The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

How Likely Is Simpson's Paradox?

Marios G. Pavlides & Michael D. Perlman

How Likely Is Simpson's Paradox?

Marios G. PAVLIDES and Michael D. PERLMAN

What proportion of all $2 \times 2 \times 2$ contingency tables exhibit Simpson's Paradox? An exact answer is obtained for large sample sizes and extended to $2 \times 2 \times \ell$ tables by Monte Carlo approximation. Conditional probabilities of the occurrence of Simpson's Paradox are also derived. If the observed cell proportions satisfy a Simpson reversal, the posterior probability that the population parameters satisfy the same reversal is obtained. This Bayesian analysis is applied to the well-known Simpson reversal of the 1995–1997 batting averages of Derek Jeter and David Justice.

KEY WORDS: Bayes factor; Bayes test; Dirichlet distribution; Multinomial distribution; Simpson's Paradox; Simpson reversal.

1. INTRODUCTION

The celebrated Simpson's Paradox occurs when two random events A and B are conditionally positively (negatively) correlated given a third event C, and also given its complement $\bar{C}$, but are unconditionally negatively (positively) correlated. That is, either

$$P(A \cap B \mid C) \geq P(A \mid C)P(B \mid C), \tag{1.1}$$

$$P(A \cap B \mid \bar{C}) \geq P(A \mid \bar{C})P(B \mid \bar{C}), \tag{1.2}$$

$$P(A \cap B) \leq P(A)P(B), \tag{1.3}$$

with at least one inequality strict, or the three inequalities are replaced by their opposites. We shall refer to (1.1)–(1.3) as a *positive Simpson reversal*, and to their opposite as a *negative Simpson reversal*.

As an example of this phenomenon, Julious and Mullee (1994) report a study of two treatments $(A, \bar{A})$ for kidney stones in which the first treatment was *more* effective (B) for large stones (C) and small stones $(\bar{C})$ but *less* effective $(\bar{B})$ when the data were aggregated over both classes of stones.

Although a Simpson reversal at first appears paradoxical, many examples occur throughout the literature, especially in the social and biomedical sciences. See, for example, Bickel, Hammel, and O'Connell (1975), Blyth (1972), and the references available at *http://en.wikipedia.org/wiki/Simpson's_paradox*.

Marios G. Pavlides is Lecturer of Mathematics and Statistics, Frederick University Cyprus, Box 24729, 1303 Nicosia, Cyprus (E-mail: *m.pavlides@frederick.ac.cy*). Michael D. Perlman is Professor, Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195 (E-mail: *michael@stat.washington.edu*). We warmly thank Fred Bookstein, Peter Hoff, Thomas Richardson, Jon Wakefield, Jon Wellner, the Editor, Associate Editor, and an anonymous referee for many valuable comments. We are especially grateful to Petros Hadjicostas for generously contributing the results in Section 5.

Thus one may ask: *Just how likely is a Simpson reversal?* We shall address several versions of this question.

First, if $[p_{ijk} \mid i, j, k = 1, 2]$ are the eight cell probabilities in the $2 \times 2 \times 2$ table corresponding to the dichotomies A or $\bar{A}$, B or $\bar{B}$, and C or $\bar{C}$, where $\sum_{i,j,k} p_{ijk} = 1$, then (1.1)–(1.3) can be expressed equivalently as

$$p_{111}p_{221} \geq p_{121}p_{211}, \tag{1.4}$$

$$p_{112}p_{222} \geq p_{122}p_{212}, \tag{1.5}$$

$$(p_{111} + p_{112})(p_{221} + p_{222}) \leq (p_{121} + p_{122})(p_{211} + p_{212}). \tag{1.6}$$

Question 1.1. Assume that the array of cell probabilities $[p_{ijk}]$ is random and distributed uniformly over the probability simplex

$$\mathcal{S}_8 \equiv \left\{ [p_{ijk}] \middle| p_{ijk} \geq 0, \sum_{i,j,k} p_{ijk} = 1 \right\} \tag{1.7}$$

(a 7-dimensional set in $\mathbb{R}^8$). What is the probability that $[p_{ijk}]$ satisfy a positive or negative Simpson reversal, that is, satisfy the inequalities (1.4)–(1.6) or their opposites? (The exact answer is $1/60$; see Sections 2 and 5.)

It follows that if $[M_{ijk} \mid i, j, k = 1, 2]$ are the cell counts and $[\hat{p}_{i,j,k}] \equiv [M_{i,j,k}/n]$ the corresponding cell probabilities in a $2 \times 2 \times 2$ contingency table, then for large $n \equiv \sum_{i,j,k} M_{ijk}$, approximately $1/60 \approx 1.67\%$ of all possible such probability tables exhibit a Simpson reversal.

Related questions have been addressed earlier by Huh (1987), Jeon, Chung, and Bae (1987), Šleževičienė-Steuding and Steuding (2006), and by Hadjicostas (1997, 1998, 2001). Hadjicostas's results are reviewed in Section 5 where they are applied to obtain the exact answer $1/60$ to Question 1.1.

In Section 3, several questions involving conditional probabilities of a Simpson reversal are addressed, such as the following: Given that the events A and B are negatively correlated (as in (1.3)), what is the conditional probability that they are positively correlated under both C and $\bar{C}$ (as in (1.1) and (1.2))? For example, if one treatment is less effective than another in the aggregate over large and small kidney stones, what is the probability that it is actually more effective for both large and small kidney stones considered separately? More precisely:

Question 1.2. Again assume that $[p_{ijk}]$ is uniformly distributed over $\mathcal{S}_8$. Given that $[p_{ijk}]$ satisfies (1.6), what is the probability that $[p_{ijk}]$ satisfies (1.4) and (1.5)? Conversely, given that $[p_{ijk}]$ satisfies (1.4) and (1.5), what is the probability that $[p_{ijk}]$ satisfies (1.6)? (The answers are easy to obtain.)

Lastly, if it is assumed that

$$[M_{ijk}] | [p_{ijk}] \sim M_8(n; [p_{ijk}]), \tag{1.8}$$

the multinomial distribution with 8 cells, n total observations, and cell probabilities $[p_{ijk}]$, then we can pose the following inferential question:

Question 1.3. Again, assume that $[p_{ijk}]$ is uniformly distributed over $\mathcal{S}_8$ a priori. If the observed cell proportions $[\hat{p}_{i,j,k}]$ exhibit a Simpson reversal, what is the posterior probability that $[p_{ijk}]$ satisfy the same Simpson reversal?

This question is addressed in Section 4 as a Bayesian decision problem under both the uniform prior and the Jeffreys prior on $\mathcal{S}_8$ for the multinomial sampling model. In fact, the method can be applied for all Dirichlet prior distributions on $\mathcal{S}_8$, which class is rich enough to include informative priors, that is, those that are intended to reflect prior information.

Questions 1.1, 1.2, and 1.3 are also addressed throughout the paper for the more general case where the third (conditioning) factor C has 3 or more levels. Section 4 contains several illustrative examples, including an analysis of the well-known Simpson reversal of the batting averages of baseball stars Derek Jeter and David Justice.

Simpson's Paradox is usually attributed to Simpson (1951). However, according to Good and Mittal (1987, p. 695), the underlying idea goes back at least as far as Pearson, Lee, and Bramley-Moore (1899) and Yule (1903). Good and Mittal (1987) attribute the term "Simpson's Paradox" to Blyth (1972).

2. THE PRIOR PROBABILITY OF SIMPSON'S PARADOX

Suppose that factors A and B each have 2 levels while factor C has ℓ levels, $\ell \in \{2, 3, \ldots\}$. As in Section 1 we could denote the $2 \times 2 \times \ell$ table of population cell probabilities by $[p_{ijk}]$. However, to avoid triple subscripts, we instead represent p_{ijk} by $p_{4k+2i+j-6}$ for $i, j = 1, 2$ and $k = 1, \ldots, \ell$. Thus, a *positive Simpson reversal* occurs when

$$p_1 p_4 \geq p_2 p_3, \qquad (R_1^+)$$

$$p_5 p_8 \geq p_6 p_7, \qquad (R_2^+)$$

$$\vdots$$

$$p_{4\ell-3} p_{4\ell} \geq p_{4\ell-2} p_{4\ell-1}, \qquad (R_\ell^+)$$

but

$$\left(\sum_{k=1}^{\ell} p_{4k-3}\right) \cdot \left(\sum_{k=1}^{\ell} p_{4k}\right) \leq \left(\sum_{k=1}^{\ell} p_{4k-2}\right) \cdot \left(\sum_{k=1}^{\ell} p_{4k-1}\right), \qquad (R_{\ell+1}^+)$$

with at least one inequality strict. A *negative Simpson reversal* occurs when these $\ell + 1$ inequalities are replaced by their opposites, which events we denote by $R_1^-, \ldots, R_\ell^-, R_{\ell+1}^-$. Thus:

$$\{\textit{positive Simpson reversal}\} = \bigcap_{k=1}^{\ell+1} R_k^+, \qquad (2.1)$$

$$\{\textit{negative Simpson reversal}\} = \bigcap_{k=1}^{\ell+1} R_k^-, \qquad (2.2)$$

$$\{\textit{Simpson's Paradox}\} = \left(\bigcap_{k=1}^{\ell+1} R_k^+\right) \cup \left(\bigcap_{k=1}^{\ell+1} R_k^-\right). \qquad (2.3)$$

The array of cell probabilities $\mathbf{p} \equiv [p_1, \ldots, p_{4\ell}]$ for the $2 \times 2 \times \ell$ table lies in the probability simplex

$$\mathcal{S}_{4\ell} := \{\mathbf{p} | p_1 \geq 0, \ldots, p_{4\ell} \geq 0, p_1 + \cdots + p_{4\ell} = 1\}, \qquad (2.4)$$

a $(4\ell - 1)$-dimensional set in $\mathbb{R}^{4\ell}$. The general Dirichlet distribution on $\mathcal{S}_{4\ell}$, denoted by $D_{4\ell}(\boldsymbol{\alpha})$, has probability density function proportional to $\prod_{r=1}^{4\ell} p_r^{\alpha_r - 1}$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{4\ell})$ and each $\alpha_r > 0$ (cf. Hartigan 1983, p. 97). Because $\bigcap_{k=1}^{\ell+1} R_k^+$ and $\bigcap_{k=1}^{\ell+1} R_k^-$ are essentially disjoint events in $\mathcal{S}_{4\ell}$,

$$\pi_\ell(\boldsymbol{\alpha}) := \Pr(\text{Simpson's Paradox under } D_{4\ell}(\boldsymbol{\alpha})) \qquad (2.5)$$

$$= P_{\boldsymbol{\alpha}}\left(\bigcap_{k=1}^{\ell+1} R_k^+\right) + P_{\boldsymbol{\alpha}}\left(\bigcap_{k=1}^{\ell+1} R_k^-\right) \qquad (2.6)$$

$$=: \pi_\ell^+(\boldsymbol{\alpha}) + \pi_\ell^-(\boldsymbol{\alpha}). \qquad (2.7)$$

The following representation (cf. Wilks 1962, section 7.7, or Hartigan 1983, p. 97) of the Dirichlet distribution in terms of independent gamma random variables simplifies the calculation of probabilities of events in $\mathcal{S}_{4\ell}$ that involve $R_1^+, \ldots, R_\ell^+, R_{\ell+1}^+$ and/or $R_1^-, \ldots, R_\ell^-, R_{\ell+1}^-$, such as $\bigcap_{k=1}^{\ell+1} R_k^+$ and $\bigcap_{k=1}^{\ell+1} R_k^-$.

Proposition 2.1. Let $V_1, \ldots, V_{4\ell}$ be independent random variables with $V_r \sim \text{Gamma}(\alpha_r, 1)$ for $r = 1, \ldots, \ell$. Then,

$$\mathbf{p} := \left(\frac{V_1}{\sum_{r=1}^{4\ell} V_r}, \ldots, \frac{V_{4\ell}}{\sum_{r=1}^{4\ell} V_r}\right) \sim D_{4\ell}(\boldsymbol{\alpha}). \qquad (2.8)$$

Thus, $R_1^+, \ldots, R_\ell^+, R_{\ell+1}^+$ can be expressed in terms of $V_1, \ldots, V_{4\ell}$ as follows:

$$V_1 V_4 \geq V_2 V_3, \qquad (R_1^+)$$

$$\vdots$$

$$V_{4\ell-3} V_{4\ell} \geq V_{4\ell-2} V_{4\ell-1}, \qquad (R_\ell^+)$$

$$\left(\sum_{k=1}^{\ell} V_{4k-3}\right) \cdot \left(\sum_{k=1}^{\ell} V_{4k}\right) \leq \left(\sum_{k=1}^{\ell} V_{4k-2}\right) \cdot \left(\sum_{k=1}^{\ell} V_{4k-1}\right). \qquad (R_{\ell+1}^+)$$

Remark 2.1. It follows from this representation that the events $R_1^+, \ldots, R_\ell^+$ are mutually independent under any Dirichlet distribution on $\mathcal{S}_{4\ell}$; the same is true for $R_1^-, \ldots, R_\ell^-$.

By simulating the independent gamma random variables $V_1, \ldots, V_{4\ell}$ we can apply Proposition 2.1 to approximate the probability of a Simpson reversal under any Dirichlet distribution $D_{4\ell}(\boldsymbol{\alpha})$ on $\mathcal{S}_{4\ell}$.

When $\boldsymbol{\alpha} = (\alpha, \ldots, \alpha)$ we write $D_{4\ell}(\boldsymbol{\alpha})$ as $D_{4\ell}(\alpha)$, a symmetric/exchangeable distribution on $\mathcal{S}_{4\ell}$. Note that $D_{4\ell}(1)$ is the uniform distribution on $\mathcal{S}_{4\ell}$, while $D_{4\ell}(0.5)$ is the Jeffreys prior distribution for the multinomial model (4.4), of which (1.8) is a special case (cf. Hartigan 1983, p. 100; Davison 2003, p. 575).

We now approximate the probability of Simpson's Paradox under $D_{4\ell}(\alpha)$ for several values of α. Here, $V_1, \ldots, V_{4\ell}$ are iid Gamma$(\alpha, 1)$ random variables, hence exchangeable, so the positive and negative Simpson reversals $\bigcap_{k=1}^{\ell+1} R_k^+$ and $\bigcap_{k=1}^{\ell+1} R_k^-$ have the same probabilities. Thus, from (2.6),

$$\pi_\ell(\alpha) := \Pr(\text{Simpson's Paradox under } D_{4\ell}(\alpha)) \tag{2.9}$$

$$= 2P_\alpha\left(\bigcap_{k=1}^{\ell+1} R_k^+\right) \tag{2.10}$$

$$= 2\pi_\ell^+(\alpha). \tag{2.11}$$

Table 1 displays 99% confidence intervals for $\pi_\ell(\alpha)$ based on 10^7 Monte Carlo simulations of $V_1, \ldots, V_{4\ell}$. The following features are noteworthy:

(a) [$\alpha = 1$]: This case corresponds to the uniform distribution on the probability simplex $\mathcal{S}_{4\ell}$. The 99% confidence interval for $\pi_2(1)$ is 0.0166 ± 0.00010, so this is the probability of Simpson's Paradox in a $2 \times 2 \times 2$ table under the uniform distribution. Note that this agrees with the exact value $1/60$ found in Section 5.

(b) [$\alpha = 0.5$]: This case corresponds to the Jeffreys distribution on $\mathcal{S}_{4\ell}$. The 99% confidence interval for $\pi_2(0.5)$ is 0.0267 ± 0.00013, so this is the probability of Simpson's Paradox in a $2 \times 2 \times 2$ table under the Jeffreys distribution. That this exceeds the value 0.0166 under the uniform distribution is expected because the Jeffreys distribution puts more of its mass near the boundary of $\mathcal{S}_{4\ell}$, where the Simpson reversals tend to be most pronounced. (See, e.g., Blyth 1972, p. 365.)

(c) For each fixed ℓ, $\pi_\ell(\alpha)$ *decreases as* α *increases*. This is plausible, as follows: If $\mathbf{p} \equiv [p_1, \ldots, p_{4\ell}] \sim D_{4\ell}(\alpha)$ then for each $r = 1, \ldots, 4\ell$, $\mathrm{E}_\alpha(p_r) = 1/4\ell$ while

$$\mathrm{Var}_\alpha(p_r) = \frac{4\ell - 1}{(4\ell)^2(4\alpha\ell + 1)} = O\left(\frac{1}{\alpha}\right) \tag{2.12}$$

(cf. Wilks 1962, p. 179), so the distribution of $\mathbf{p}$ concentrates around the central point $(1/4\ell, \ldots, 1/4\ell)$ of $\mathcal{S}_{4\ell}$ as α increases. However, Blyth (1972) notes that the most extreme Simpson reversals occur near the boundary of $\mathcal{S}_{4\ell}$. This suggests that Simpson's Paradox should become less likely as α increases.

(d) For each fixed α, $\pi_\ell(\alpha)$ *decreases as* ℓ *increases*. This, too, is plausible because a reversal from ℓ positive conditional correlations to one negative unconditional correlation seems increasingly unlikely for large ℓ.

In fact, it is easy to show that $\pi_\ell(\alpha)$ decreases at least exponentially fast in ℓ: From (2.10), Remark 2.1, and the exchangeability of $V_1, \ldots, V_{4\ell}$,

$$\pi_\ell(\alpha) \le 2P_\alpha\left(\bigcap_{k=1}^{\ell} R_k^+\right) = \frac{1}{2^{\ell-1}}. \tag{2.13}$$

We conjecture that $\pi_\ell(\alpha)$ *decreases at an exponential rate*.

Conjecture 2.1. For each $\alpha > 0$ there exists $h(\alpha) > 0$ such that

$$\pi_\ell(\alpha) \approx \pi_2(\alpha) \cdot \exp\left\{-h(\alpha)\left(\frac{\ell}{2} - 1\right)\right\},$$
$$\ell = 2, 3, \ldots. \tag{2.14}$$

As evidence for this conjecture we estimated $\pi_\ell(\alpha)$ for $\alpha \in \{0.5, 1\}$ and $\ell \in \{2, 3, \ldots, 30\}$, again based on 10^7 Monte Carlo simulations of $V_1, \ldots, V_{4\ell}$. Denote by $\hat{\pi}_\ell(\alpha)$ the point estimate of $\pi_\ell(\alpha)$ thus obtained. We found that for $\ell > 10$, $\hat{\pi}_\ell(\alpha) \approx 0$ to the accuracy of our simulation, so we restrict attention to $\ell \in \{2, 3, \ldots, 10\}$.

In Figure 1 we plot $\log(\hat{\pi}_\ell(\alpha)/\hat{\pi}_2(\alpha))$ against $(\ell/2) - 1$ for the cases $\alpha = 0.5$ and $\alpha = 1$. The plotted points agree well with the conjectured exponential rates of convergence. Point estimates $\hat{h}(\alpha)$ were obtained by averaging

$$-\frac{\log(\hat{\pi}_\ell(\alpha)/\hat{\pi}_2(\alpha))}{(\ell/2) - 1}$$

across $\ell \in \{3, 4, \ldots, 10\}$, yielding $\hat{h}(0.5) \approx 1.85$ and $\hat{h}(1) \approx 2.27$.

3. SOME CONDITIONAL PROBABILITIES OF SIMPSON REVERSALS

In Section 2 we approximated the unconditional probability $\pi_\ell(\alpha)$ (recall (2.10)) of the occurrence of Simpson's Paradox in a $2 \times 2 \times \ell$ table under several symmetric Dirichlet distributions $D_{4\ell}(\alpha)$ on the probability simplex $\mathcal{S}_{4\ell}$. We now consider the

Table 1. 99% confidence intervals for $\pi_\ell(\alpha)$ under the symmetric Dirichlet prior $D_{4\ell}(\alpha)$. The confidence intervals are based on 10^7 Monte Carlo simulations of the iid Gamma$(\alpha, 1)$ random variables $V_1, \ldots, V_{4\ell}$ (cf. (2.8)).

$\pi_t(\alpha)$	$\boldsymbol{\ell = 2}$	$\boldsymbol{\ell = 3}$	$\boldsymbol{\ell = 4}$	$\boldsymbol{\ell = 5}$	$\boldsymbol{\ell = 6}$
$\boldsymbol{\alpha = 0.5}$	0.02670 ± 0.000131	$0.01161 \pm 8.73\text{e–}05$	$0.004506 \pm 5.46\text{e–}05$	$0.001712 \pm 3.37\text{e–}05$	$0.0006246 \pm 2.04\text{e–}05$
$\boldsymbol{\alpha = 1.0}$	0.01659 ± 0.000104	$0.005656 \pm 6.11\text{e–}05$	$0.001785 \pm 3.44\text{e–}05$	$0.0005434 \pm 1.90\text{e–}05$	$0.0001636 \pm 1.04\text{e–}05$
$\boldsymbol{\alpha = 2.0}$	$0.009270 \pm 7.81\text{e–}05$	$0.002390 \pm 3.98\text{e–}05$	$0.0005764 \pm 1.96\text{e–}05$	$0.0001406 \pm 9.66\text{e–}06$	$3.640\text{e–}05 \pm 4.91\text{e–}06$
$\boldsymbol{\alpha = 3.0}$	$0.006282 \pm 6.44\text{e–}05$	$0.001375 \pm 3.02\text{e–}05$	$0.0002830 \pm 1.37\text{e–}05$	$6.360\text{e–}05 \pm 6.50\text{e–}06$	$1.080\text{e–}05 \pm 2.68\text{e–}06$
$\boldsymbol{\alpha = 4.0}$	$0.004809 \pm 5.64\text{e–}05$	$0.0008955 \pm 2.44\text{e–}05$	$0.0001630 \pm 1.04\text{e–}05$	$3.160\text{e–}05 \pm 4.58\text{e–}06$	$7.800\text{e–}06 \pm 2.28\text{e–}06$
$\boldsymbol{\alpha = 5.0}$	$0.003879 \pm 5.06\text{e–}05$	$0.0007016 \pm 2.16\text{e–}05$	$0.0001118 \pm 8.61\text{e–}06$	$1.640\text{e–}05 \pm 3.30\text{e–}06$	$2.400\text{e–}06 \pm 1.26\text{e–}06$

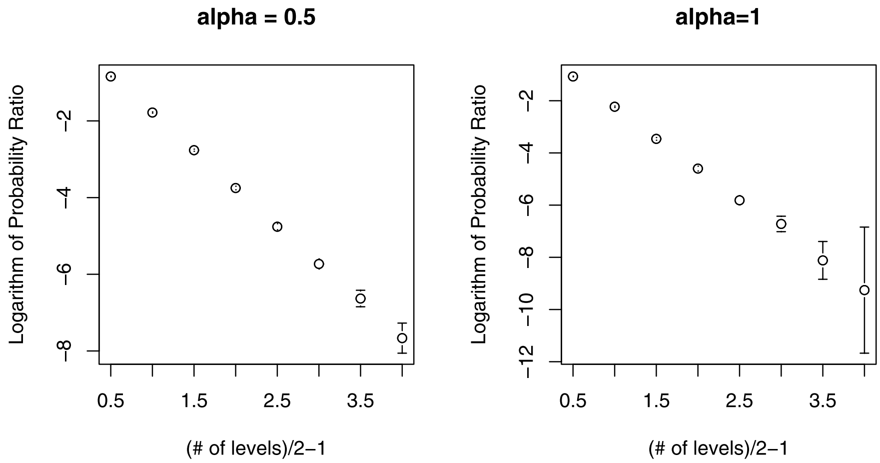

Figure 1. The values of $\log(\hat{\pi}_\ell(\alpha)/\hat{\pi}_2(\alpha))$ are plotted against $(\ell/2)-1$, for the cases $\alpha = 0.5$ and $\alpha = 1$. The bracketing intervals represent 99% confidence intervals for $\log(\pi_\ell(\alpha)/\pi_2(\alpha))$.

following four *conditional* probabilities of Simpson reversals and related events under $D_{4\ell}(\alpha)$.

$$\beta_\ell(\alpha) := P_\alpha\left(\bigcap_{k=1}^{\ell} R_k^+ \middle| R_{\ell+1}^+\right) = \pi_\ell(\alpha), \tag{3.1}$$

$$\gamma_\ell(\alpha) := P_\alpha\left(R_{\ell+1}^+ \middle| \bigcap_{k=1}^{\ell} R_k^+\right) = 2^{\ell-1}\pi_\ell(\alpha), \tag{3.2}$$

$$\delta_\ell(\alpha) := P_\alpha\left(\bigcap_{k=1}^{\ell} R_k^+ \middle| R_{\ell+1}^-\right) = \frac{1}{2^{\ell-1}} - \pi_\ell(\alpha), \tag{3.3}$$

$$\epsilon_\ell(\alpha) := P_\alpha\left(R_{\ell+1}^- \middle| \bigcup_{k=1}^{\ell} R_k^+\right) = \frac{1-\pi_\ell(\alpha)}{2-\frac{1}{2^{\ell-1}}}. \tag{3.4}$$

Result (3.1) follows easily from the definition of conditional probability, from (2.10), and from the symmetry of the prior $D_{4\ell}(\alpha)$ which implies that

$$P_\alpha(R_k^+) = P_\alpha(R_k^-) = \tfrac{1}{2}, \qquad k = 1, \ldots, \ell+1. \tag{3.5}$$

Result (3.2) follows from (3.1) by Bayes' formula, independence, and symmetry.

Result (3.3) is derived as follows:

$$\begin{aligned} P_\alpha\left(\bigcap_{k=1}^{\ell} R_k^+ \middle| R_{\ell+1}^-\right) &= 2P_\alpha\left(R_{\ell+1}^- \cap \bigcap_{k=1}^{\ell} R_k^+\right) \\ &= 2\left[P_\alpha\left(\bigcap_{k=1}^{\ell} R_k^+\right) - P_\alpha\left(\bigcap_{k=1}^{\ell+1} R_k^+\right)\right] \\ &= \frac{1}{2^{\ell-1}} - \pi_\ell(\alpha), \end{aligned}$$

where we use the fact that $\bar{R}_{\ell+1}^- = R_{\ell+1}^+$ almost surely, the independence of $R_1^\pm, \ldots, R_\ell^\pm$ (Remark 2.1), and the symmetry of $D_{4\ell}(\alpha)$.

Result (3.4) is derived by Bayes formula, then complementation.

By the symmetry of $D_{4\ell}(\alpha)$, (3.1)–(3.4) remain valid if R_k^+ and R_k^- are interchanged throughout.

Because the conditional probabilities in (3.1)–(3.4) are simple functions of $\pi_\ell(\alpha)$, they can be evaluated using the entries in Table 1. For example, the answers to the two parts of Question 1.2 appear in (i) and (iii) below:

(i) If it is known that Treatment 1 (T1) is *less* effective than Treatment 2 (T2) in the aggregate over large and small kidney stones, then under the uniform distribution $D_8(1)$ [resp. the Jeffreys distribution $D_8(0.5)$] the conditional probability that T1 is *more* effective than T2 for large and small stones separately is $\beta_2(1) = \pi_2(1) \approx 0.0166$ [resp. $\beta_2(0.5) = \pi_2(0.5) \approx 0.0267$].

(ii) If T1 is *more* effective than T2 for large and small stones separately, then under $D_8(1)$ [resp. $D_8(0.5)$] the conditional probability that T1 is *less* effective in the aggregate is $\gamma_2(1) = 2\pi_2(1) \approx 0.0332$ [resp. $\gamma_2(0.5) = 2\pi_2(0.5) \approx 0.0534$].

(iii) If T1 is *more* effective than T2 in the aggregate over large and small stones, then under $D_8(1)$ [resp. $D_8(0.5)$] the conditional probability that T1 is *more* effective for large and small stones separately is only $\delta_2(1) = 1/2 - \pi_2(1) \approx 0.483$ [resp. $\delta_2(0.5) = 1/2 - \pi_2(0.5) \approx 0.473$].

(iv) If T1 is *more* effective than T2 for large or small stones or both, then under $D_8(1)$ [resp. $D_8(0.5)$] the conditional probability that T1 is *more* effective in the aggregate is $\epsilon_2(1) \approx 0.656$ [resp. $\epsilon_2(0.5) \approx 0.649$].

4. THE POSTERIOR PROBABILITY OF SIMPSON'S PARADOX

We now return to Question 1.3: Given that the observed cell proportions $[\hat{p}_{ijk}]$ in a $2 \times 2 \times \ell$ contingency table exhibit a Simpson reversal, how likely is it that the *population* cell probabilities $[p_{ijk}]$ satisfy the same Simpson reversal?

It is easy to answer this question under the Bayesian Dirichlet-multinomial model. As in Section 2, represent the cell probabilities, cell counts, and cell proportions as

$$[p_{ijk}] \leftrightarrow [p_{4k+2i+j-6}] =: \mathbf{p}, \tag{4.1}$$

$$[M_{ijk}] \leftrightarrow [M_{4k+2i+j-6}] =: \mathbf{M}, \tag{4.2}$$

$$[\hat{p}_{ijk}] \leftrightarrow [\hat{p}_{4k+2i+j-6}] =: \hat{\mathbf{p}}, \tag{4.3}$$

respectively, where $i, j = 1, 2$ and $k = 1, \ldots, \ell$. It is well known (cf. Hartigan, 1983, p. 97) that the Dirichlet distributions constitute a conjugate family of prior distributions for the multinomial model

$$\mathbf{M} \mid \mathbf{p} \sim M_{4\ell}(n; \mathbf{p}) \qquad \text{and} \qquad \mathbf{p} \sim D_{4\ell}(\boldsymbol{\alpha}) \tag{4.4}$$

imply

$$\mathbf{p} \mid \mathbf{M} \sim D_{4\ell}(\boldsymbol{\alpha} + \mathbf{M}). \tag{4.5}$$

It follows from (4.5) and (2.5)–(2.7) that under this Bayesian model, the posterior probabilities that $\mathbf{p}$ satisfies a positive or negative Simpson reversal are simply $\pi_\ell^+(\boldsymbol{\alpha} + \mathbf{M})$ and $\pi_\ell^-(\boldsymbol{\alpha} + \mathbf{M})$, respectively. In the symmetric case where $\boldsymbol{\alpha} = (\alpha, \ldots, \alpha)$ we denote these by $\pi_\ell^+(\alpha + \mathbf{M})$ and $\pi_\ell^-(\alpha + \mathbf{M})$. As in Section 2, we can apply Proposition 2.1 to estimate these probabilities by simulating independent gamma random variables.

We now provide several examples where $\hat{\mathbf{p}}$ (equivalently, $\mathbf{M}$) exhibits a positive Simpson reversal and obtain estimates of the posterior probability $\pi_\ell^+(\boldsymbol{\alpha} + \mathbf{M})$ that $\mathbf{p}$ also satisfies a positive reversal. These estimates are then used to estimate the Bayes factor

$$\mathrm{BF}^+(\boldsymbol{\alpha}, \mathbf{M}) := \frac{\pi_\ell^+(\boldsymbol{\alpha} + \mathbf{M})/(1 - \pi_\ell^+(\boldsymbol{\alpha} + \mathbf{M}))}{\pi_\ell^+(\boldsymbol{\alpha})/(1 - \pi_\ell^+(\boldsymbol{\alpha}))}, \tag{4.6}$$

the ratio of the posterior odds to the prior odds of a positive Simpson reversal versus its complement. When $\boldsymbol{\alpha} = (\alpha, \ldots, \alpha)$ we write $\mathrm{BF}^+(\alpha, \mathbf{M})$. The Bayes factor can be viewed as a measure of the strength of evidence favoring a hypothesis versus its complement; cf. Kass and Raftery (1995).

As before, we shall consider two "uninformative" priors, the Jeffreys prior $D_{4\ell}(0.5)$ and the uniform prior $D_{4\ell}(1)$. Our method can also be applied with any informative Dirichlet prior $D_{4\ell}(\boldsymbol{\alpha})$, for example, with $\boldsymbol{\alpha}$ chosen to reflect prior information about the first and second moments of $\mathbf{p}$; see Example 4.4.

Example 4.1. In this small example, factor C has 2 levels ($\ell = 2$); the multinomial data $\mathbf{M}$ is displayed in Table 2. The reader should verify that $\mathbf{M}$ exhibits a positive Simpson reversal.

Again, using 10^7 Monte Carlo simulations we obtained the point estimates $\hat{\pi}_2^+(0.5 + \mathbf{M}) = 0.1964$ and $\hat{\pi}_2^+(1 + \mathbf{M}) = 0.1745$, which yield the estimates $\widehat{\mathrm{BF}}^+(0.5, \mathbf{M}) = 18.1$ and $\widehat{\mathrm{BF}}^+(1, \mathbf{M}) = 25.3$ for the Bayes factors in (4.6) under the priors $D_8(\alpha)$ for $\alpha = 0.5$ and 1, respectively. According to Kass and Raftery (1995), these provide "positive" to fairly "strong" posterior evidence in favor of a positive Simpson reversal.

Table 2. [Example 4.1.] Multinomial data $\mathbf{M}$ with a positive Simpson reversal in a $2 \times 2 \times 2$ contingency table with factors A, B, and C each having two levels. Also shown are the aggregated cell counts for factors A and B when summed over the two levels of factor C.

	C_1			C_2			Aggregate over C	
Level	B_1	B_2	Level	B_1	B_2	Level	B_1	B_2
A_1	2	1	A_1	2	3	A_1	4	4
A_2	5	3	A_2	1	2	A_2	6	5

Example 4.2. Here, Example 4.1 is extended (see Table 3) by adding a third level (for simplicity, we just repeat the second level) of factor C (so $\ell = 3$) in which factors A and B are again conditionally positively correlated, while they remain negatively correlated when aggregated over all three levels of C. How will this affect the posterior evidence for a positive Simpson reversal?

On the one hand, the evidence for a positive Simpson reversal may be stronger because A and B are positively correlated for three levels of C rather than for only two levels as in Example 4.1. On the other hand, the estimated prior probabilities of a positive Simpson reversal have decreased from $\hat{\pi}_2^+(0.5) = 0.01335$ and $\hat{\pi}_2^+(1) = 0.00830$ in Example 4.1 to $\hat{\pi}_3^+(0.5) = 0.00586$ and $\hat{\pi}_3^+(1) = 0.00283$ under the priors $D_{12}(0.5)$ and $D_{12}(1)$ here.

Now we obtain the point estimates $\hat{\pi}_3^+(0.5 + \mathbf{M}) = 0.2191$ and $\hat{\pi}_3^+(1 + \mathbf{M}) = 0.1973$, which in turn yield the estimated Bayes factors $\widehat{\mathrm{BF}}^+(0.5, \mathbf{M}) = 47.6$ and $\widehat{\mathrm{BF}}^+(1, \mathbf{M}) = 86.6$. Thus, the evidence for a positive Simpson reversal is *stronger* than in Example 4.1—the occurrence of an observed positive Simpson reversal in a third conditioning level outweighs the rapid decrease (recall Conjecture 2.1) of the prior probability of a Simpson reversal caused by the addition of the third level.

Example 4.3. Our third example (see Table 4), a modification of the university admissions example in Moore (1991, example 2, p. 246), demonstrates the effect of increased sample size on the posterior evidence for a Simpson reversal.

The estimated prior probabilities of a positive Simpson reversal are again $\hat{\pi}_2^+(0.5) = 0.01335$ under the Jeffreys prior $D_8(0.5)$ and $\hat{\pi}_2^+(1) = 0.00830$ under the uniform prior $D_8(1)$. As N ranges from 1 to 10 the estimated posterior probability $\hat{\pi}_2^+(0.5 + \mathbf{M} \cdot N)$ of this reversal ranges from 0.2071 to 0.7742 under the Jeffreys prior, while $\hat{\pi}_2^+(1 + \mathbf{M} \cdot N)$ ranges from 0.1881 to 0.7634 under the uniform prior. The estimated Bayes factors shown in Table 5 range from 19.3 to 253.4 for the Jeffreys prior and from 27.7 to 385.5 for the uniform prior. Thus, for $N = 10$ and according to Kass and Raftery (1995), these data would provide "decisive" evidence for a positive Simpson reversal.

Example 4.4. Ross (2004, p. 12) noted the occurrence of a Simpson reversal in the comparative batting averages of Major League Baseball players David Justice and Derek Jeter for the three years 1995, 1996, and 1997. The count data $\mathbf{M}$ are presented in Table 6.

The data in Table 6 exhibit a positive Simpson reversal: Justices's batting averages for the three years were 0.253, 0.321, and 0.329, each *greater* than Jeter's averages of 0.250, 0.314, and 0.291, but Justices's average aggregated over the three years was 0.298, which is *less* than Justice's three-year average of 0.300. While baseball fans continue to debate who was the better hitter in the face of this apparent paradox, as in the previous examples we can estimate the posterior probability that

Table 3. [Example 4.2.] Multinomial data $\mathbf{M}$ with a positive Simpson reversal in a $2 \times 2 \times 3$ contingency table with factors A, B, and C. Also shown are the aggregated cell counts for factors A and B when summed over the three levels of factor C.

	C_1			C_2			C_3			Aggregate over C	
Level	B_1	B_2	Level	B_1	B_2	Level	B_1	B_2	Level	B_1	B_2
A_1	2	1	A_1	2	3	A_1	2	3	A_1	6	7
A_2	5	3	A_2	1	2	A_2	1	2	A_2	7	7

Table 4. [Example 4.3.] Multinomial data $\mathbf{M} \cdot N$ with a positive Simpson reversal in a $2 \times 2 \times 2$ contingency table with three factors: Application Status, Gender, and Department. The multiplier N takes on values $\{1, 5, 10\}$. The third table gives the cell counts for factors "Application Status" and "Gender" when aggregated over the two levels of "Department."

	Physics Dept.			English Dept.			Aggregate over Dept.	
Level	**Female**	**Male**	Level	**Female**	**Male**	Level	**Female**	**Male**
Admit	$6N$	$5N$	**Admit**	$25N$	$2N$	**Admit**	$31N$	$7N$
Deny	$4N$	$5N$	**Deny**	$75N$	$8N$	**Deny**	$79N$	$13N$

Table 5. [Example 4.3.] Point estimates for the Bayes factor $\widehat{\mathrm{BF}}^{+}(\alpha, \mathbf{M} \cdot N)$ for a positive Simpson reversal under the Dirichlet priors $D_8(0.5)$ and $D_8(1)$. This indicates the rate of increase of $\widehat{\mathrm{BF}}^{+}(\alpha, \mathbf{M} \cdot N)$ as the sample size multiplier N increases from 1 to 10.

Point estimate	$\alpha = 0.5$	$N = 1$	19.3
		$N = 5$	100.7
		$N = 10$	253.4
$\widehat{\mathrm{BF}}^{+}(\alpha, \mathbf{M} \cdot N)$	$\alpha = 1.0$	$N = 1$	27.7
		$N = 5$	153.1
		$N = 10$	385.5

the innate cell probabilities for the $2 \times 2 \times 3$ table in Table 6 actually satisfy the positive Simpson reversal suggested by the data.

For this purpose, we shall use two "semi-informative" Dirichlet priors $D_{12}(\boldsymbol{\alpha})$ where $\boldsymbol{\alpha}$ is chosen so that the first moments of $D_{12}(\boldsymbol{\alpha})$ match the *expected* averages of Justice and Jeter in each year. It is generally accepted that the batting average 0.300 represents the standard for good hitters. (This is borne out by the three-year aggregated averages of these two players.) Thus, we expect Justice to have had $(0.3) \times 411 = 123.3$ safe hits in 1995, $(0.3) \times 140 = 42.0$ safe hits in 1996, and $(0.3) \times 495 = 148.5$ safe hits in 1997; the other entries in Table 7 are obtained similarly.

The 12 expected counts in Table 7, denoted by $[E_{ijk}]$, are then each divided by 2330, the total three-year sample size, yielding 12 estimated a priori cell probabilities $[\tilde{p}_{ijk}]$. The Dirichlet parameters $\boldsymbol{\alpha} \equiv [\alpha_{ijk}]$ are selected to be proportional to $\tilde{\mathbf{p}} \equiv [\tilde{p}_{ijk}]$, that is, $\alpha_{ijk} = c\,\tilde{p}_{ijk}$ for some $c > 0$. We consider two values of c: $c = 6$ so that $\sum \alpha_{ijk} = 6$ (as in the Jeffreys prior $D_{12}(0.5)$), and $c = 12$ [so $\sum \alpha_{ijk} = 12$ as in the uniform prior $D_{12}(1)$].

Monte Carlo estimates of the prior and posterior probabilities for $c = 6$ and $c = 12$ are obtained via Proposition 2.1: $\hat{\pi}_3^{+}(6\tilde{\mathbf{p}}) = 0.01571$, $\hat{\pi}_3^{+}(12\tilde{\mathbf{p}}) = 0.01312$, $\hat{\pi}_3^{+}(6\tilde{\mathbf{p}} + \mathbf{M}) = 0.07463$, and $\hat{\pi}_3^{+}(12\tilde{\mathbf{p}} + \mathbf{M}) = 0.07458$ where $\mathbf{M}$ is the count data in Table 6. These yield the estimated Bayes factors $\widehat{\mathrm{BF}}^{+}(6\tilde{\mathbf{p}}, \mathbf{M}) = 5.05$ and $\widehat{\mathrm{BF}}^{+}(12\tilde{\mathbf{p}}, \mathbf{M}) = 6.06$, which provide only mildly positive evidence of an innate positive Simpson reversal for Justice and Jeter for the years 1995–1997.

Ross (2004, p. 13) makes the following intriguing statement regarding the occurrence of Simpson's Paradox in these data (considering only the two years 1995 and 1996): "This

Table 6. [Example 4.4.] Batting data for David Justice and Derek Jeter over three consecutive years. H = "hit safely," O = "out." The aggregated data appear in Table 4.

	1995			1996			1997			1995–1997	
Level	**H**	**O**	Level	**H**	**O**	Level	**H**	**O**	Level	**H**	**O**
Justice	104	307	**Justice**	45	95	**Justice**	163	332	**Justice**	312	734
Jeter	12	36	**Jeter**	183	399	**Jeter**	190	464	**Jeter**	385	899

Table 7. [Example 4.4.] The expected counts $[E_{ijk}]$ used to construct the semi-informative Dirichlet priors $D_{12}(c\tilde{\mathbf{p}})$.

	1995			1996			1997	
Level	**H**	**O**	Level	**H**	**O**	Level	**H**	**O**
Justice	123.3	287.7	**Justice**	42	98	**Justice**	148.5	346.5
Jeter	14.4	33.6	**Jeter**	174.6	407.4	**Jeter**	196.2	457.8

bizarre phenomenon seems to occur for some pair of interesting ballplayers about once a year." Ross (2004) does not define "interesting" but we might speculate as follows.

If I denotes the number of interesting players, then $\binom{I}{2}$ is the number of possible occurrences of Simpson's Paradox among them. Under the uniform prior $D_8(1)$ the probability of Simpson's Paradox is $\hat{\pi}_2(1) = 1/60$, so we should expect to see about $\binom{I}{2}/60$ occurrences per year. Thus for $I = 12$ "interesting" hitters, we would expect about $\binom{12}{2}/60 \approx 1.1$ yearly occurrences of Simpson's Paradox, which would agree with Ross's estimate. For $I = 30$ "interesting hitters" (one for each of the 30 Major League teams), however, about $\binom{30}{2}/60 = 7.25$ yearly occurrences would be expected.

Remark 4.1. To define the occurrence of Simpson's Paradox in a $2 \times 2 \times \ell$ contingency table $A \times B \times C$, we have used two-way odds ratios of the form $p_{11}p_{22}/p_{12}p_{21}$ (see (1.4)–(1.6)) to determine positive or negative dependencies in the 2×2 conditional (given C) and marginal (aggregated over C) tables for the factors A and B. Our results can be extended to a $2 \times 2 \times 2 \times \ell$ table $A \times B \times C \times D$ by considering three-way odds ratios of the form

$$\frac{p_{111}p_{221}p_{122}p_{212}}{p_{121}p_{211}p_{112}p_{222}}, \tag{4.7}$$

which determine positive or negative three-way interactions in the $2 \times 2 \times 2$ conditional (given D) and marginal (aggregated over D) tables for A, B, and C. From this, positive and negative Simpson reversals can be defined similarly to $R_1^+, \ldots, R_\ell^+, R_{\ell+1}^+$ and $R_1^-, \ldots, R_\ell^-, R_{\ell+1}^-$, respectively, and their prior, conditional, and posterior probabilities evaluated as in Sections 2–4.

5. THE EXACT VALUE OF $\pi_2(1)$

In Hadjicostas (1998, 2001) it is assumed that a 2×2 table $[a, b; c, d]$ of known counts is divided *additively and at random* into ℓ 2×2 subtables. It is proved that as $N := a+b+c+d \to \infty$, such that $ad/(bc) \to t \neq 1$, and under mild assumptions, the asymptotic probability of Simpson's Paradox occurring is

$$q(\ell, t) := \begin{cases} [(\ell-1)!]^4 \cdot \int_{S(\ell,t)} \mathrm{d}\lambda_{4(\ell-1)}, & \text{if } 0 < t < 1 \\ [(\ell-1)!]^4 \cdot \int_{S'(\ell,t)} \mathrm{d}\lambda_{4(\ell-1)}, & \text{if } 1 < t < \infty, \end{cases} \tag{5.1}$$

where λ_m is the Lebesgue measure on $\mathbb{R}^m$; $S(\ell, t)$ is the set of all $((\alpha_i, \beta_i, \gamma_i, \delta_i) \mid i = 1, 2, \ldots, \ell-1) \in (0,1)^{4(\ell-1)}$ such that

$$\sum_{m=1}^{\ell-1} \alpha_m < 1, \qquad \sum_{m=1}^{\ell-1} \beta_m < 1,$$
$$\sum_{m=1}^{\ell-1} \gamma_m < 1, \qquad \sum_{m=1}^{\ell-1} \delta_m < 1,$$
$$\frac{\alpha_i \delta_i}{\beta_i \gamma_i} \geq \frac{1}{t}, \qquad i = 1, 2, \ldots, \ell-1,$$

and

$$\frac{\left(1 - \sum_{m=1}^{\ell-1} \alpha_m\right)\left(1 - \sum_{m=1}^{\ell-1} \delta_m\right)}{\left(1 - \sum_{m=1}^{\ell-1} \beta_m\right)\left(1 - \sum_{m=1}^{\ell-1} \gamma_m\right)} \geq \frac{1}{t};$$

and $S'(\ell, t)$ is defined similarly but with both $\geq$ replaced with $\leq$. Hadjicostas shows that

$$q(2, t) = \begin{cases} -\frac{(1-t)^2}{6t} \ln(1-t) + \frac{3t-2}{12}, & \text{if } 0 < t < 1 \\ -\frac{(t-1)^2}{6t} \ln(1-1/t) + \frac{3-2t}{12t}, & \text{if } 1 < t < \infty, \end{cases} \tag{5.2}$$

and that $\sup_t q(2, t) = \lim_{t \to 1} q(2, t) = 1/12$.

The above results can be applied to obtain the exact value of $\pi_2(1)$, the prior probability of Simpson's Paradox in a $2 \times 2 \times 2$ table under the uniform distribution on the probability simplex $\mathcal{S}_8$. The following result has been contributed by Petros Hadjicostas:

Proposition 5.1 (Recall (2.9) on page 228). $\pi_2(1) = 1/60$.

Proof. It follows from (2.11) and Proposition 2.1 that

$$\pi_2(1) = 2 \int_{K_1} \exp\left\{ -\sum_{j=1}^{8} v_j \right\} \mathrm{d}v_1 \cdots \mathrm{d}v_8, \tag{5.3}$$

where

$$K_1 = \big\{ (v_1, \ldots, v_8) \in (0, \infty)^8 \,|\, v_1 v_4 \geq v_2 v_3, v_5 v_8 \geq v_6 v_7, \\ \text{but } (v_1 + v_5) \cdot (v_4 + v_8) \leq (v_2 + v_6) \cdot (v_3 + v_7) \big\}.$$

Consider the 1–1 transformation $T_1 : (v_1, \ldots, v_8) \mapsto (w_1, \ldots, w_4, \delta_1, \delta_2, \delta_3, \delta_4)$ given by $w_j := v_j/(v_j + v_{j+4})$ and $\delta_j := v_j + v_{j+4}$, for $j \in \{1, 2, 3, 4\}$, that maps K_1 onto

$$T_1(K_1) = \Bigg\{ (w_1, \ldots, w_4, \delta_1, \ldots, \delta_4) \in (0,1)^4 \times (0, \infty)^4 \Bigg| \\ t := \frac{\delta_1 \delta_4}{\delta_2 \delta_3} \leq 1, \frac{w_1 w_4}{w_2 w_2} \geq \frac{1}{t} \text{ and} \\ \frac{(1 - w_1) \cdot (1 - w_4)}{(1 - w_2) \cdot (1 - w_3)} \geq \frac{1}{t} \Bigg\}.$$

It is easy to see that that inverse transformation T_1^{-1} is defined by $v_j = w_j \cdot \delta_j$ and $v_{4+j} = (1 - w_j) \cdot \delta_j$, for $j \in \{1, 2, 3, 4\}$, and that the Jacobian of T_1 is equal to $\delta_1 \delta_2 \delta_3 \delta_4 =: J_1$. By changing variables according to T_1 in (5.3) we obtain

$$\pi_2(1) = 2 \int_{K_2} \exp\left\{ -\sum_{j=1}^{4} \delta_j \right\} \cdot \left(\prod_{j=1}^{4} \delta_j \right) \\ \times q\left(2, \frac{\delta_1 \delta_4}{\delta_2 \delta_3}\right) \mathrm{d}\delta_1 \cdots \mathrm{d}\delta_4, \tag{5.4}$$

where $K_2 = \{(\delta_1, \delta_2, \delta_3, \delta_4) \in (0, \infty)^4 \mid \delta_1 \delta_4/(\delta_2 \delta_3) \leq 1\}$.

Now, apply another change of variables according to the 1–1 transformation $T_2 \colon (\delta_1, \delta_2, \delta_3, \delta_4) \mapsto (x, y, z, t)$ defined by $x = \delta_1, y = \delta_2, z = \delta_3, t = \delta_1 \delta_4/(\delta_2 \delta_3)$. Equivalently, $\delta_1 = x, \delta_2 = y, \delta_3 = z, \delta_4 = tyz/x$. Then T_2 maps K_2 onto $T_2(K_2) = (0, \infty)^3 \times (0, 1]$ and has Jacobian $yz/x =: J_2$. By this change of variables and Fubini's Theorem, (5.4) becomes

$$\pi_2(1) = 2 \int_0^1 t q(2, t) I(t) \, \mathrm{d}t, \tag{5.5}$$

where

$$I(t) = \int_0^\infty \int_0^\infty \int_0^\infty \frac{(yz)^3}{x} \times \exp\left\{-\left(x+y+z+\frac{tyz}{x}\right)\right\} \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z.$$

For $0 < t < 1$ the integral $I(t)$ can be obtained in the following closed form by invoking Mathematica:

$$I(t) = 12\big[3\ln(1/t)(1+t)(t^2+8t+1) - (1-t) \times (11t^2+38t+11)\big]/(1-t)^7, \quad (5.6)$$

while $q(2,\cdot)$ was computed in closed form in Hadjicostas (1998), as provided in (5.2). Substituting (5.2) and (5.6) into (5.5) and invoking Mathematica once more to calculate the resulting integral, we obtain $\pi_2(1) = 1/60$.

[Received January 2009. Revised June 2009.]

REFERENCES

Bickel, P. J., Hammel, E. A., and O'Connell, J. W. (1975), "Sex Bias in Graduate Admissions: Data From Berkeley," *Science*, 187 (4275), 398–404.

Blyth, C. R. (1972), "On Simpson's Paradox and the Sure-Thing Principle," *Journal of the American Statistical Association*, 67, 364–366.

Davison, A. C. (2003), *Statistical Models*, Cambridge, U.K.: Cambridge University Press.

Good, I. J., and Mittal, Y. (1987), "The Amalgamation and Geometry of Two-by-Two Contingency Tables," *The Annals of Statistics* 15 (2), 694–711.

Hadjicostas, P. (1997), "Copositive Matrices and Simpson's Paradox," *Linear Algebra and Its Applications*, 264, 475–488.

——— (1998), "The Asymptotic Proportion of Subdivisions of a 2 × 2 Table That Result in Simpson's Paradox," *Combinatorics, Probability and Computing*, 7 (4), 387–396. Corr.: 8 (1999), 599.

——— (2001), "Asymptotic Upper Bounds for the Proportion of Simpson Subdivisions of a 2 × 2 Table," *Communications in Statistics Simulation and Computation*, 30 (4), 1031–1051.

Hartigan, J. A. (1983), *Bayes Theory*, New York: Springer-Verlag.

Huh, M.-H. (1987), "Simpson's Paradox and Randomization," *Journal of the Korean Statistical Society*, 16 (2), 113–116.

Jeon, J. W., Chung, H. Y., and Bae, J. (1987), "Chances of Simpson's Paradox," *Journal of the Korean Statistical Society*, 16 (2), 117–127.

Julious, S. A., and Mullee, M. A. (1994), "Confounding and Simpson's Paradox," *British Medical Journal*, 209 (6967), 1480–1481.

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

Moore, D. S. (1991), *Statistics: Concepts and Controversies* (3rd ed.), New York: Freeman.

Pearson, K., Lee, A., and Bramley-Moore, L. (1899), "Genetic (Reproductive) Selection: Inheritance of Fertility in Man," *Philosophical Transactions of the Royal Statistical Society, Ser. A*, 173, 534–539.

Ross, K. (2004), *A Mathematician at the Ballpark: Odds and Probabilities for Baseball Fans*, New York: Pi Press.

Simpson, E. H. (1951), "The Interpretation of Interaction in Contingency Tables," *Journal of the Royal Statistical Society, Ser. B*, 13, 238–241.

Šleževičienė-Steuding, R., and Steuding, J. (2006), "Simpson's Paradox in the Farey Sequence," *Integers*, 6, A4, 9 pp. (electronic).

Wilks, S. S. (1962), *Mathematical Statistics*, New York: Wiley.

Yule, G. U. (1903), "Notes on the Theory of Association of Attributes in Statistics," *Biometrika*, 2, 121–134.