详细解答

STEP I 1988，第 15 题

概率与统计 • 二项分布的假设检验

STEP I 1988，第 15 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/88-s1-q15

关键思路

把“超过一半的得分来自犯规”化为犯规次数与入球数之间的一个简单不等式（inequality）。以入球数作为条件（conditioning）之后，余下的工作只是数一数有多少个等可能的犯规次数符合要求。至于第二部分，要问的是：按照 B 分布，在 300 次试验中预期会出现多少次九次犯规的情况？假如甚么都没有改变，观察到的次数又有多出人意料？

解答

概率

就某场比赛中的某一支球队而言，设 G 为该队的入球数，F 为对方向该队犯规的次数。那么 G 与 F 独立（independent），且

$$\mathrm{P}(G=0),\ \ldots,\ \mathrm{P}(G=4) \ = \ \tfrac{1}{9},\ \tfrac{2}{9},\ \tfrac{3}{9},\ \tfrac{2}{9},\ \tfrac{1}{9}, \qquad \mathrm{P}(F=f) = \tfrac{1}{10} \quad (f=0,1,\ldots,9).$$

该队从入球得 $2G$ 分，从犯规得 F 分，合共 $2G+F$ 分。该队超过一半的得分来自犯规，当且仅当

$$F > \tfrac{1}{2}(2G+F) \iff F > 2G.$$

（假如该队一分也没有得到，则 $F = 0 = 2G$，这种情况不应计算在内，而严格不等式正好把它排除了。）

给定 $G = g$，符合要求的 F 值为 $2g+1,\ldots,9$，即十个等可能的值当中的 $9-2g$ 个。因此，利用 F 与 G 的独立性，

$$\mathrm{P}(F > 2G \mid G = g) = \frac{9-2g}{10}, \qquad g = 0,1,2,3,4,$$

得出 $\frac{9}{10}, \frac{7}{10}, \frac{5}{10}, \frac{3}{10}, \frac{1}{10}$。根据全概率公式（law of total probability），

$$\mathrm{P}(F > 2G) = \frac{1\cdot 9 + 2\cdot 7 + 3\cdot 5 + 2\cdot 3 + 1\cdot 1}{9\times 10} = \frac{45}{90} = \frac{1}{2}.$$

答案如此简洁，是有原因的：入球数的分布关于 2 对称，而 $(9-2g)/10$ 是 g 的线性函数，所以它的平均值等于它在 $g=2$ 的值，即 $\frac{1}{2}$。

犯规次数的分布有没有改变?

一支球队的犯规次数，就是对手被犯规的次数，所以在旧规则下，每支球队在一场比赛中犯规 9 次的概率为 $\frac{1}{10}$。设 X 为 300 个球队场次当中，球队犯规 9 次的场次数目。我们检验

$$H_0:\ p = \tfrac{1}{10}\ \text{（没有改变）} \qquad \text{对} \qquad H_1:\ p < \tfrac{1}{10},$$

其中 p 是一支球队在一场比赛中犯规 9 次的概率。备择假设（alternative hypothesis）是单侧（one-sided）的，因为修改规则的目的，是令犯规次数偏高的情况变得*较不*可能出现。

> **想一想** 这里采用二项分布（binomial）模型是否合理?

我们需要这 300 个球队场次互相独立，而且每个场次的概率 p 都相同。题目已说明同一场比赛之内的独立性，而假设不同场次的比赛之间也互相独立，亦属合理。因此在 H_0 之下，$X \sim \mathrm{B}(300, \frac{1}{10})$，且

$$\mathrm{E}(X) = 30, \qquad \mathrm{Var}(X) = 300\times\tfrac{1}{10}\times\tfrac{9}{10} = 27, \qquad \text{标准差}\sqrt{27}\approx 5.20.$$

我们观察到 $X=12$，比期望值 30 少了 18，相当于约 3.5 个标准差（standard deviation）。采用正态近似（normal approximation）并加上连续性校正（continuity correction），

$$\mathrm{P}(X\le 12) \approx \Phi\left(\frac{12.5-30}{\sqrt{27}}\right) = \Phi(-3.37) \approx 0.0004.$$

精确的二项概率还要更小，约为 0.0001。无论采用哪一个数值，假如甚么都没有改变，这样极端的结果在一千个球季之中也出现不到一次。它远远超出常用的 5% 甚至 1% 显著性水平（significance level），所以我们拒绝 H_0。这*确实*是有力的证据，显示犯规次数的分布已经改变，而且是朝着所期望的方向改变：犯规九次的场次变得较为罕见。

有两点保留值得说明。这个检验只显示 P(9 次犯规) 下降了，并没有说明分布的其余部分如何改变（尽管那是最自然的解释）。此外，统计学也无法证明这个下降是由修改规则*造成*的，因为两个球季之间可能还有其他事情改变了。

验算

▶ 入球数的概率 $\frac{1}{9}(1+2+3+2+1)$ 之和为 1，理应如此。

- 以精确分数逐一列举全部 5×10 个数对 (g, f)，得出 $\mathrm{P}(F > 2G) = \frac{1}{2}$；而模拟 200,000 场比赛的结果为 0.4997。
- 直接把二项概率加起来，对于 $X \sim \mathrm{B}(300, 0.1)$ 得出 $\mathrm{P}(X \leq 12) = 9.2 \times 10^{-5}$。正态近似（$3.8 \times 10^{-4}$）与平均值为 30 的泊松近似（Poisson approximation）（1.7×10^{-4}）都一致显示这个概率极小。

常见错误

- 把 F 与 G 比较，而不是与 $2G$ 比较：一个入球值两分。
- 计算 $F \geq 2G$ 而不是 $F > 2G$。“超过一半”是严格的不等关系，而一分未得的球队，并没有任何得分来自犯规。
- 以 150 场比赛作为样本大小，而不是 300 个球队场次；或者在计算之前忘记写明零假设（null hypothesis）和二项分布模型。

STEP 之外

第二部分是对二项比例所作的假设检验（hypothesis test），而我们求得的尾部概率就是 p 值（p-value）。到了大学，你会学到似然比（likelihood ratio）检验和卡方检验（chi-squared test），它们会把观察所得的整个犯规次数分布与旧分布作比较，而不只是比较其中一个值。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率论入门）把概率方面的内容（条件概率、二项分布与正态近似）讲解得十分清楚。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。