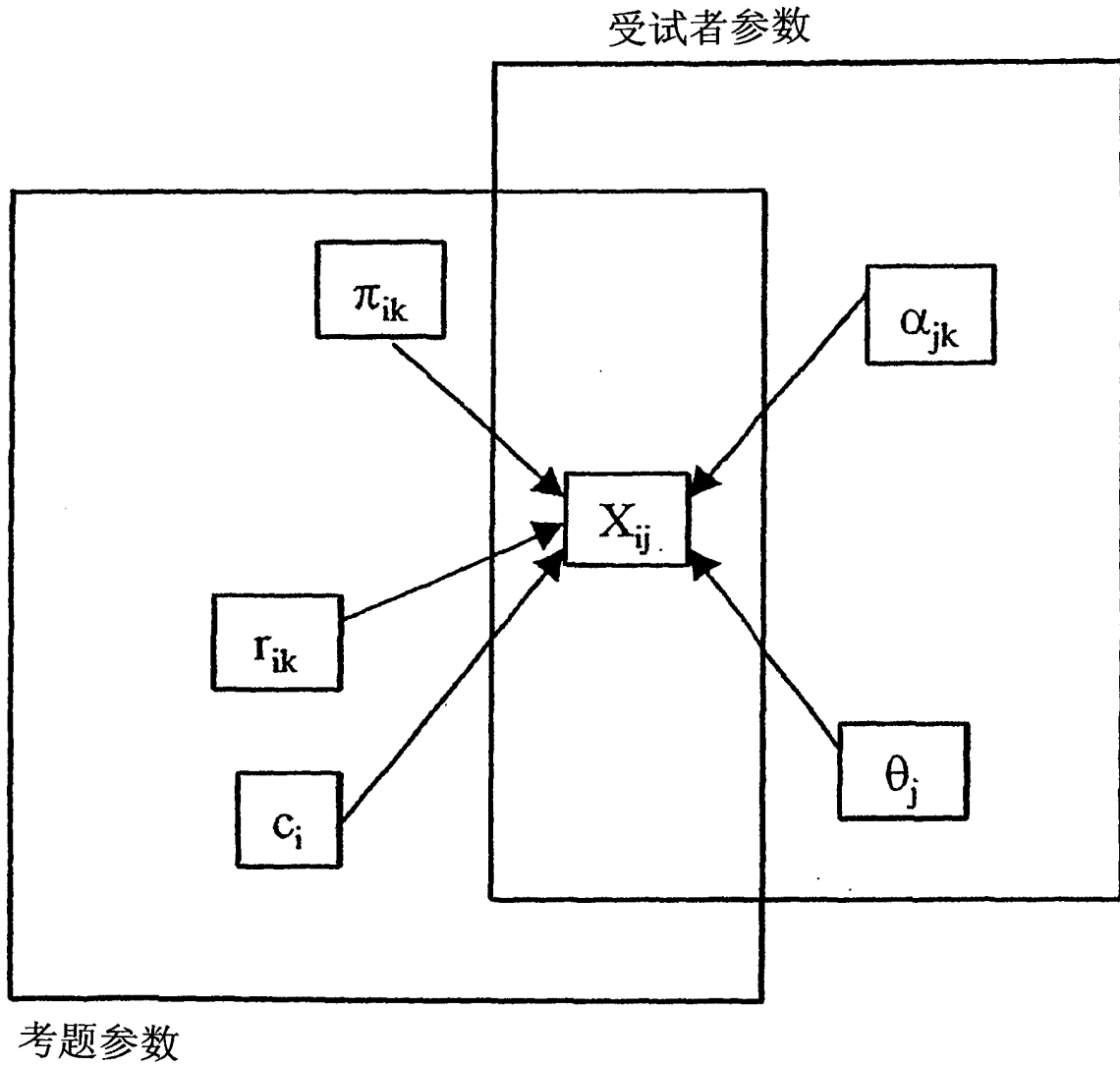


图 4  
(现有技术)

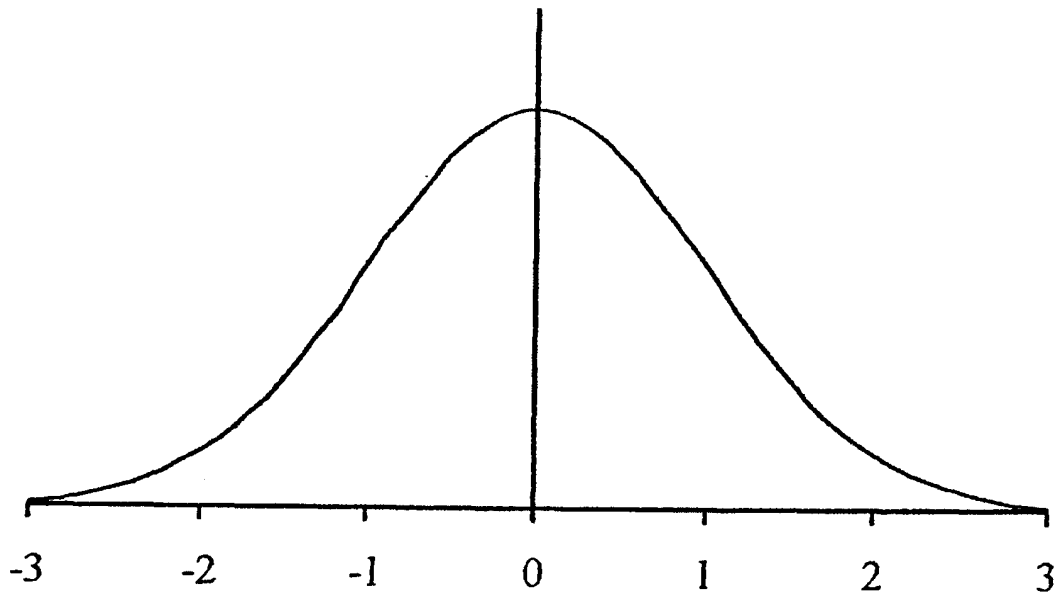


图 5  
(现有技术)

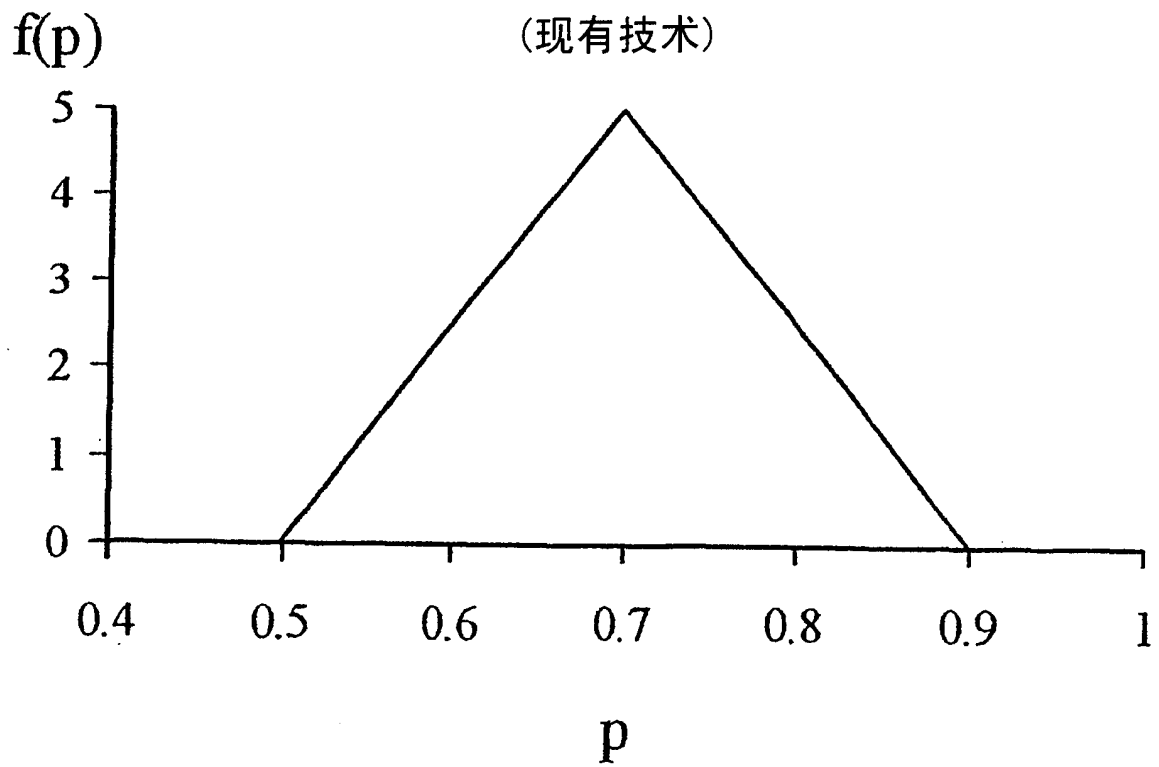




图 6

(现有技术)

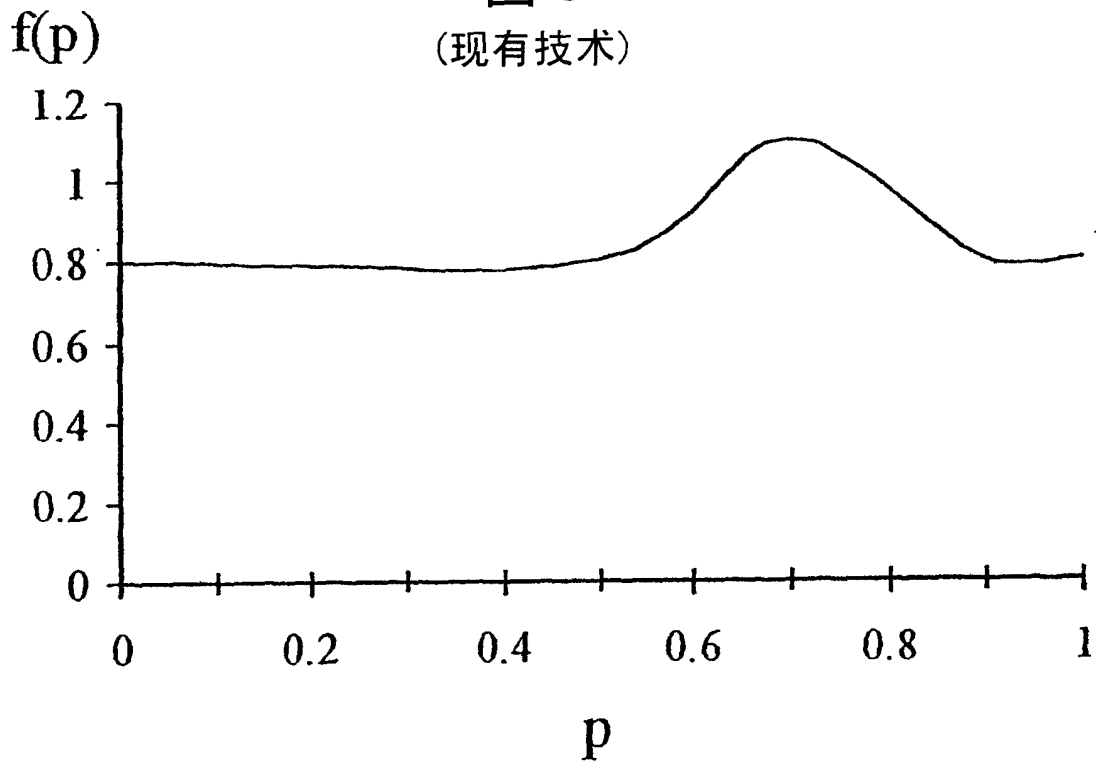


图 7

(现有技术)

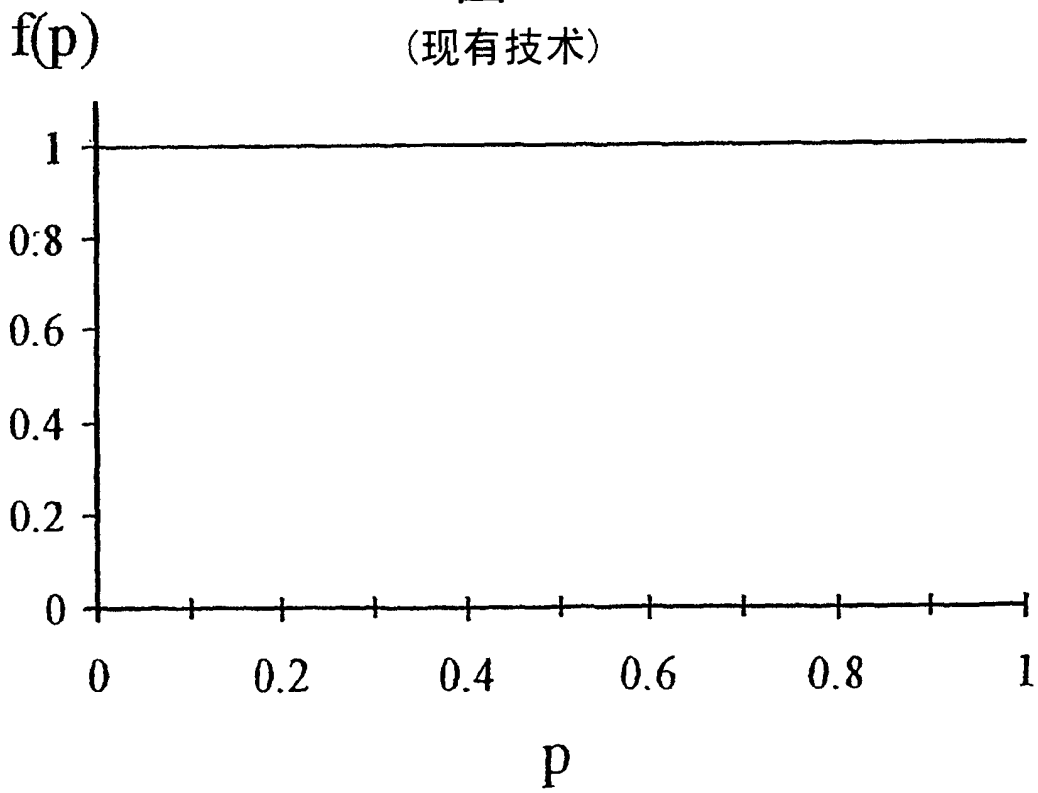


图 8

(现有技术)

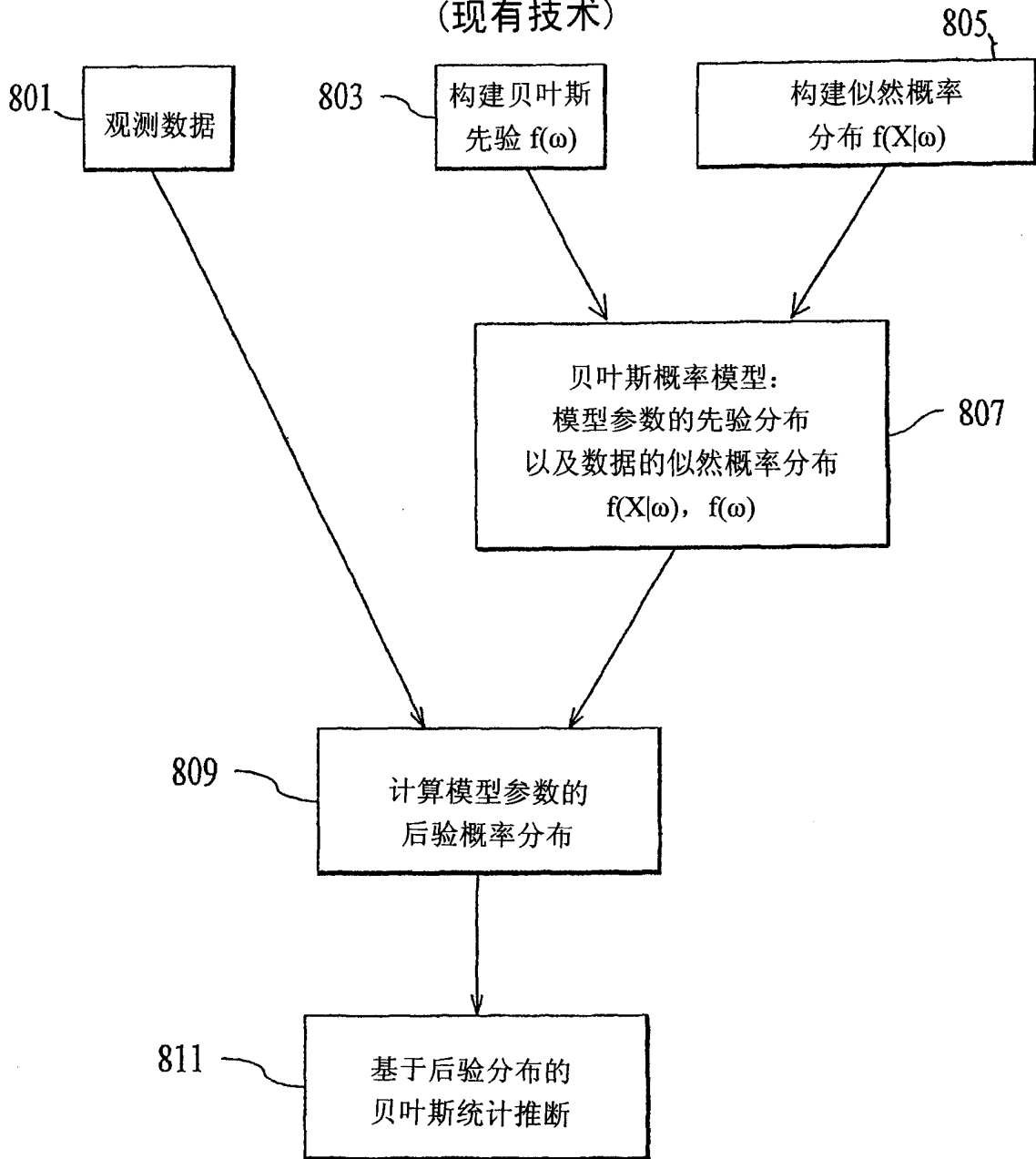


图 9

(现有技术)

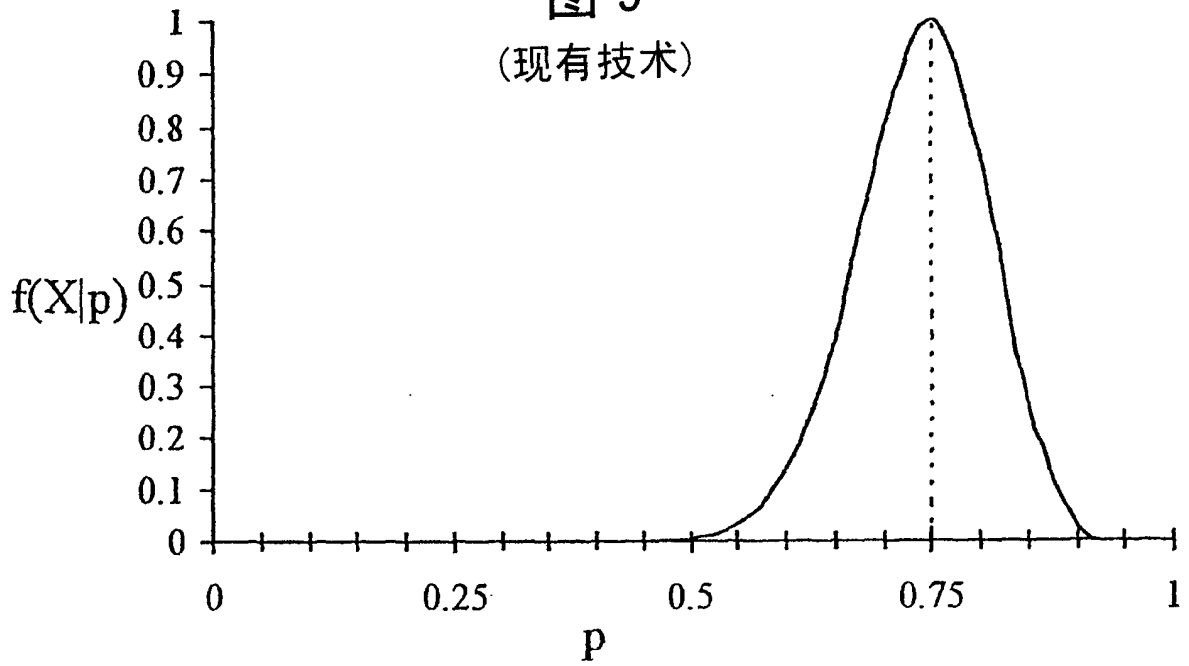


图 11

(现有技术)

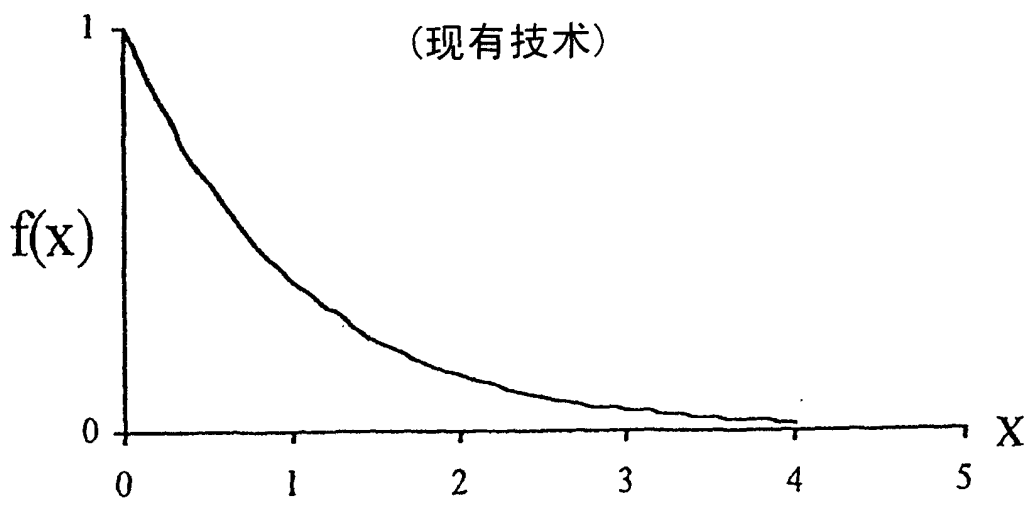


图 10  
(现有技术)

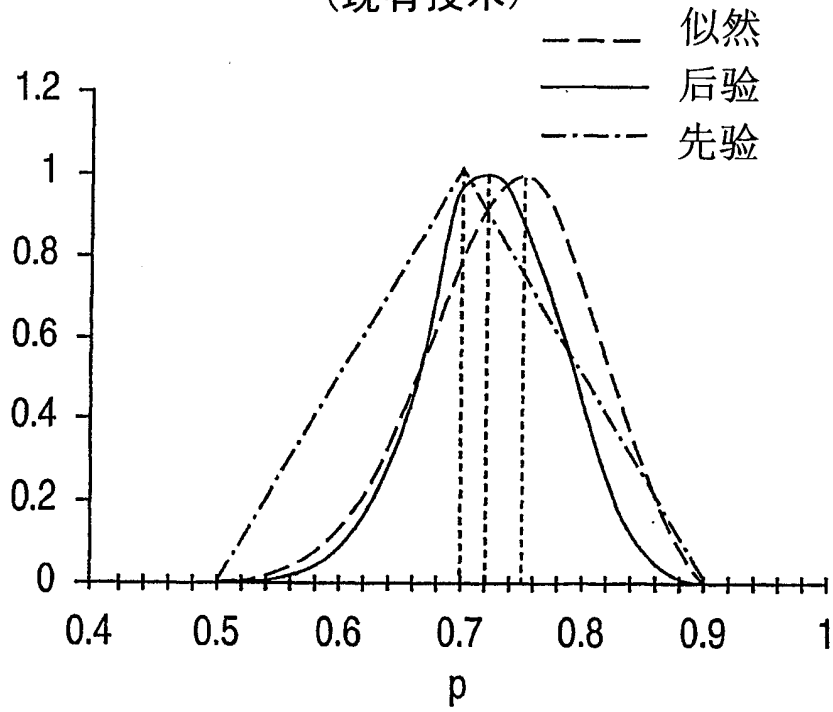


图 12  
(现有技术)

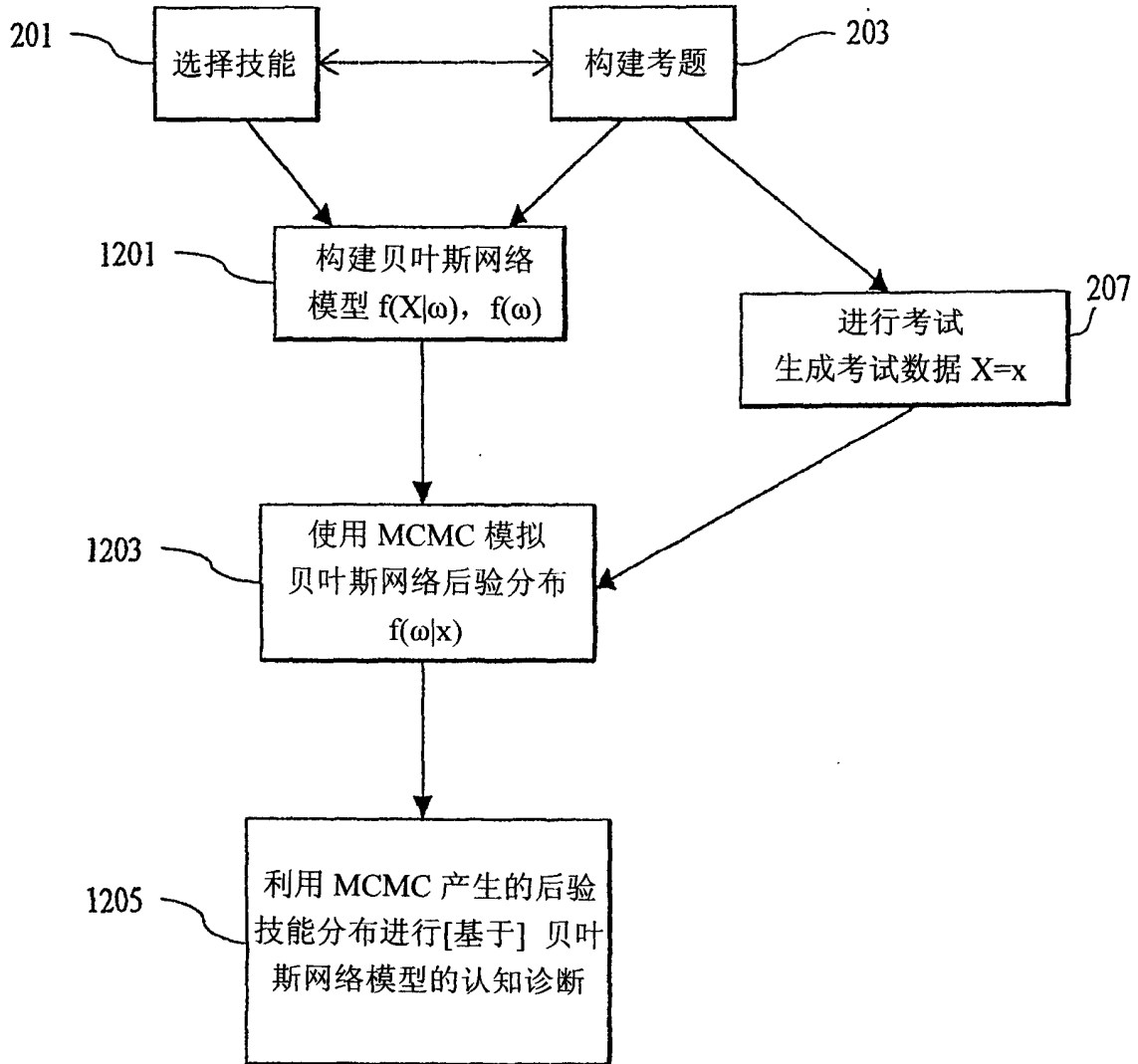


图 13  
(现有技术)

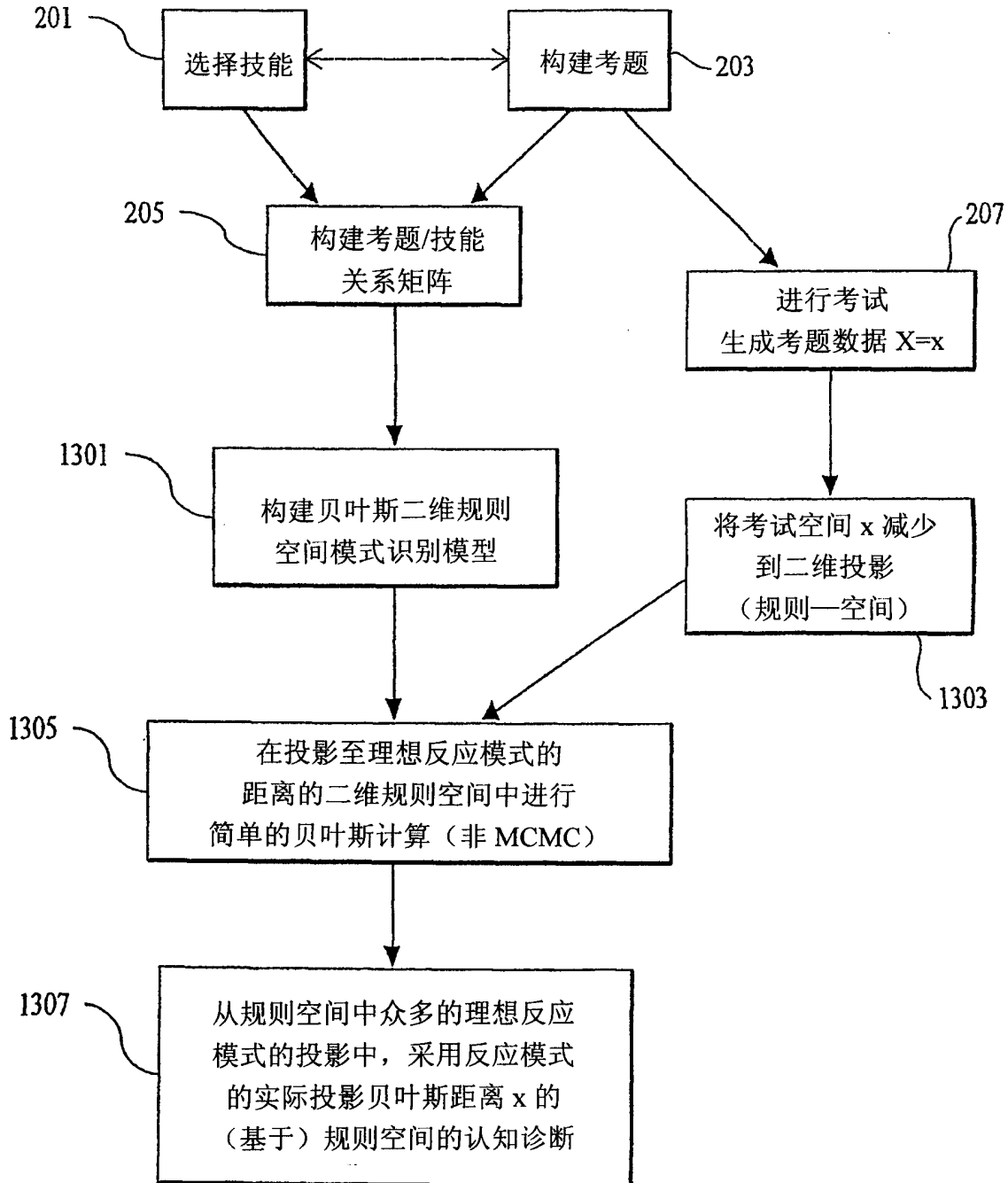


图 14  
(现有技术)

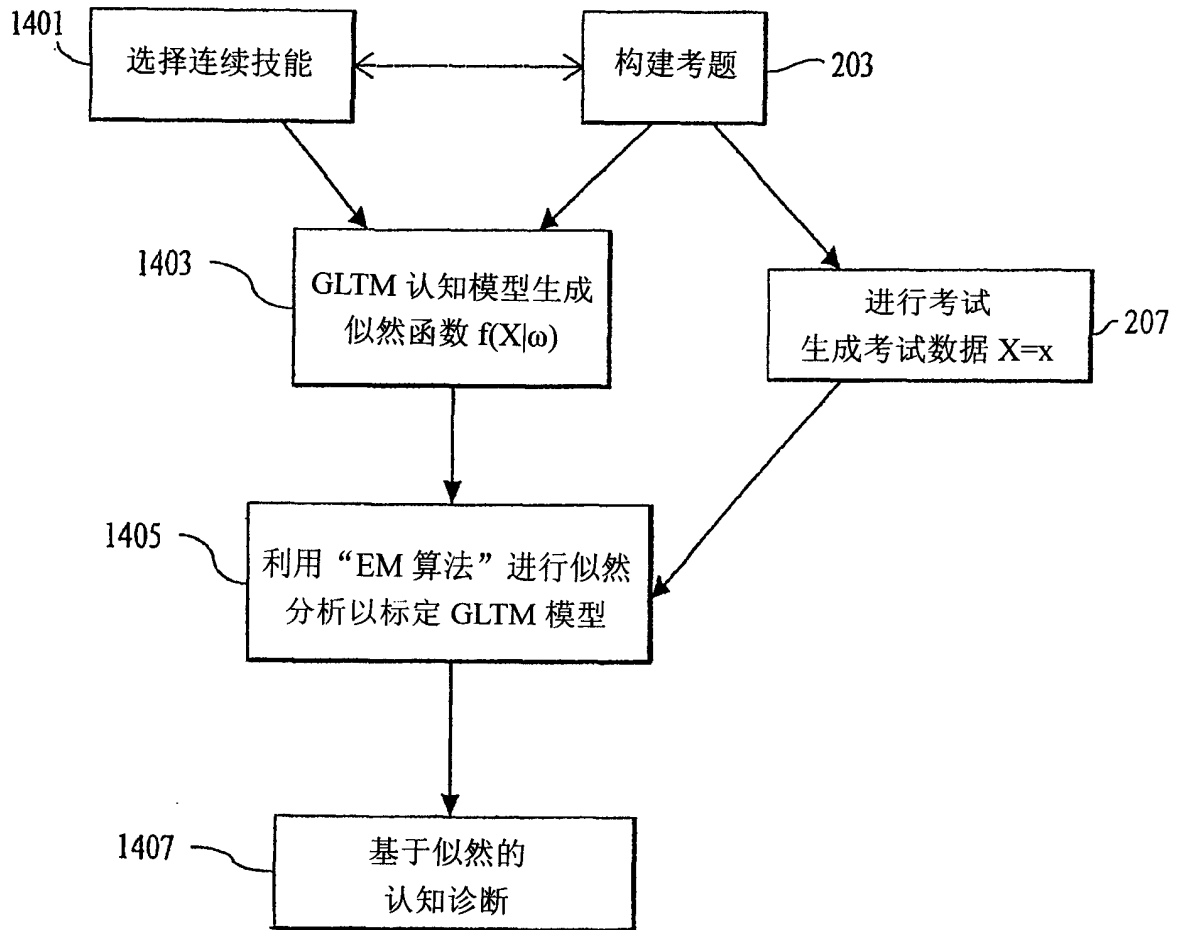
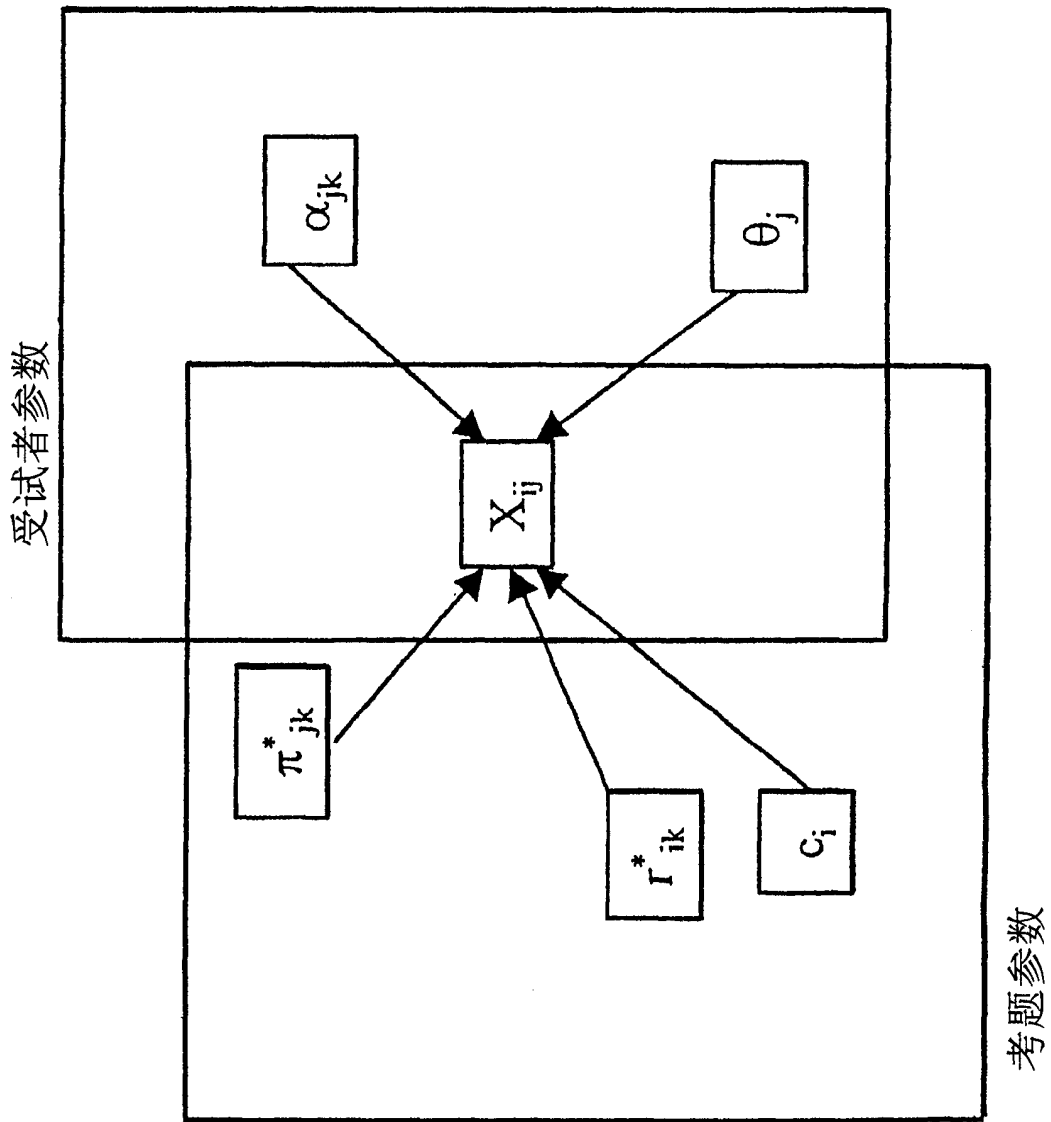


图 15





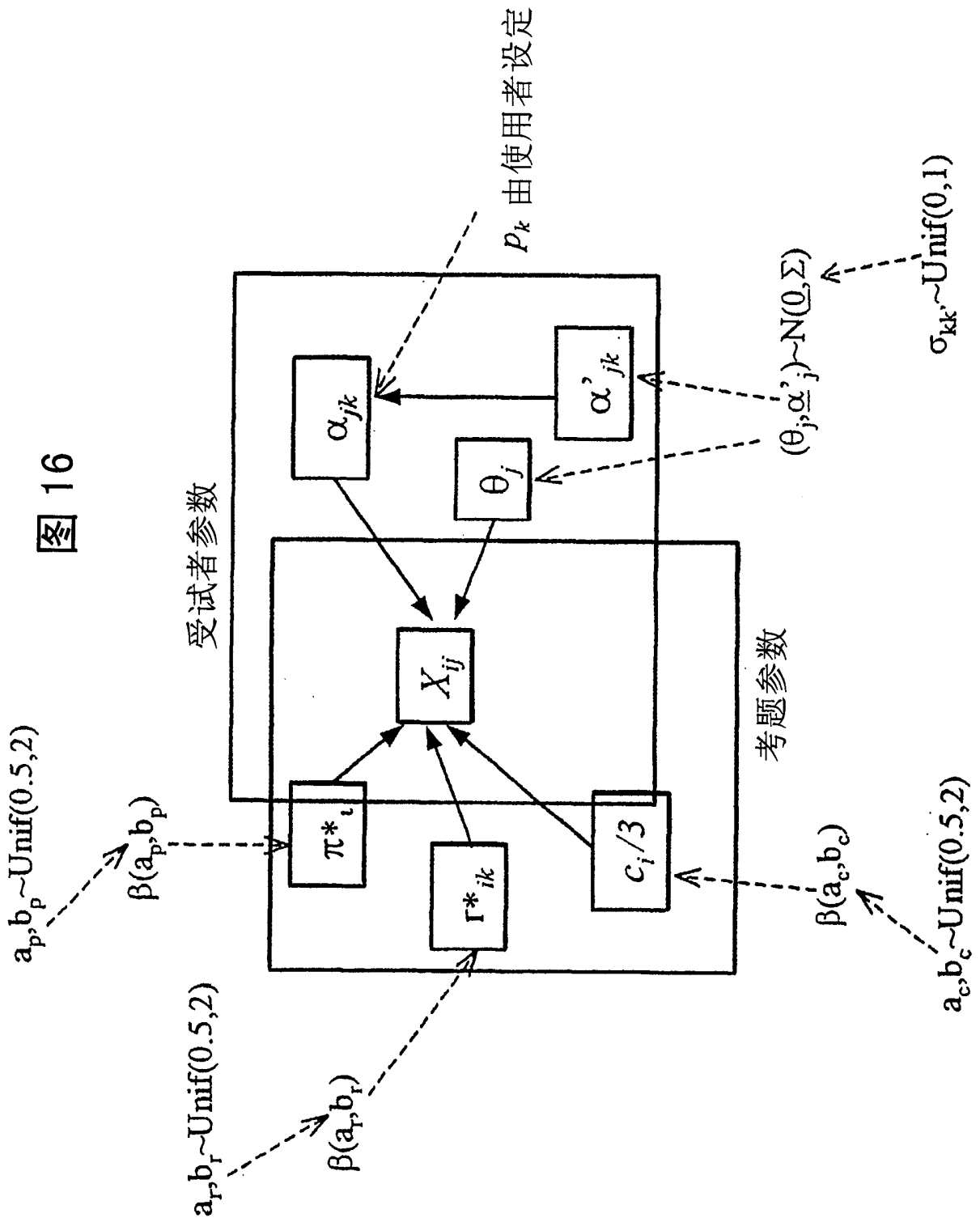


图 17a

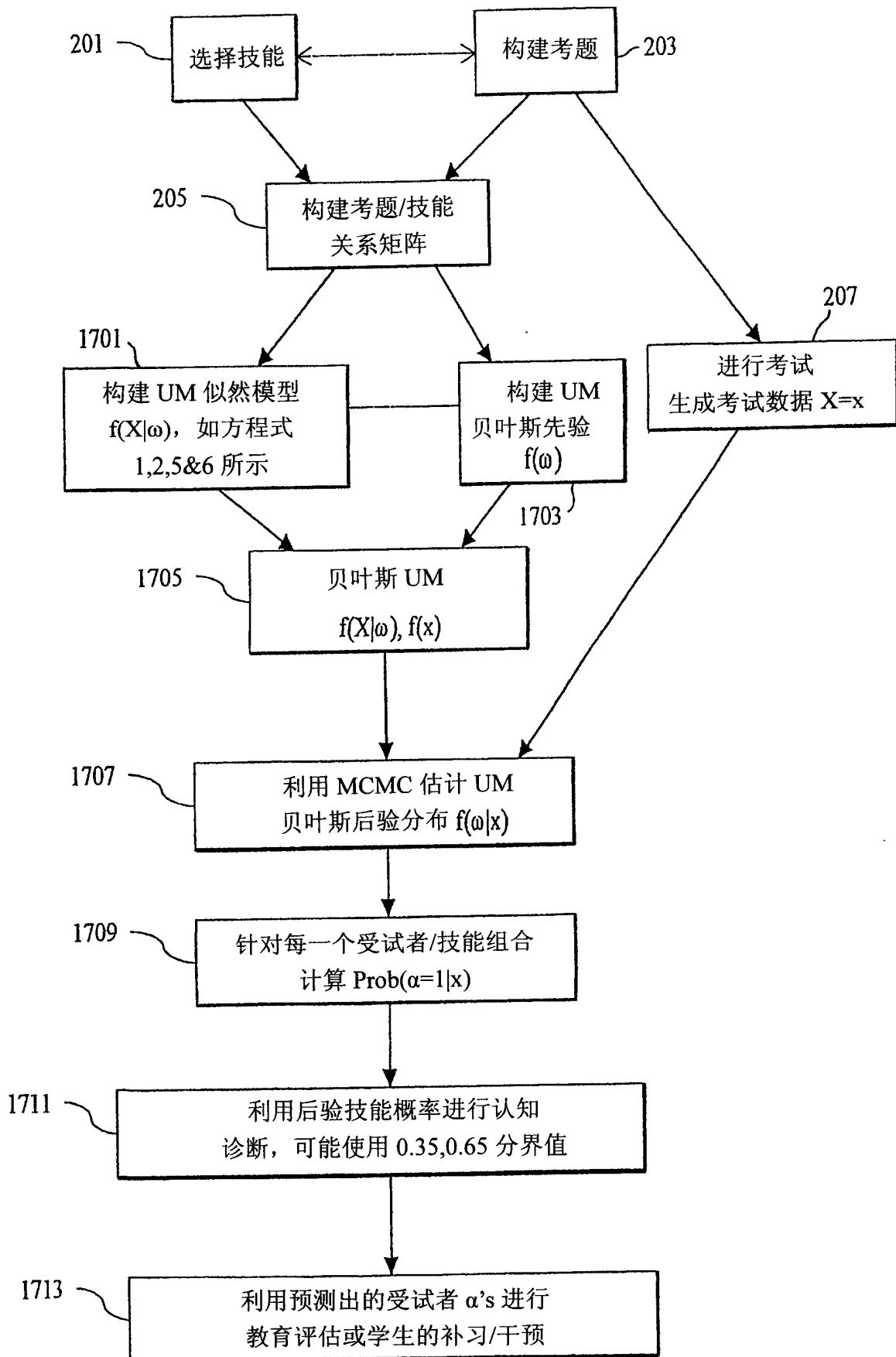


图 17b

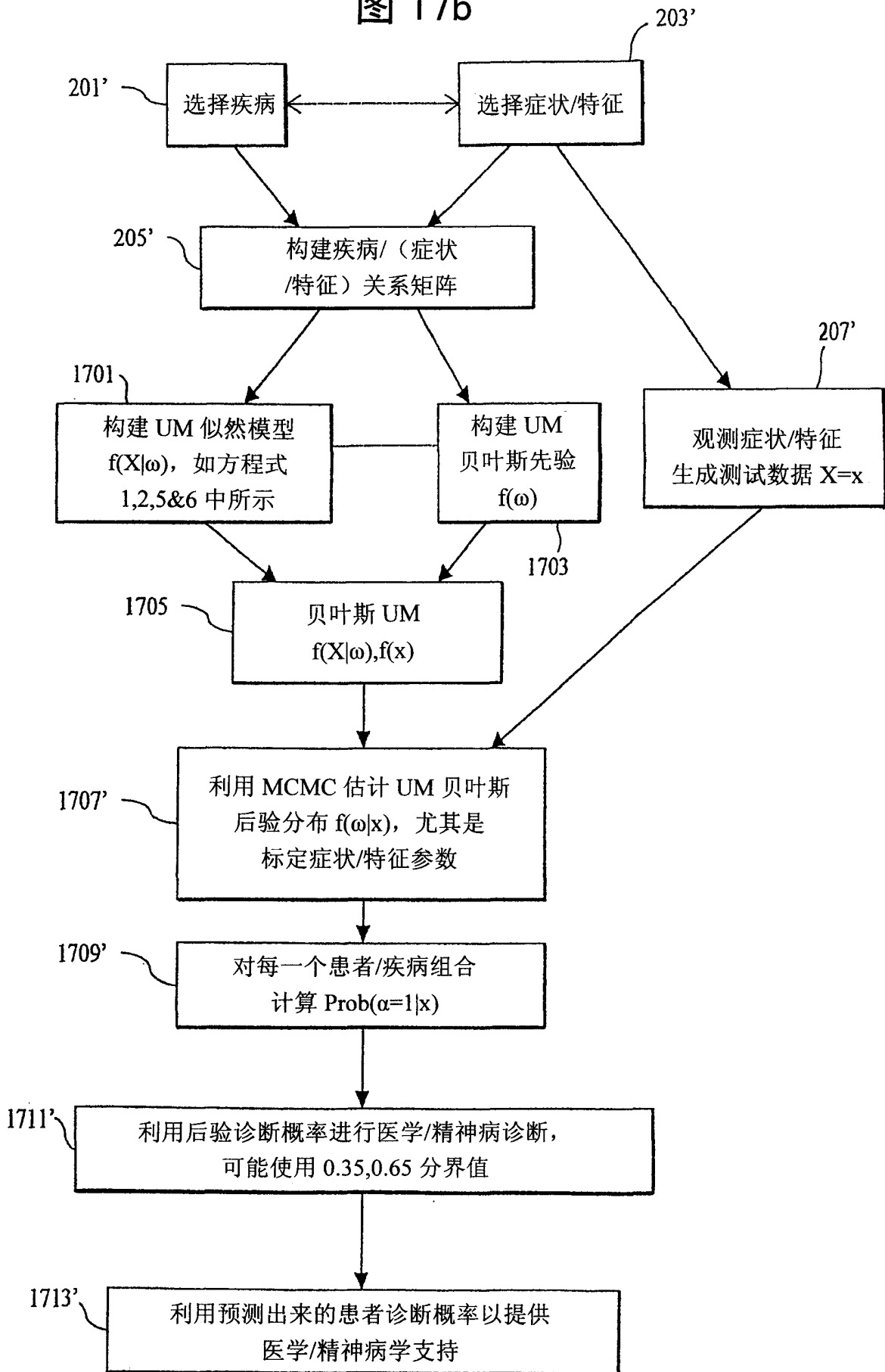
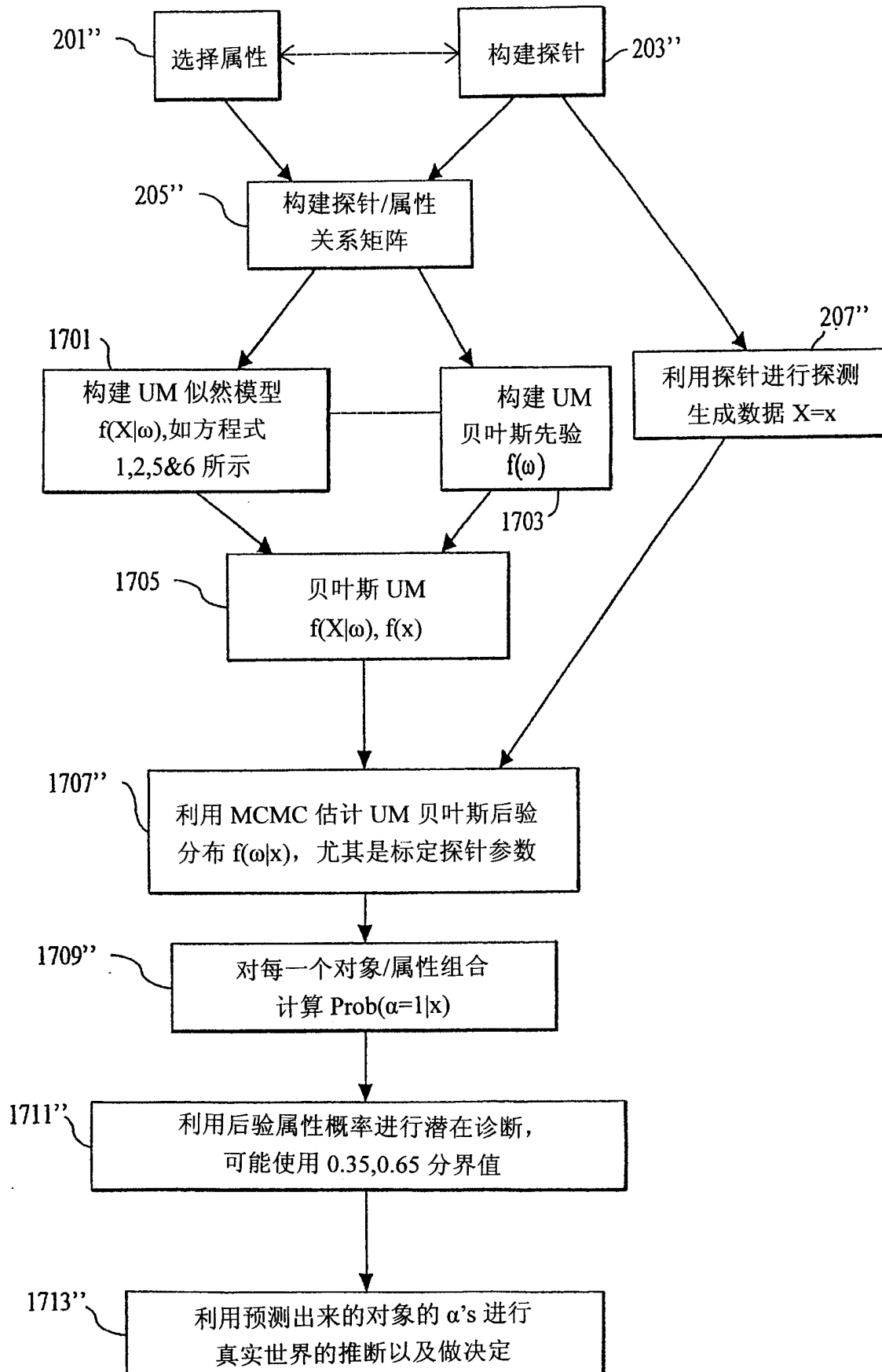


图 17c



## 图 18

[II] 考虑 7 个观察值：10, 4, 2, 5, 4, 2, 8。

9. 平均值是

- (a) 2                      (b) 3                      (c) 4                      (d) 5                      (e) 8

10. 标准差是

- (a) 2                      (b) 3                      (c) 4                      (d) 5                      (e) 8

11. 中值是

- (a) 2                      (b) 3                      (c) 4                      (d) 5                      (e) 8

12. 第一四分位数是

- (a) 2                      (b) 3                      (c) 4                      (d) 5                      (e) 8

13. 第三四分位数是

- (a) 5                      (b) 6.5                      (c) 8                      (d) 9                      (e) 10

[III] 以下四个问题一般地涉及数据集和它们的直方图。

14. 所有大的数据集都具有钟形的直方图。这种说法是否正确？

- (a) 正确                      (b) 错误

15. 对于大的数据集，非常接近 50%的数据小于平均值且非常接近 50%的数据大于平均值。

- (a) 对于每一个这样大的数据集是正确的  
(b) 对于某些大的数据集是错误的

16. 中值比作为中间的测量值的平均值更可取，当数据集

- (a) 是大的。  
(b) 有离开本体的值（异乎寻常地大或异乎寻常地小的值）。  
(c) 是对称的但不是钟形的。  
(d) 有奇数个点。

17. 假设一个大的数据集具有一个大致上是钟形的直方图。假设没有离开本体的值。则

- (a) 大约 68%的数据处于  $\bar{x} \pm s$  之内。  
(b) 正好 68%的数据处于  $\bar{x} \pm s$  之内。  
(c) 大约 95%的数据处于  $\bar{x} \pm 3s$  之内。  
(d) 正好 95%的数据处于  $\bar{x} \pm 3s$  之内。

18. 中值比作为中间的测量值的平均值更可取，当数据集

- (a) 是小的。  
(b) 是对称的但不是钟形的。  
(c) 显著地向左侧倾斜。  
(d) 包含生物的或金融的数据。

图 19

考题	技能							
	1	2	3	4	5	6	7	8
1	1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0
5	1	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0
7	1	1	0	0	0	0	0	0
8	1	0	1	0	0	0	0	0
9	0	0	1	0	0	0	0	0
10	0	0	0	1	0	0	0	0
11	0	1	0	0	0	0	0	0
12	0	1	0	0	0	0	0	0
13	0	1	0	0	0	0	0	0
14	1	0	0	0	0	0	0	0
15	0	1	0	0	0	0	0	0
16	0	0	1	0	0	0	0	0
17	1	0	1	1	0	0	0	0
18	0	1	1	0	0	0	0	0
19	0	1	1	0	0	0	0	0
20	1	0	1	0	0	0	0	0
21	1	0	0	1	0	0	0	0
22	1	0	0	1	0	0	0	0
23	0	0	0	0	0	1	0	0
24	0	0	0	0	0	1	0	0
25	0	0	0	0	0	1	0	0
26	0	0	1	0	0	0	0	0
27	0	0	1	0	0	0	0	0
28	0	0	0	1	0	0	0	0
29	0	0	0	1	0	0	0	0
30	0	0	0	0	0	1	0	0
31	0	0	0	0	0	0	1	0
32	0	0	0	0	0	0	1	0
33	0	0	0	0	0	0	1	1
34	0	0	0	0	0	0	1	1
35	0	0	0	0	0	0	1	1
36	0	0	0	0	0	0	1	1
37	0	0	0	0	1	0	0	0
38	0	0	0	0	1	0	0	0
39	0	0	0	0	1	0	0	0
40	0	0	0	0	0	1	0	0
总计	13	7	9	6	3	5	6	4