A PRECISE PROBABILITY RELATED TO SIMPSON'S PARADOX

ALEX RICE

Abstract. Suppose two friends are about to participate in a two-day contest in which they repeatedly attempt a task with a clear success/failure outcome (such as shooting free throws on a basketball court). We have no specific prior knowledge of the participants' skills, how the change in day will impact their success, or how many attempts they will take each day, so we suppose that each participant's success rate for day 1, success rate for day 2, and proportion of attempts that take place on day 1 are all chosen uniformly at random between 0 and 1. What is the probability that the same person has a higher success rate each of the two individual days, but the other person has a higher success rate for the two-day period? This can be thought of as a prior probability of a simple case of *Simpson's paradox*, and we show that this probability is $(\pi^2 - 9)/36 = .0241556778\ldots$.

1. Introduction

The following is a problem composed by the author, versions of which have been included in a variety of math competitions at the middle school, high school, and collegiate levels:

> Robbie and Julia compete in a two-day coin-flipping contest. Each day, each person flips between 1 and 100 coins, inclusive. Each day, Robbie flips a higher proportion of heads than Julia. For the two-day contest, Robbie flips $R\%$ heads, while Julia flips $J\%$ heads. What is the least possible value of $R - J$, rounded to the nearest integer?

For the sake of the intrigued reader, we postpone the precise answer to the very end of the note, but the key is to divorce oneself from the intuition that the answer must be nonnegative. The fact that R can be less than J is a simple example of a broader statistical phenomenon known as *Simpson's paradox*, in which a statistical trend that exists in multiple sets of data is reversed when the sets are combined. The effect was so-named by Blyth [1] in 1972 owing to its description by Simpson [7] in 1951, but analogous effects had been described earlier by Pearson, Lee, and Bramley-Moore [5] and Yule [8].

But even if we accept the possibility that R can be less than J, how *likely* is such a scenario? The coin-flip example could be constrained by the assumption that the coin is fair, or at least the same for each player and the same on each day. Instead, we eschew such prior knowledge, and consider a two-day contest in which two players repeatedly attempt a task with a clear success/failure outcome, and we suppose that each participant's success rate for day 1, success rate for day 2, and proportion of attempts that take place on day 1 are all chosen uniformly[1] at random between 0 and 1. Under these assumptions, what is the probability that the same person has a higher success rate each of the two days, but the other person has a higher success rate for the two-day period? A geometric interpretation of the question is shown in Figure 1 below.

Similar questions have been considered previously, most notably by Pavlides and Perlman [4] who verified a solution of Hadjicostas [2] to the following question: if outcomes are uniformly distributed amongst the eight possibilities determined by the success or failure of three events A, B, C, what is the probability that A and B are positively correlated conditioned on C, and positively correlated conditioned on $\overline{C}$ (where the bar denotes the complement), but negatively correlated overall (or the analogous scenario with all correlations reversed)? In this setting, the precise probability is $1/60$. The likelihood of a similar formulation of Simpson's paradox in quantum mechanics is further discussed in Section 5 of [6].

[1]The motivating examples of free throws and coin flips suggest a binomial distribution is more appropriate than uniform. However, rather than thinking of a player's "skill level" (success probability on each independent trial) as fixed, we seek a prior probability by averaging over all possible skill levels, and considering all possible sample sizes. In particular, we leave it as an exercise that if $f(\delta, N, \epsilon)$ denotes the probability of at most ϵN successes from N independent trials with probability δ, then $\lim_{N\to\infty} \int_0^1 f(\delta, N, \epsilon) d\delta = \epsilon$ for every $\epsilon \in [0, 1]$.

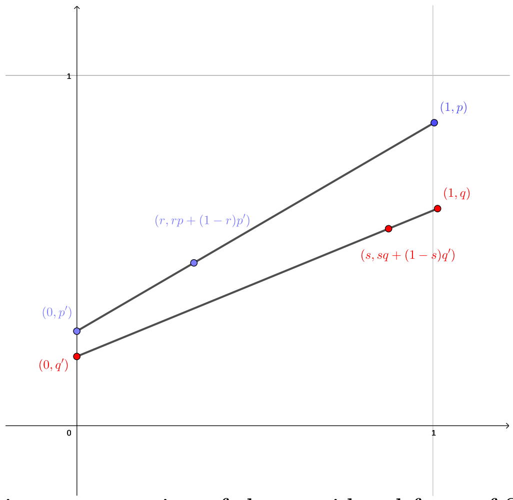

FIGURE 1. A geometric representation of the considered form of Simpson's paradox, with $x=0$ corresponding to day 2 and $x=1$ corresponding to day 1. While both blue endpoints are higher than the respective red endpoints, the intermediate red point, whose height is given by a convex combination of the two red endpoint heights, is higher than the intermediate blue point. The central question here is the likelihood of a configuration like this (or with the colors reversed) if the four colored endpoint heights, followed by the two positions along the two resulting line segments, are all chosen uniformly at random.

2. Main Computation

To begin with our formulation of the reversal phenomenon, suppose $\mathbf{v}=(p,p',q,q',r,s)\in[0,1]^6$. We say that $\mathbf{v}$ is a *Simpson reversal* if $p>q$, $p'>q'$, and $rp+(1-r)p'<sq+(1-s)q'$, or if the same system holds with all inequalities reversed. Our central question is to determine the probability of a Simpson reversal, provided all six of the defined probabilities are chosen uniformly at random between 0 and 1. For $n\in\mathbb{N}$, we let μ_n denote n-dimensional Lebesgue measure (since the sets we are working with are relatively nice, the uninitiated reader can think of this as *volume* in the appropriate number of dimensions). Our main result, the computation of the desired probability, is as follows.

Theorem 1. $\mu_6\left(\left\{\mathbf{v}\in[0,1]^6 : \mathbf{v} \text{ is a Simpson reversal}\right\}\right)=\dfrac{\pi^2-9}{36}=0.0241556778\ldots$

The previous work that aligns most closely with our computation is that of Jones and Wilson [3], with the novelty coming from our averaging over all possible choices of (p,p',q,q') to obtain a single prior probability.

Proof. Let $S=\left\{\mathbf{v}\in[0,1]^6 : \mathbf{v} \text{ is a Simpson reversal}\right\}$. We first note that if $(p,p',q,q',r,s)\in S$ with $p>q$ and $p>p'$, then we must have $q>p'$, otherwise any convex combination of p and p' would exceed any convex combination of q and q'. Further, the three maps sending (p,p',q,q',r,s) to $(p',p,q',q,1-r,1-s)$, (q,q',p,p',s,r) and $(q',q,p',p,1-s,1-r)$, respectively, are measure-preserving bijections between

$$S_1=\{(p,p',q,q',r,s)\in S : q'<p'<q<p\}$$

and the analogous subsets of S defined by switching the roles of the "players" and/or the "days". In particular, $\mu_6(S)=4\mu_6(S_1)$. Further the map $(p,p',q,q',r,s)\mapsto(1-q',1-q,1-p',1-p,1-s,1-r)$ is a measure-preserving bijection between

$$T=\{(p,p',q,q',r,s)\in S_1 : p-p'>q-q'\}$$

and the analogous subset of S_1 with $q-q'>p-p'$. Therefore, we have $\mu_6(S_1)=2\mu_6(T)$, hence

$$\mu_6(S)=8\mu_6(T). \tag{1}$$

Making the further measure-preserving changes of variables $x = p - p'$ and $y = q - q'$, $\alpha = 1/2 - r$, and $\beta = s - 1/2$ we have that T has the same measure as

$$T' = \{(p, x, q, y, \alpha, \beta) \in \mathbb{R}^6 : 0 < q - y < p - x < q < p < 1,\ x > y,\\ -1/2 \leq \alpha, \beta \leq 1/2,\ 2x\alpha + 2y\beta > 2p - x - 2q + y\},$$

where the final inequality is a rearrangement of the condition

$$(1/2 - \alpha)p + (1/2 + \alpha)(p - x) < (1/2 + \beta)q + (1/2 - \beta)(q - y),$$

or equivalently

$$x\alpha + y\beta > \left(\frac{p + p'}{2}\right) - \left(\frac{q + q'}{2}\right),$$

where the right-hand side is the vertical distance between the blue and red lines in Figure 1 at $x = 1/2$. Crucially, for each fixed element of

$$\tilde{T} = \left\{(p, x, q, y) \in \mathbb{R}^4 : 0 < q - y < p - x < q < p < 1,\ x > y\right\},$$

we have

$$\mu_2\left(\left\{(\alpha, \beta) \in [-1/2, 1/2]^2 : 2x\alpha + 2y\beta > 2p - x - 2q + y\right\}\right) = \frac{1}{4xy} \cdot \mu_2\left(T_{(p,x,q,y)}\right),$$

where $T_{(p,x,q,y)} = \{(u, v) \in [-x, x] \times [-y, y] : u + v > 2p - x - 2q + y\}$, as shown in the figure below.

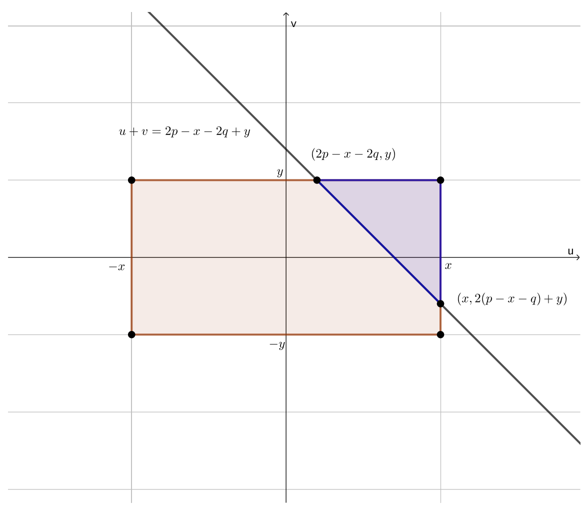

FIGURE 2. For fixed $(p, x, q, y) \in \tilde{T}$, the blue region is $T_{(p,x,q,y)}$, which has area $2(q-p+x)^2$. This figure is effectively equivalent to Figure 2 in [3]. Dividing $2(q-p+x)^2$ by $4xy$, the area of the larger rectangle, gives a probability equivalent to the conclusion of the proposition on page 296 of that paper.

Combining this with (1), we have

$$\mu_6(S) = 8 \iiiint\limits_{\tilde{T}} \frac{1}{4xy} \cdot \mu_2\left(T_{(p,x,q,y)}\right) dq dp dy dx = 4 \iiiint\limits_{\tilde{T}} \frac{(q - p + x)^2}{xy} dq dp dy dx. \tag{2}$$

We partition $\tilde{T}$ into three pieces: $\tilde{T}_1 = \{(p, x, q, y) \in \tilde{T} : p > x + y\}$, $\tilde{T}_2 = \{(p, x, q, y) \in \tilde{T} : p < x + y < 1\}$, and $\tilde{T}_3 = \{(p, x, q, y) \in \tilde{T} : x + y > 1\}$. Let $R = \{(x, y) \in \mathbb{R}^2 :\ x > y > 0,\ x + y < 1\}$. Then, we have

$$\iiiint\limits_{\tilde{T}_1} \frac{(q-p+x)^2}{xy} dqdpdydx = \iint\limits_R \int_{x+y}^1 \int_{p-x}^{p-x+y} \frac{(q-p+x)^2}{xy} dqdpdydx$$
$$= \frac{1}{3} \iint\limits_R \frac{1-x}{x} y^2 - \frac{y^3}{x} dydx.$$

In order to save the dx integration for last, we decompose the integration over R as

$$\iint\limits_R (\cdot) dydx = \int_0^{1/2} \int_0^x (\cdot) dydx + \int_{1/2}^1 \int_0^{1-x} (\cdot) dydx,$$

and we encourage the reader to integrate away the first three variables and verify that

$$\iiiint\limits_{\tilde{T}_1} \frac{(q-p+x)^2}{xy} dqdpdydx = \frac{1}{36}\left(\frac{11}{192} + I_1\right), \quad \text{where} \quad I_1 = \int_{1/2}^1 \frac{(1-x)^4}{x} dx, \tag{3}$$

$$\iiiint\limits_{\tilde{T}_2} \frac{(q-p+x)^2}{xy} dqdpdydx = \iint\limits_R \int_x^{x+y} \int_y^{p-x+y} \frac{(q-p+x)^2}{xy} dqdpdydx = \frac{1}{4} \iint\limits_R \frac{y^3}{x} dydx = \frac{1}{1024} + \frac{1}{16} I_1, \tag{4}$$

and

$$\iiiint\limits_{\tilde{T}_3} \frac{(q-p+x)^2}{xy} dqdpdydx = \int_{1/2}^1 \int_{1-x}^x \int_x^1 \int_y^{p-x+y} \frac{(q-p+x)^2}{xy} dqdpdydx = \frac{1}{12} I_2, \tag{5}$$

where

$$I_2 = \int_{1/2}^1 \frac{(x-1)^4 \log\left(\frac{x}{1-x}\right) + 4(x-1)^3(2x-1) + 3(x-1)^2(2x-1)}{x} dx$$

is a convergent improper integral. Having reduced matters to single integrals, we find (with computer assistance, verified by hand) that

$$I_1 = \log 2 - \frac{131}{192}, \ I_2 = \frac{\pi^2}{12} - \frac{13 \log 2}{12} - \frac{1}{24}. \tag{6}$$

Combining (2)-(6), we have $\mu_6(S) = 4\left(\frac{11}{36 \cdot 192} + \frac{1}{1024} + \left(\frac{1}{36} + \frac{1}{16}\right) I_1 + \frac{1}{12} I_2\right) = \frac{\pi^2 - 9}{36}$, as claimed. □

3. Concluding Remarks

For the patient reader, we now return with the answer to the problem posed at the very beginning of the introduction. Up to the ordering of the days, the most extreme scenario is as follows. Day 1: Robbie flips 1/1 head, while Julia flips 99/100 heads. Day 2: Robbie flips 1/100 heads, while Julia flips 0/1 heads. By percentage, Robbie wins both days, but for the two-day period, $R = 100 * (2/101) \approx 1.98$, while $J = 100 * (99/101) \approx 98.02$. This gives a minimum value for $R - J$, rounded to the nearest integer, of -96.

Regarding future directions, the calculation provided in this note could, with sufficient patience and care, be generalized with respect to the number of days or number of players. It would also be interesting to provide a more intuitive, perhaps geometric explanation for the final answer $\frac{\pi^2-9}{36} = \frac{\pi+3}{6} \cdot \frac{\pi-3}{6}$.

References

[1] C. Blyth, *On Simpson's paradox and the sure-thing principle*, Jour. of the Amer. Stat. Assoc. **67** (1972), no. 338, 364-366.

[2] P. Hadjicostas, *The asymptotic proportion of subdivisions of a* 2×2 *table that result in Simpson's paradox*, Combinatorics, Probability, and Computing **7** (1999), 387-396.

[3] M. A. Jones, J. Wilson, *A two-dimensional perspective on Simpson's paradox and its likelihood*, The College Mathematics Journal **50** (2019), no. 4, 295-297.

[4] M. Pavlides, M. Perlman, *How likely is Simpson's paradox?*, The American Statistician **63** (2009), no. 3, 226-233.

[5] K. Pearson, A. Lee, L. Bramley-Moore, *Genetic (reproductive) selection: inheritance of fertility in man, and of fecundity in thoroughbred racehorses*, Philosophical Transactions of the Royal Society A, **192** (1899), 257-330.

[6] A. Selvitella, *The ubiquity of Simpson's paradox*, Jour. of Statistical Distributions and Applications **4** (2017), Article 2.

[7] E. Simpson, *The interpretation of interaction in contingency tables*, Jour. of the Royal Stat. Soc. B **13** (1951), 238-241.

[8] G. U. Yule, *Notes on the theory of association of attributes in statistics*, Biometrika **2** (1903), no. 2, 121-134.

Department of Mathematics, Millsaps College, Jackson, MS 39210

Email address: `riceaj@millsaps.edu`