University of Waterloo W. H. Cherry

PARADOXES: Efron's Dice

Experience shows that probability is a topic many people find difficult, in that mistakes are easier to make than in (some) other subject areas. This is one of six Statistical Highlights (#46 to #51) which discuss probabilistic subtleties and mistakes (which sometimes lead to so-called 'paradoxes'), with a view to helping the reader recognize and deal correctly with such matters. [These and related statistical issues are also discussed in Figure 7.12 of the STAT 220 Course Materials.]

A conventional die is a cube (with rounded corners and edges) whose six faces are made distinguishable by labels like 1 to 6 or these numbers of dots. For the four Efron dice, labels are often more restricted – in Section 1 below, the dice have at most *two* different labels chosen from *seven*, here 0, 1, 2, 3, 4, 5, 6. Two more numerous label sets, with fewer die restrictions, are mentioned in SOURCE overleaf on page HL48.2.

1. The Context

Four (six-sided) dice, here designasted **A**, **B**, **C** and **D**, have outcomes denoted 0 to 6 and (carefully chosen) associated probabilities, most of which are zero. As is common in situations like this, we represent these abstractions by the four (univariate) probability functions of the random variables A, B, C and D as shown in Tables (1), (2), (3) and (4) at the right. The Question of interest is which die is 'better' in four pairwise comparisons **A** *vs* **B**, **B** *vs* **C**, **C** *vs* **D** and **D** *vs* **A**? – that is, which die in each pairwise comparison is more likely to show the higher outcome?

a	0	1	2	3	4	5	6
$f(a)$	0	0	0	1	0	0	0

-----(1)

b	0	1	2	3	4	5	6
$f(b)$	0	0	⅔	0	0	0	⅓

-----(2)

c	0	1	2	3	4	5	6
$f(c)$	0	½	0	0	0	½	0

-----(3)

d	0	1	2	3	4	5	6
$f(d)$	⅓	0	0	0	⅔	0	0

-----(4)

We can answer this question using the four *joint* (bivariate) p.f.s of A and B, B and C, C and D, D and A; these are shown at the right in Tables (5), (6), (7) and (8); under the assumptuion of probabilistic independence, their 49 values are each the *product* of the relevant two values from each pair of relevant individual p.f.s – for example, ⅔ when $A=3$ and $B=2$ is 1×⅔ and is the probability of the event $A=3 \cap B=2$. [$\cap$ denotes 'intersection'.]

b \ a	0	1	2	3	4	5	6	
0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
2	0	0	0	⅔	0	0	0	⅔
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	⅓	0	0	0	⅓
	0	0	0	1	0	0	0	

-----(5)

c \ b	0	1	2	3	4	5	6	
0	0	0	0	0	0	0	0	0
1	0	0	⅓	0	0	0	⅙	½
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	⅓	0	0	0	⅙	½
6	0	0	0	0	0	0	0	0
	0	0	⅔	0	0	0	⅓	

-----(6)

d \ c	0	1	2	3	4	5	6	
0	0	⅙	0	0	0	⅙	0	⅓
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	⅓	0	0	0	⅓	0	⅔
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
	0	½	0	0	0	½	0	

-----(7)

a \ d	0	1	2	3	4	5	6	
0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	⅓	0	0	0	⅔	0	0	1
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
	⅓	0	0	0	⅔	0	0	

-----(8)

- The probability in Table (5) of ⅔ of the event $A=3 \cap B=2$ (in which **A**s outcome of 3 is higher than **B**s 2) is *higher* than the ⅓ for the event $A=3 \cap B=6$ (in which **A**s outcome of 3 is lower than **B**s 6); thus, in the **A** *vs* **B** pairwise comparison, **A** is 'better' than **B**.
- The sum of the three probabilities in Table (6) of ⅔ of the events $B=2 \cap C=1$, $B=6 \cap C=1$ and $B=6 \cap C=5$ (in each of which **B**s outcome is higher than **C**s) is *higher* than the ⅓ for the event $B=2 \cap C=5$ (in which **B**s outcome of 2 is lower than **C**s 5); thus, in the **B** *vs* **C** pairwise comparison, **B** is 'better' than **C**.
- The sum of the three probabilities in Table (7) of ⅔ of the events $C=1 \cap D=0$, $C=5 \cap D=0$ and $C=5 \cap D=4$ (in each of which **C**s outcome is higher than **D**s) is *higher* than the ⅓ for the event $C=1 \cap D=4$ (in which **C**s outcome of 1 is lower than **D**s 4); thus, in the **C** *vs* **D** pairwise comparison, **C** is 'better' than **D**.
- The probability in Table (8) of ⅔ of the event $D=4 \cap A=3$, (in which **D**s outcome of 4 is higher than **A**s 3) is *higher* than the ⅓ for the event $D=0 \cap A=3$ (in which **D**s outcome of 0 is lower than **A**s 3); thus, in the **D** *vs* **A** pairwise comparison, **D** is 'better' than **A**.

This is the 'paradox' generated by Efron's 'dice': **A** is 'better' than **B** is 'better' than **C** is 'better' than **D** *but* **D** is 'better' than **A**.

2. How can we account for the 'paradox'?

*Un*like the Apple Pie Paradox discussed in Statistical Highlight #47, where comparison is among *different* (trivariate and bivariate) distributions, the apparently anomalous ranking among the four Efron dice arises from bivariate distributions which are generated by the *same* process. however, the *common* theme is the potential difficulty of ranking multivariate information.

(*continued overleaf*)

- The probabilites that yield the die rankings are generated from *multiple* components and so it is *un*surprisung if (carefully chosen) sets of these components can be made to yield a seemingly anomalous ranking. Most notable among these choices:
  - there are seven possible die labels, not the usual six;
  - the resulting seven possible die outcomes allow the (few) non-zero probabilities to be assigned as a single 'central' 3 with a probability of 1 to die **A** with the remaining six outcomes in pairs arranged systematically around the central '3' to **B**, **C** and **D**;
  - the outcomes with *non*-zero probabilities are *distinct* among the four dice (and cover all seven outcomes) [but see SOURCE below and Table HL48.2].
- As for the Apple Pie Paradox of Statistical Highlight #47, the probability calculations overleaf on page HL48.1 involve a *modelling* assumption of probabilistic independence, a mathematical idealization whose real-world implications are routinely troublesome.
- The respective *means* of the four univariate p.f.s (1), (2), (3) and (4), at the upper right overleaf on page HL48.1, are 3, 3⅓, 3 and 2⅔, which yield a 'ranking' of **B** 'better' than **A** which is the same' as **C** which is 'better' than **D**; this differs from the previous (bivariate) 'probability' ranking and so (like Statistical Highlight #47) reminds us that, when probability distributions are involved, different approaches to a Question may yield different Answers.

Despite these caveats, the Efron 'dice' are notable, all the more so because of their achievement of the *same* odds of 2:1 for the four comparisons.

- ∗ Anomalous ranking is *not* confined to a *technical* context like that of this Highlight #48; it may also arise in more familiar settings; for instance:
  - Is the world's best quarterback or the world's best shot puter the better athlete?
  - Who is the greatest tennis (or other sport) player of all time?

3. Practical Considerations

The *mathematical* appeal of Efron's dice may be dimiinished from a practical perspective by the 'unnatural' characteristics of the dice; in particular:

- ∗ seven labels for a set of what are normally thought of as *six*-sided objects;
- ∗ restricted outcomes (arising from zero probabilities), most notably for **A**, that seems inconsistent with the underlying *concept* of a die.

Although the second characteristic is markedly reduced in another (larger) label set that also generate anomalous ranking of the four dice (see Table HL48.2 below), an individual unsympathetic to even the *idea* of 'paradoxes' might use considerations like these to argue their case.

SOURCE: The four probability functions (1), (2), (3) and (4) at the upper right overleaf on page HL48.1 are based on diagrams given by Wolfram MathWorld, identified in a Google search for *Efron dice* in December, 2021.

This source gives two other sets of dice diagrams and two references.

- The first set has thirteen die outcomes of 0 to 12, with the respective dice having 2, 2, 4 and 5 outomes with non-zero probabilities and no overlap across the four dice; this set preserves the 2:1 odds ratio for all four comparisons.
- The second set has the same thirteen die outcomes but the non-zero probabilities are more numerous (3, 5, 5, 6) for each die and so (obviously) overlap among the (odd) outcomes; this overlap allows *ties* to occur. The odds ratio of 11:6 *ex*cludes ties (which involve only *odd* outcomes).

The p.f.s for *these* two dice sets are summarized in Tables HL48.1 and 48.2 below.

Table HL48.1 (13 labels, 2:1 odds)

Label	0	1	2	3	4	5	6	7	8	9	10	11	12
$f(a)$	0	0	0	0	0	⅓	⅔	0	0	0	0	0	0
$f(b)$	0	0	0	0	⅔	0	0	0	0	0	0	0	⅓
$f(c)$	⅙	⅙	0	0	0	0	0	⅙	½	0	0	0	0
$f(d)$	0	0	⅙	⅓	0	0	0	0	0	⅙	⅙	⅙	0

Table HL48.2 (13 labels, 11:6 odds)

Label	0	1	2	3	4	5	6	7	8	9	10	11	12
$f(a)$	0	0	0	0	0	⅓	⅓	⅓	0	0	0	0	0
$f(b)$	⅙	⅙	0	0	0	0	0	⅙	⅓	⅙	0	0	0
$f(c)$	0	0	0	⅙	⅓	⅙	0	0	0	0	0	⅙	⅙
$f(d)$	0	⅙	⅙	⅙	0	0	0	0	0	⅙	⅙	⅙	0

The two references are:

- Gardner, M.: *Mathematical Games: The Paradox of the Nontransitive Dice and the Elusive Principle of Indifference.* Scientific American **223**, 110-114, December, 1970.
- Honsberger, R.: *Some Surprises in Probability*; Chapter 5 in: *Mathematical Plums* (Ed. R. Honsberger), Washington, DC: Math. Assoc. Amer., pages 94-97, 1979.

NOTE: If we were to change Sttatistical Highlight #47 to *four* flavours of pie and *seven* values of 0 to 6 for the pie enjoyment index, the probabilistic structure of the 'paradox' in *this* Highlight #48 would then be involved, with (hopefully) a plausibly modified context to justify ranking only pairwise comparisons of pie enjoyment.