

# Classification Consistency and Accuracy

## 分类一致性和准确性



汇报人：黄颖诗

2019-05-28

# 基础知识

## ☑ 分类背景、意义

## ■ 分类一致性和准确性

### ➤ 1.1 分类背景与意义

#### (1) 使用测验的目的:

- 考察考生得分在群体中的相对位置;

➡ 高考、期末考试...

- 考察考生是否达到某一能力水平标准。

➡ 学业水平考试、四六级考试...



不通过 < 425 < 通过

测量结果直接决定学习进程、考生选拔、教学质量评价

# 基础知识

## ■ 分类背景、意义

## ☑ 分类一致性和准确性

## ➤ 不同测验目的下如何进行测验质量评价？

### (1) 传统信效度

- **信度**：采取同样的方法对同一对象重复进行测量时，其所得结果相一致的程度（例如：分半信度、 $KR-20$ 、 $KR-21$ 、 $\alpha$ 系数）；
- **效度**：测量工具或手段能够准确测出所需测量的事物的程度。

➡ 真分数方差与观察分数方差之比

### (2) 分类考试：

- 整个测验的得分变异较小；
- 类别内的分数高低无意义。

**传统信效度指标不再适用**

# 基础知识

## ■ 分类背景、意义

### ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

(1) 分类一致性（信度）：

- 指考生在同一测验的两个平行复本上被一致分类的比例。

		英语六级A卷	
		≥425(通过)	<425(不通过)
英语六级B卷	通过	一致	不一致
	不通过	不一致	一致

# 基础知识

## ■ 分类背景、意义

### ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

### (2) 分类准确性（效度）：

- 将考生进行正确分类的比例，即基于观察分数分布和基于真分数分布的类别划分一致性程度。

		英语六级考试得分	
		≥425(通过)	<425(不通过)
真实英语六级水平	通过	一致	不一致
	不通过	不一致	一致

# 基础知识

## ■ 分类背景、意义

### ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算：

### ■ $P$ 指标

分类一致性：考生在同一测验的两个平行复本上被一致分类的比例。

		英语六级A卷	
		≥425(通过)	<425(不通过)
英语六级B卷	通过	$p_{11}$	$p_{10}$
	不通过	$p_{01}$	$p_{00}$

$$P = \sum_{j=1}^J P_{jj} = p_{00} + p_{11}$$

↑  
类别总数

↓  
两个平行复本都划分到第 $j$ 类的考生比例

# 基础知识

## ■ 分类背景、意义

### ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算：

### ■ $P$ 指标

分类准确性：可以表示为假阳性概率与假阴性概率。

		英语六级考试得分	
		$\geq 425$ (通过)	$< 425$ (不通过)
真实英语六级水平	通过	$p_{11}$	$p_{10}$
	不通过	$p_{01}$	$p_{00}$

假阳性( $p_{01}$ ): 卷面分比真实能力高

假阴性( $p_{10}$ ): 真实能力比卷面分高

# 基础知识

## ■ 分类背景、意义

## ☑ 分类一致性和准确性

### ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算：

#### ■ $P$ 指标

考生需要作答同一测验的两个平行复本。

➡ 考试时间有限

➡ 构建两个平行测验的费用较高

重测在实际中难以实施

➡ 如何通过单次测验获得分类一致性与准确性指标？



# 基础知识

## ■ 分类背景、意义

## ☑ 分类一致性和准确性

### ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算：

#### ■ 分半方法

计算所有分半情况下 $p$ 的均值，表示为 $\beta$ （与 $\alpha$ 系数类似）

→ 通过Spearman-Brown 公式获得对完整测验长度的估计值

(Marshall & Haertel, 1975)

➡ 仅根据经验得出，有待完整的理论证明

# 基础知识

## ■ 分类背景、意义

## ☑ 分类一致性和准确性

### ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算:

		英语六级A卷	
		≥425(通过)	<425(不通过)
?? 英语六级B卷??	通过	?	?
	不通过	?	?

Q: 只进行一次测验, 如何得到两份平行测验的数据?

A: 通过假设观察分数分布服从不同的模型/分布, 生成一个假设的、完整的、重复平行测验。

→ 分类一致性 $P$

# 基础知识

## ■ 分类背景、意义

### ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算：

		英语六级考试得分	
		≥425(通过)	<425(不通过)
真实英语六级水平	通过	?	?
	不通过	?	?

**A:** 用已有的数据的分类结果作为真实分类的估计值，与生成的观察分数分布分类结果的一致程度。

→ 分类准确性 $\gamma$

# 基础知识

## ■ 分类背景、意义

## ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算:

测量理论	模型/分布假设	参考文献
	二项分布	Huynh, 1976; Subkoviak, 1976; Hanson & Brennan, 1990; Livingston & Lewis, 1995
经典测量理论 (CTT)	多项分布	Lee, 2007; Lee, Brennan & Wan, 2009
	复合多项分布	Lee, 2007; Lee, Brennan & Wan, 2009
	双变量正态分布	Peng & Subkoviak, 1980; Woodruff & Sawyer, 1989; Breyer & Lewis, 1994
项目反应理论 (IRT)	Rasch 模型	Huynh, 1990
	多级计分模型	Wang, Kolen, & Harris, 2000

# 基础知识

## ■ 分类背景、意义

### ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

### (3) 分类一致性与分类准确性的计算:

#### ■ 二项分布 (0, 1计分)

假设考生的观察分数独立且服从二项分布，首先估计出单个考生在测验上被一致分类的概率，然后通过求平均获得测验分类一致性指标的估计值。

		英语六级A卷	
		≥425(通过)	<425(不通过)
英语六级B卷	通过	$p_{11}$	$p_{10}$
	不通过	$p_{01}$	$p_{00}$

$$P^{(\text{小明})}_C = P(\text{小明在A卷的得分} \geq 425, \text{小明在B卷的得分} \geq 425) \\ + P(\text{小明在A卷的得分} < 425, \text{小明在B卷的得分} < 425)$$

$$P_C = \sum_{i=1}^N P^{(\text{考生}i)}_C / \text{参加六级考试的学生总人数}$$

(Subkoviak, 1976)

# 基础知识

## ■ 分类背景、意义

## ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算:

### ■ 二项分布

$$P^{(\text{小明})}_{C=425} = P(\text{A卷} \geq 425, \text{B卷} \geq 425) + P(\text{A卷} < 425, \text{B卷} < 425)$$



两份平行测验独立 (作答互不影响)

$$P^{(\text{小明})}_{C=425} = P(\text{A卷} \geq 425) \times P(\text{B卷} \geq 425) + P(\text{A卷} < 425) \times P(\text{B卷} < 425)$$



均服从特定的二项分布

$$P^{(\text{小明})}_{C=425} = P(\text{A卷} \geq 425)^2 + P(\text{A卷} < 425)^2$$

$$P(X_i \geq C) = \sum_{X_i=C}^n \binom{n}{X_i} p^{X_i} (1-p)^{n-X_i}$$

(Subkoviak, 1976)

# 基础知识

## ■ 分类背景、意义

## ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算:

### ■ 二项分布

$$P(X_i \geq C) = \sum_{X_i=C}^n \binom{n}{X_i} p_i^{X_i} (1-p_i)^{n-X_i}$$



小明答对一道题的概率（正确比率真分数）



- 极大似然估计:  $\hat{p}_i = X_i/n$
- 线性回归:  $\hat{p}_i = \alpha_{21/x} \left(\frac{x_i}{n}\right) + (1 - \alpha_{21/x}) \left(\frac{M_x}{n}\right)$

$$\alpha_{21/x} = \frac{n}{n-1} \left[ 1 - \frac{M_x(n - M_x)}{nS_x^2} \right]$$

# 基础知识

## ■ 分类背景、意义

### ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

### (3) 分类一致性与分类准确性的计算:

#### ■ 二项分布

假设共有10名考生，试卷共有5题（0, 1计分），答对4题即通过。

	考生	得分/答对题数	正确比率真分数估计值
均值 = 2.30 方差 = 2.61	1	0	0.19
	2	4	0.66
	...	...	...
	10	5	0.77

$$\alpha_{21/x} = \frac{n}{n-1} \left[ 1 - \frac{M_x(n-M_x)}{nS_x^2} \right] = 0.58 ; \hat{p}_i = \alpha_{21/x} \left( \frac{x_i}{n} \right) + (1 - \alpha_{21/x}) \left( \frac{M_x}{n} \right) = 0.19$$

$$P(X_i \geq C) = \sum_{X_i=C}^n \binom{n}{X_i} p_i^{X_i} (1-p_i)^{n-X_i} = 0.0055 ; P^{(1)}_4 = 0.9890$$

$$P_C = \sum_{i=1}^N P^{(i)}_C / N = 0.75$$

(Subkoviak, 1976)



# 基础知识

## ■ 分类背景、意义

## ☑ 分类一致性和准确性

### ➤ 1.2 分类一致性与准确性

#### (3) 分类一致性与分类准确性的计算:

##### ■ Beta 二项分布 (二参数)

假设正确比率真分数分布服从  $beta$  分布, 以求联合概率的方式估计测验的分类一致性。

正确比率真分数  $\pi$  服从  $beta$  分布 (先验)



条件观察分数  $x$   $f(x) = \binom{n}{x} B(\alpha + x, n + \beta - x) / B(\alpha, \beta)$

$B(\alpha, \beta)$  为 Beta 函数  
 $\alpha$ : 均值;  $\beta$ : 方差  
(由 KR-21 和观察分数分布估计得到)

##### ■ Beta 二项分布 (四参数)

2PB 中  $\pi$  取值范围为  $[0, 1]$ , 4PB 中设定取值范围为  $[a, b]$

(Huynh, 1976; Hanson & Brennan, 1990)

# 基础知识

## ■ 分类背景、意义

### ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算:

■ 多项分布 (多级计分; 共有  $n$  道题)

假设有  $k$  个类别, 每个类别在一道题目上的得分分别为  $c_1, c_2, \dots, c_k$

$c_1, c_2, \dots, c_k$



$\pi_1, \pi_2, \dots, \pi_k$  (正确比率真分数)



1分  $p_i$



0分  $1 - p_i$



$\pi_k$

# 基础知识

## ■ 分类背景、意义

## ☑ 分类一致性和准确性

### ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算:

■ 多项分布 (多级计分; 共有  $n$  道题)

得分为  $c_1$  的题目数为  $x_1$ , 得分为  $c_2$  的题目数为  $x_2$ , 以此类推

$$x_1 + x_2 + \dots + x_k = n$$



服从多项分布, 则  $n$  道题中  $x_1, x_2, \dots, x_k$  出现的概率为

$$\Pr(X_1 = x_1, X_2 = x_2, \dots, X_K = x_K | \vec{\pi}) = \frac{n!}{x_1! x_2! \dots x_K!} \pi_1^{x_1} \pi_2^{x_2} \dots \pi_K^{x_K}$$



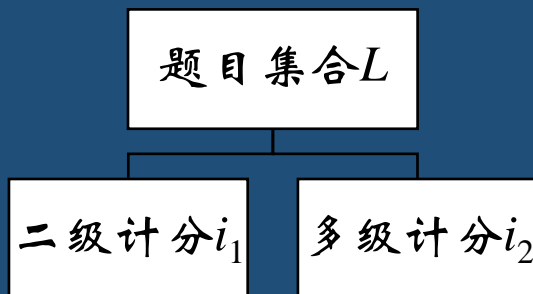
测验总分为  $Y$  的概率

$$\Pr(Y = y | \vec{\pi}) = \sum_{c_1 x_1 + c_2 x_2 + \dots + c_K x_K = y} \Pr(X_1 = x_1, X_2 = x_2, \dots, X_K = x_K | \vec{\pi})$$

# 基础知识

## ■ 分类背景、意义

### ☑ 分类一致性和准确性



## ➤ 1.2 分类一致性与准确性

### (3) 分类一致性与分类准确性的计算:

#### ■ 复合多项分布（混合题型；共有 $n$ 道题）

让计分规则相同的题目属于同一个集合 $i$ ，共有 $L$ 个集合

$\pi_{ij}$ 为第 $i$ 个集合中考生得分为 $c_{ij}$ 的正确比率真分数



每个题目集合均服从多项分布，则集合 $i$ 的总分为 $Y$ 的概率为

$$\Pr(Y_1 = y_1, \dots, Y_L = y_L | \vec{\pi}_1, \dots, \vec{\pi}_L) = \prod_{i=1}^L \Pr(Y_i = y_i | \vec{\pi}_i)$$



测验总分为 $Z$ 的概率( $Z = \sum_{i=1}^L w_i Y_i$ )

$$\Pr(Z = z | \vec{\pi}_1, \dots, \vec{\pi}_L) = \sum_{y_1, \dots, y_L: \sum w_i y_i = z} \Pr(Y_1 = y_1, \dots, Y_L = y_L | \vec{\pi}_1, \dots, \vec{\pi}_L)$$

# 基础知识

## ■ 分类背景、意义

## ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算:

### ■ 项目反应理论

计算能力 $\theta$ 条件下的作答反应向量 $U_n$



通过复合二项分布/多项分布/...计算条件观察分数分布



$$P(U_1, U_2, \dots, U_n | \theta) = \prod_{i=1}^n P(U_i | \theta)$$

$$P(X | \theta) = \sum_{U_1, \dots, U_n: \sum w_i U_i = X} P(U_1, U_2, \dots, U_n | \theta)$$

分类一致性与准确性指标

# 基础知识

## ■ 分类背景、意义

## ☑ 分类一致性和准确性

## ➤ 1.2 分类一致性与准确性

(3) 分类一致性与分类准确性的计算：

