On Simpson's Paradox for Discrete Lifetime Distributions

Daniel Lebowitz
Department of Mathematics
Oregon State University

June 10, 2013

Abstract

In probability and statistics, Simpson's paradox is an apparent paradox in which a trend is present in different groups, but is reversed when the groups are combined. Joel Cohen (1986) has shown that continuously distributed lifetimes can never have a Simpson's paradox. We investigate the same question for discrete random variables to see if a Simpson's paradox is possible. With discrete random variables, we first look at those that have equally spaced values and show that Simpson's paradox does not occur. Next, when observing the discrete lifetimes that are unequally space with identical supports, we similarly discover that a Simpson's paradox still cannot occur. When the two random variables do not have identical supports, which allows for the flexibility to compare a broad range of different random variables, we discover that a Simpson's paradox can occur.

1 Halley's Life Table

In 1662, John Graunt developed one of the first life tables. He gathered his data from London's bills of mortality, but unfortunately the data lacked consistency, since there was no set structure of recording the births and deaths. Since Graunt's data was so unorganized and incomplete, he had no other choice but to guess all of the unknown entries. Another characteristic that made Graunt's data less desirable is that London's population growth was largely affected by migration, Halley (1693).

Casper Neumann (1648-1715), a German minister in Breslau, Silesia, had possession of complete records of births and the ages of deaths of people from Breslau from the years 1687-1691. Not only was the data accurate and complete, unlike the data Graunt used in 1662, but it also had several properties that made the data favorable for the uses of a life table. These characteristics included conclusions that the number of births and deaths were approximately equivalent, there was very little migration, and death rates for all age remained approximately constant. With these characteristics, it is fair to assume that Breslau had a near stationary population. Neumann sent these demographic records to Gottfried Leibniz, who then sent them to the Royal Society in London. The Royal Society asked Edmund Halley (1656-1742) to analyze the data. Halley published his analysis in 1693 in the *Philosophical Transactions*, Ciecka (2008).

Before discussing Halley's uses of a life table, it is necessary to define a few variables. For this we employ standard actuarial notation. The youngest age at which everyone in the population has died is denoted by ω, the population at age $x \in \{0, 1, 2, \ldots, \omega\}$ is l_x, and $L_x = 0.5(l_x + l_{x+1})$, Ciecka (2008).

The table on the next page was produced by Halley and represents the combined male and female survivors. Halley adjusted and smoothed the data, including $L_0 = 1000$, where the actual value from the Breslau data would have resulted in $L_0 = 0.5(1238 + 890) = 1064$. It seems that this specific rounded value was for the convenience of having a radix of 1000, Ciecka (2008).

Age x	L_{x-1}	Age x	L_{x-1}	Age x	L_{x-1}	Age x	L_{x-1}
1	1000	23	579	45	397	67	172
2	855	24	573	46	387	68	162
3	798	25	567	47	377	69	152
4	760	26	560	48	367	70	142
5	732	27	553	49	357	71	131
6	710	28	546	50	346	72	120
7	692	29	539	51	335	73	109
8	680	30	531	52	324	74	98
9	670	31	523	53	313	75	88
10	661	32	515	54	302	76	78
11	653	33	507	55	292	77	68
12	646	34	499	56	282	78	58
13	640	35	490	57	272	79	49
14	634	36	481	58	262	80	41
15	628	37	472	59	252	81	34
16	622	38	463	60	242	82	28
17	616	39	454	61	232	83	23
18	610	40	445	62	222	84	20
19	604	41	436	63	212		
20	598	42	427	64	202	85-100	107
21	592	43	417	65	192		
22	586	44	407	66	182	Total	34000

The following observations, and their explanations, were obtained by Halley and published in his classic papers, Ciecka (2008).

1. *”Proportion of men able to bear arms” (ages 18 to 56).*

Assuming that half of the population consists of men, Halley derived

$$\frac{L_{17} + L_{18} + \cdots + L_{55}}{2 * 34000} = 0.265$$

where the population of Breslau was 34,000.

2. *The odds of survival between two ages.*

This compares the number of deaths to the number that survived between two ages

$$\frac{L_{x+t}}{L_x - L_{x+t}}$$

Ex: The odds of surviving through the teenage years is

$$\frac{L_{19}}{L_{12} - L_{19}} = \frac{598}{640 - 598} = 598 : 42 = 14.2 : 1$$

3. *The age where it is an ”even wager” that a person at a current age would survive or die.*

Ex: If we were to look at a 52 year old and wanted to determine the age of survival needed to have an even wager, then we would observe that $L_{51} = 324$ and $L_{67} = 162 = L_{51}/2$. Therefore, an even wager for a 52 year old would be living an additional 16 years or reaching the age of 68.

4. *To regulate the price of term insurance.*

The price is determined by the odds of survival, which Halley observed the odds of survival of 1 year.

5. *Life annuity*

Halley used the following formula for a life annuity:

$$a_x = \sum_{t=1}^{\omega-x-1} (1+i)^{-t} \frac{L_{x+t-1}}{L_{x-1}}$$

This calculation determines the expected present value of a life annuity of one pound at a given present age and a fixed discount rate (Halley used 6%).

Ex: The expected present value of a life annuity of one pound for a 70 year old, based on a fixed discount rate of 6%, would be (using Halley's life table)

$$a_{70} = \sum_{t=1}^{15} (1.06)^{-t} \frac{L_{79+t}}{L_{79}} = 5.32$$

6. *Joint life annuity of two lives.*

Suppose that L_x represents lives at age x and L_y represents lives at age y. Furthermore, consider ${}_tD_x$ and ${}_tD_y$, which represents the number of deaths from L_x and L_y within t years, respectively. Therefore, $L_x = L_{x+t} +_t D_x$ and $L_y = L_{y+t} +_t D_y$. Halley was able to derive the probability of at least one life surviving as

$$1 - \frac{({}_tD_x)({}_tD_y)}{L_x L_y} = \frac{(L_{x+t})(L_{y+t}) + (L_{x+t})({}_tD_y) + (L_{y+t})({}_tD_y)}{L_x L_y}$$

Therefore, the expected present value of the life annuity that pays when at least one of the two individuals survives would be

$$a_{xy} = \sum_{t=1}^{\omega-x-1} (1+i)^{-t} \left(1 - \frac{({}_tD_x)({}_tD_y)}{L_x L_y}\right)$$

7. *Joint life annuity on three lives.*

Similarly to the joint life annuity on two lives, consider three lives at ages x, y, and z. Halley concluded that the value of a life annuity that pays when at least one of the three individuals is alive would be

$$a_{xyz} = \sum_{t=1}^{\omega-x-1} (1+i)^{-t} \left(1 - \frac{({}_tD_x)({}_tD_y)({}_tD_z)}{L_x L_y L_z}\right)$$

Also, consider an annuity that pays the youngest of the three individuals only after the older two die. Let $x = \min\{x, y, z\}$. Halley determined that the value of this annuity is

$$a_{yz|x} = \sum_{t=1}^{\omega-x-1} (1+i)^{-t} \left(\frac{(L_{x+t})({}_tD_y)({}_tD_z)}{L_x L_y L_z}\right)$$

There are a couple notable observations about Halley's work. First, life tables similar to Halley's are still being used this present day. Secondly, a major difference between Halley's calculations and the calculations used today is that Halley uses average survivors (L_x) rather than exact survivors (l_x), although some interpret Halley as using the exact survivors in his calculations. Regardless, it was fair to say that Edmond Halley was a pioneer in the development of survival analysis, Ciecka (2008). He is often referred to as the 'Father of Actuarial Science', Waymire (1988) and references therein.

2 Stochastic, Hazard Rate, and Relative Log-Concavity Orders

In this section we will look at some important concepts that are used throughout survival analysis. The definitions that will be introduced are going to be utilized throughout the remaining sections.

Definition 1. *Let L be a positive random variable representing the age at death of an individual. The **survival function** of L is defined as*

$$S_L(t) = P(L > t) = 1 - F_L(t)$$

Notice the density function, $f_L(t)$, when it exists is the negative derivative of the survival function.

Definition 2. *Let L be the age of death of a random individual being a continuous distribution with density $f_L(t)$. Then the **hazard rate** of L is defined as*

$$\lambda_L(t) = \lim_{\Delta \to 0} \frac{P(t < L \leq t + \Delta | L > t)}{\Delta} = \frac{f_L(t)}{1 - F_L(t)} = \frac{-S'_L(t)}{S_L(t)}$$

The following are standard forms of hazard rates in actuarial science

(a) $\lambda(t) = Bc^t$, where $B, c > 0$ and $0 \leq t < \infty$;

(b) $\lambda(t) = A/(B - t)^{c+1}$, where $A, B, c > 0$ and $t \leq B$;

(c) $\lambda(t) = H(t - B)^{c-1}$, where $c, H > 0$ and $t \geq B$.

Brillinger (1961) provides a theoretical basis using extreme value theory. Form (a) is known as Gompertz's Law, which shows the hazard rate increases geometrically over time, which is a sufficient model limited for specific time periods of an individual's life. Forms (b) and (c) are mentioned by Brillinger (1961) as being Makeham's First and Second Law, respectively. Although, others (H. L. Rietz) express Makeham's First and Second Law differently. Rietz (1986) expresses Makeham's First Law as $\lambda(t) = A + Bc^t$, which is identical to Gompertz's Law, but contains a term independent of t. Then with Makeham's Second Law, Rietz (1986) expresses it as $\lambda(t) = A + Dt + Bc^t$, which is identical to Makeham's First Law, but with a linear term.

Notice, one may either (a) fit continuous hazard rates to data, or (b) address the data explicitly with discrete hazard rates. Consider the lifetime of a random individual to be discretely measured, which aligns more closely to the data gathered in Halley's life table. This means the time periods until the death of any individual is discrete instead of continuous. If a lifetime is discretely measured, then with comparison to the definition above, notice $\Delta \in \mathbb{N}$, which is the distance between two consecutive values that this lifetime may obtain. Furthermore, if n_i

is the ith possible lifetime of an individual, then $\Delta = n_i - n_{i-1}$ and notice

$$\frac{P(n_{i-1} < L \leq n_{i-1} + \Delta | L > n_{i-1})}{\Delta} = \frac{P(n_{i-1} < L \leq n_i | L > n_{i-1})}{n_i - n_{i-1}}$$
$$= \frac{P(L = n_i | L \geq n_i)}{n_i - n_{i-1}}$$
$$= \frac{P(L = n_i)}{(n_i - n_{i-1}) P(L \geq n_i)}$$

Grimshaw et al. (2005) defines the hazard rate of a discrete random variable, where $n_i - n_{i-1} = 1$ for all i, and we will be considering $n_i - n_{i-1} \in \mathbb{N}$. However, we wish to also include the case of unequal spacing as follows.

Definition 3. *Let L be a positive random variable with possible values $\{n_1, n_2, \dots\}$. Then the **hazard rate** of L is defined as*

$$\lambda_L(n_i) = \frac{P(L = n_i | L \geq n_i)}{n_i - n_{i-1}}$$

where $n_{i-1}, n_i \in \mathbb{N}$, $n_{i-1} < n_i$ for all i, and $n_0 = 0$.

We will define hazard rate order, which will be used throughout sections 3, 4, and 5. Stochastic order will also be defined and we will see that these two orders will share relations when distributions have specific conditions.

Definition 4. *Let L_A and L_B be continuous random variables on $\mathbb{R}_+$ with density functions, or discrete random variables on $\mathbb{N}$ with mass functions, $f_{L_A}(t)$ and $f_{L_B}(t)$, respectively. Also, denote their respective cumulative distribution functions by $F_{L_A}(t)$ and $F_{L_B}(t)$.*

- *L_A is said to be smaller than L_B in the **stochastic order**, or $L_A \leq_{st} L_B$, if $S_{L_A}(t) \leq S_{L_B}(t)$ for all t.*
- *L_A is said to be smaller than L_B in the **hazard rate order**, or $L_A \leq_{hr} L_B$, if $f_{L_A}(t)/S_{L_A}(t) \geq f_{L_B}(t)/S_{L_B}(t)$ for all t.*

It can be shown that $\leq_{hr}$ implies $\leq_{st}$ for both discrete and continuous random variables. The discrete case can proven using Proposition 2 in Section 4.

Definition 5. *Let L be a continuous random variable on $\mathbb{R}_+$ with density function, or discrete random variables on $\mathbb{N}$ with mass function, $f_L(t)$. Then the **support** of L is defined as*

$$supp(L) = \{t : f_L(t) > 0\}$$

The inequality used to define log-concave in Definition 6 (below) is well known as Newton's inequality. The definition of log-concave is addressed by Yu (2010). Log-concavity order provides the ability to derive conditions that imply both stochastic ordering and hazard rate ordering as illustrated in Example 2.

Definition 6. *A non-negative sequence $u = \{u_j, j \geq 1\}$ is **log-concave**, or equivalently, $\log(u_j)$ is **concave** in $supp(u)$ if*

- *the $supp(u)$ is a set of consecutive integers in $\mathbb{N}$; and*
- $u_j^2 \geq u_{j+1}u_{j-1}$ *for all j.*

Definition 7. *Let L_A and L_B be continuous random variables on $\mathbb{R}_+$ with density functions, or discrete random variables on $\mathbb{N}$ with mass functions, $f_{L_A}(t)$ and $f_{L_B}(t)$, respectively. We say L_A is* ***log-concave relative*** *to L_B, denoted $L_A \leq_{lc} L_B$, if*

- *$supp(L_A)$ and $supp(L_B)$ are both intervals in $\mathbb{R}_+$ if L_A and L_B are continuous random variables or consecutive integers in $\mathbb{N}$ if L_A and L_B are discrete random variables;*
- *$supp(L_A) \subset supp(L_B)$;*
- $\log(f_{L_A}(t)/f_{L_B}(t))$ *is a concave function on $supp(L_A)$.*

Although all three orderings can be applied to both discrete and continuous random variables, the main purpose of this paper is to analyze the characteristics of discrete lifetimes, so we will emphasize the discrete portion of the definitions. Information on the continuous properties involved and more details of discrete random variables can found in Yu (2009).

Theorem 1 (Yu)**.** *Let random variables L_A and L_B have mass functions $f_{L_A}(t)$ and $f_{L_B}(t)$ respectively, both supported on the same $\mathbb{N}$ or $\{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$. Assume $L_A \leq_{lc} L_B$. Then $L_A \leq_{st} L_B$ and $L_A \leq_{hr} L_B$ are equivalent, and each holds if and only if $f_{L_A}(1)/f_{L_B}(1) \geq 1$.*

Example 1. Suppose $L_A \sim Geom(p_a)$ and $L_B \sim Geom(p_b)$. First there is a need to check when $L_A \leq_{lc} L_B$. Notice $supp(L_A) = supp(L_B) = \mathbb{N}$. By the definition of log-concave

$$\left(\frac{(1-p_a)^{n-1}p_a}{(1-p_b)^{n-1}p_b}\right)^2 \geq \left(\frac{(1-p_a)^n p_a}{(1-p_b)^n p_b}\right)\left(\frac{(1-p_a)^{n-2}p_a}{(1-p_b)^{n-2}p_b}\right)$$

for all $n \in \mathbb{N}$, which simplifies to $1 \geq 1$, so there are no restrictions to the values of p_a and p_b based on log-concavity. Consider, $L_A \leq_{st} L_B$. Notice, for all $n \in \mathbb{N}$

$$(1-p_a)^{n-1} = P(\text{first } n-1 \text{ are failures}) = P(L_A \geq n) \leq P(L_B \geq n) = (1-p_b)^{n-1}$$

Therefore $p_a \geq p_b$. Notice, for all $n \in \mathbb{N}$

$$\lambda_{L_A}(n) = \frac{P(L_A = n)}{P(L_A \geq n)} = \frac{(1-p_a)^{n-1}p_a}{(1-p_a)^{n-1}} = p_a$$

Since $p_a \geq p_b$, then $L_A \leq_{hr} L_B$ and it is easy to see that $P(L_A = 1)/P(L_B = 1) \geq 1$. Therefore, we can see that Theorem 1 is true with geometric random variables.

In the next example, Theorem 1 plays a more essential role in verifying hazard rate ordering.

Example 2. Suppose $L_A \sim Pois(\theta_a)$ and $L_B \sim Pois(\theta_b)$, where θ_a and θ_b are positive. First there is a need to check when $L_A \leq_{lc} L_B$. Notice $supp(L_A) = supp(L_B) = \mathbb{N}$. By the definition of log-concave

$$\left(\frac{\theta_a^{n-1}e^{-\theta_a}}{\theta_b^{n-1}e^{-\theta_b}}\right)^2 \geq \left(\frac{\theta_a^{n}e^{-\theta_a}}{\theta_b^{n}e^{-\theta_b}}\right)\left(\frac{\theta_a^{n-2}e^{-\theta_a}}{\theta_b^{n-2}e^{-\theta_b}}\right)$$

for all $n \in \mathbb{N}$, which simplifies to $1 \geq 1$, so there are no restrictions to the values of θ_a and θ_b based on log-concavity, similarly to Example 1. Notice, if we consider $P(L_A = 1)/P(L_B = 1) \geq 1$, which is equivalent to $e^{-\theta_a} \geq e^{-\theta_b}$, then $\theta_a \leq \theta_b$, since θ_a and θ_b are both positive.

In order to show that $L_A \leq_{st} L_B$ we need for all $n \in \mathbb{N}$

$$\sum_{k=1}^{n} \frac{\theta_a^{k-1} e^{-\theta_a}}{(k-1)!} \geq \sum_{k=1}^{n} \frac{\theta_b^{k-1} e^{-\theta_b}}{(k-1)!}$$

Since $\theta_a \leq \theta_b$, then take $\theta_b = \theta_a + \gamma$, where $\gamma \geq 0$. Therefore, the above inequality is equivalent to showing for all $\gamma \geq 0$

$$\begin{aligned} g(\gamma) &= \sum_{k=1}^{n} \frac{\theta_a^{k-1} e^{-\theta_a}}{(k-1)!} - \sum_{k=1}^{n} \frac{(\theta_a + \gamma)^{k-1} e^{-(\theta_a+\gamma)}}{(k-1)!} \\ &= e^{-\theta_a} \left(\sum_{k=1}^{n} \frac{\theta_a^{k-1}}{(k-1)!} - \sum_{k=1}^{n} \frac{(\theta_a + \gamma)^{k-1} e^{-\gamma}}{(k-1)!} \right) \geq 0 \end{aligned}$$

Notice, $g(0) = 0$, so if we show that $g'(\gamma) \geq 0$ for all $\gamma \geq 0$, then $g(\gamma) \geq 0$ for all $\gamma \geq 0$. Therefore for all $n \in \mathbb{N}$

$$\begin{aligned} g'(\gamma) &= e^{-\theta_a} \left(\sum_{k=1}^{n} \frac{e^{-\gamma}(\theta_a + \gamma)^{k-1}}{(k-1)!} - \sum_{k=2}^{n} \frac{e^{-\gamma}(k-1)(\theta_a + \gamma)^{k-2}}{(k-1)!} \right) \\ &= e^{-\theta_a} \left(\sum_{k=1}^{n} \frac{e^{-\gamma}(\theta_a + \gamma)^{k-1}}{(k-1)!} - \sum_{k=1}^{n-1} \frac{e^{-\gamma}(\theta_a + \gamma)^{k-1}}{(k-1)!} \right) \\ &= e^{-\theta_a} \frac{e^{-\gamma}(\theta_a + \gamma)^{n-1}}{(n-1)!} \\ &= e^{-\theta_b} \frac{(\theta_b)^{n-1}}{(n-1)!} > 0 \end{aligned}$$

Therefore, $L_A \leq_{st} L_B$.

In order to show that $L_A \leq_{hr} L_B$ we need for all $n \in \mathbb{N}$

$$\frac{\theta_a^n e^{-\theta_a}}{1 - \sum_{k=1}^{n-1} \frac{\theta_a^{k-1} e^{-\theta_a}}{(k-1)!}} \geq \frac{\theta_b^n e^{-\theta_b}}{1 - \sum_{k=1}^{n-1} \frac{\theta_b^{k-1} e^{-\theta_b}}{(k-1)!}}$$

To show this is true was nontrivial. Fortunately, since we have shown that when assuming $L_A \leq_{lc} L_B$ and $P(L_A = 1)/P(L_B = 1) \geq 1$, then the conditions result in $L_A \leq_{st} L_B$. Therefore, by applying Theorem 1, we know that $L_A \leq_{hr} L_B$.

3 Absence of Simpson's Paradox for Continuous Lifetime Distributions

In probability and statistics, Simpson's paradox is an apparent paradox in which a trend is present in different groups, but is reversed when the groups are combined. The following are well-known illustrations of Simpson's paradox.

Example 3. An instance where the Simpson's paradox may occur is with baseball players' batting averages. Consider the batting averages of two consecutive years of two fictional baseball players. Suppose that Mark was up to bat 100 times in 2011 and 2012 combined, during which he makes 28 hits for a proportion of 0.280. On the other hand, Sam was also up to bat 100 times in 2011 and 2012 combined, but had 29 hits for a proportion of 0.290. Notice in the table below, that Mark had a higher batting average than Sam when comparing each year individually.

	Mark	Sam
2011	**0.340 (17/50)**	0.320 (24/75)
2012	**0.220 (11/50)**	0.200 (5/25)
Total	0.280 (28/100)	**0.290 (29/100)**

Example 4. A real life example of an occurrence of Simpson's paradox was in a medical study involving the success rate of two kidney stone treatments. Treatment A includes all open procedures and treatment B was a process called percutaneous nephrolithomy. It was determined that treatment B was more successful then treatment A. When considering the variable of the size of the stone for each patient it was determined that treatment A was more successful for both small and large sized stones. The resulting data from the study can be seen in the table below.

	Treatment A	Treatment B
Small Stones	**93% (81/87)**	87% (234/270)
Large Stones	**73% (192/263)**	69% (55/80)
Total	78% (273/350)	**83% (289/350)**

In order to quantify Simpson's paradox in the context of actuarial lifetime data, Joel Cohen (1986) made the following observations.

Definition 8. *Let L be the age of death of a random individual, which is non-negative. The* ***crude death rate*** *of L is defined as*

$$d_L = \frac{1}{E[L]}$$

The following well-known relation between the hazard rate and survival probability will be used for the continuous case.

Theorem 2. *Let L be the age of death of a random individual, which is non-negative. Then*

$$S_L(t) = \exp\left(-\int_0^t \lambda_L(s)ds\right)$$

Proof. Suppose L is the age of death of a random individual, which is non-negative. Using Definition 2,

$$\lambda_L(t) = \frac{f_L(t)}{1 - F_L(t)}$$

$$= -\frac{d}{dt}\ln\left(1 - F_L(t)\right)$$

$$= -\frac{d}{dt}\ln S_L(t)$$

Then apply the antiderivative and obtain

$$S_L(t) = \exp\left(-\int_0^t \lambda_L(s)ds\right)$$

□

The following result is also well-known from probability theory.

Theorem 3. *Let L be the age of death of a random individual being a continuous distribution with density $f_L(t)$. Then*

$$E[L] = \int_0^\infty S_L(t)dt$$

Proof. Since the lifetime of an individual is strictly positive, the expectation of a random variable L would be obtained by multiplying t by the density $f_L(t)$ and integrating, so

$$E[L] = \int_0^\infty t f_L(t)dt$$

By integration by parts, recognizing that $-f_L(t)$ is the derivative of $S_L(t)$, $S_L(0) = 1$, $S_L(\infty) = 0$, and $\lim_{t\to\infty} tS_L(t) = 0$ notice

$$\begin{aligned} E[L] &= \int_0^\infty t f_L(t)dt \\ &= (-tS_L(t))|_0^\infty - \int_0^\infty -S_L(t)dt \\ &= \int_0^\infty S_L(t)dt \end{aligned}$$

□

Robert Parke asked if Simpsons paradox could happen when comparing age-specific (hazard rate) and crude death rates of two populations. This question was prompted due to the stratification of sex when determining the rates for annuities, when other factors, such as being Mormon, an individual's ethnicity, being a smoker, etc..., have a greater impact on mortality, but were being ignored when determining an individual's rates, Cohen (1986).

In the context of continuously distributed actuarial life data Cohen (1986) formulated Simpson's paradox as follows:

Definition 9. *Consider populations A and B with random individuals with lifetimes L_A and L_B, respectively. Let $\lambda_{L_A}(t) \geq \lambda_{L_B}(t)$ for all $t > 0$. A **Simpson's paradox** exists if $d_{L_A} < d_{L_B}$.*

Theorem 4 (Cohen)**.** *Consider populations A and B with random individuals with continuously distributed lifetimes L_A and L_B, respectively. If $\lambda_{L_A}(t) \geq \lambda_{L_B}(t)$, $\forall t > 0$, then $d_{L_A} > d_{L_B}$.*

Proof. Suppose L_A and L_B are the age of death of a random individual from populations A and B, respectively, such that $\lambda_{L_A}(t) > \lambda_{L_B}(t)$, $\forall t > 0$. This implies that

$$\int_0^t \lambda_{L_A}(s)ds > \int_0^t \lambda_{L_B}(s)ds$$

for all $t > 0$, then by Theorem 2

$$S_{L_A}(t) = \exp\left(-\int_0^t \lambda_{L_A}(s)ds\right) < \exp\left(-\int_0^t \lambda_{L_B}(s)ds\right) = S_{L_B}(t)$$

for all $t > 0$. Then by Theorem 3

$$E[L_A] = \int_0^\infty S_{L_A}(s)ds < \int_0^\infty S_{L_B}(s)ds = E[L_B]$$

and then by the definition of crude death rate

$$d_{L_A} = \frac{1}{E[L_A]} > \frac{1}{E[L_B]} = d_{L_B}$$

Therefore, a Simpson's paradox could not happen when comparing age-specific and crude death rates of two stationary populations.

□

Thus, one may conclude that if one uses the procedure of fitting a continuous hazard rate to data, then Simpson's paradox will not be an issue. However fitting a discrete hazard rate to data may introduce other issues.

4 Absence of Simpson's Paradox for Equally Spaced Discrete Lifetime Distributions and Identical Support

It has been established that Simpson's paradox cannot occur for continuous lifetime distributions. This still leaves the possibility of Simpson's paradox occurring when lifetimes are discretely measured, since Cohen's proof depended on properties that are present strictly for continuous distributions. When considering lifetimes that are discretely measured, each random variable considered will have support in $\mathbb{N} = \{1, 2, \ldots\}$. If a random variable L_0 has support in $\{0, 1, 2, \ldots\}$, consider the transformed random variable $L_1 = L_0 + 1$, which would have the desired support in $\mathbb{N}$. Notice, if we have discrete random variables L_A and L_B, where $supp(L_A) = supp(L_B)$ (identical supports), it could be easily shown that $\lambda_{L_A}(n) \geq \lambda_{L_B}(n)$ for all $n \in \mathbb{N}$ is equivalent to $L_A \leq_{hr} L_B$. The following examples are presented as warm up to the general theorem.

Example 5. A Simpson's paradox cannot occur when lifetimes are geometrically distributed.

Proof. Suppose that $L_A \sim Geom(p_a)$, $L_B \sim Geom(p_b)$, and $L_A \leq_{hr} L_B$. From Example 1 we determined that when $L_A \leq_{hr} L_B$, then $p_a \geq p_b$.

Therefore,

$$d_{L_A} = \frac{1}{E[L_A]} = \frac{1}{1/p_a} = p_a \geq p_b = \frac{1}{1/p_b} = \frac{1}{E[L_B]} = d_{L_B}$$

Hence, there is no Simpson's paradox when lifetimes are geometrically distributed.

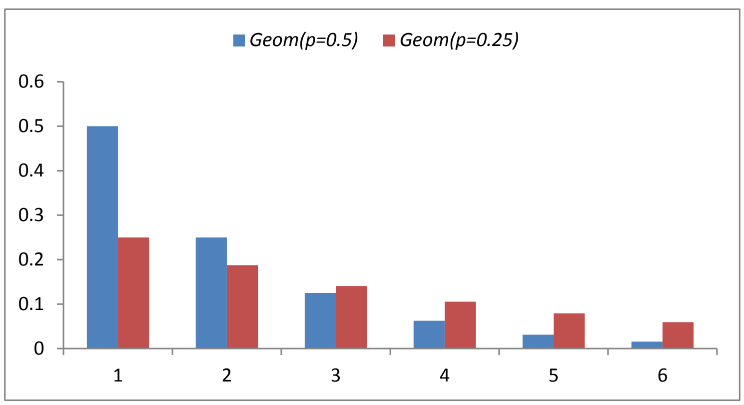
□

Consider $L_A \sim Geom(p_a)$ and $L_B \sim Geom(p_b)$, where $p_a > p_b$. If interested in determining for what values of n such that $P(L_A = n) > P(L_B = n)$, notice

$$(1-p_a)^{n-1}p_a > (1-p_b)^{n-1}p_b$$

$$\left(\frac{1-p_a}{1-p_b}\right)^{n-1} > \frac{p_b}{p_a}$$

Since $p_a > p_b$, it is evident that for at least $n = 1$ the inequality will hold. But it is inevitable that there exists $m \in \{2, 3, \dots\}$ such that for all $k \in \{m, m+1, \dots\}$ the inequality will not hold since the left side of the inequality is decaying exponentially. Therefore, if the inequality does not hold, then for all k $P(L_A = k) < P(L_B = k)$ (possibly at $k = m$ there may be equality). View the graph below, which exhibits the case for when $p_a = 0.5$ and $p_b = 0.25$.



Example 6. A Simpson's paradox cannot occur when lifetimes have a Poisson distribution.

Proof. Suppose that $L_A \sim Pois(\theta_a)$, $L_B \sim Pois(\theta_b)$, and $L_A \leq_{hr} L_B$. From Example 2 we determined that when $L_A \leq_{hr} L_B$, then $\theta_a \leq \theta_b$. So,

$$d_{L_A} = \frac{1}{E[L_A]} = \frac{1}{\theta_a + 1} \geq \frac{1}{\theta_b + 1} = \frac{1}{E[L_B]} = d_{L_B}$$

Hence, no Simpson's paradox can occur when lifetimes have a Poisson distribution.

□

Example 7. A Simpson's paradox cannot occur when lifetimes have a Bernoulli distribution.

Proof. Suppose that L_A is a Bernoulli random variable such that $P(L_A = 1) = p_a$ and $P(L_A = 2) = 1 - p_a$, and L_B is a Bernoulli random variable such that $P(L_B = 1) = p_b$ and $P(L_B = 2) = 1 - p_b$. Also, suppose $L_A \leq_{hr} L_B$. Notice

$$p_a = \frac{P(L_A = 1)}{P(L_A \geq 1)} = \lambda_{L_A}(1) \geq \lambda_{L_B}(1) = \frac{P(L_B = 1)}{P(L_B \geq 1)} = p_b$$

Also,

$$E[L_A] = p_a + 2(1 - p_a) = 2 - p_a$$

$$E[L_B] = 2 - p_b$$

$$d_{L_A} = \frac{1}{E[L_A]} = \frac{1}{2 - p_a} \geq \frac{1}{2 - p_b} = \frac{1}{E[L_B]} = d_{L_B}$$

Hence, no Simpson's paradox can occur when lifetimes have a Bernoulli distribution.

□

Example 8. A Simpson's paradox cannot occur when lifetimes have support $\{1, 2, 3\}$.

Proof. Suppose there are random variables L_A and L_B such that $P(L_A = j) = a_j$, $P(L_B = j) = b_j$, for $j \in \{1, 2, 3\}$, where $a_1 + a_2 + a_3 = b_1 + b_2 + b_3 = 1$. Furthermore, consider $L_A \leq_{hr} L_B$. For purposes of contradiction suppose there can be a Simpson's paradox. So $d_{L_A} < d_{L_B}$, which is equivalent to $E[L_A] > E[L_B]$.

Since $L_A \leq_{hr} L_B$, then

$$a_1 \geq b_1 \tag{1}$$

$$a_2 b_3 \geq a_3 b_2 \tag{2}$$

Since $E[L_A] > E[L_B]$, then $a_1 + 2a_2 + 3a_3 > b_1 + 2b_2 + 3b_3$. Therefore

$$a_1 + 2(1 - a_1 - a_3) + 3a_3$$

$$\implies -a_1 + a_3 > -b_1 + b_3$$

From (1) we then obtain $a_3 > b_3$.

Then, by (2) we know that $a_2 > b_2$, but we know that this would lead to a contradiction since $a_1 + a_2 + a_3 = b_1 + b_2 + b_3$. Therefore, there is no Simpson's paradox when lifetimes have support $\{1, 2, 3\}$.

□

Proposition 1. *A Simpson's paradox cannot occur when lifetimes have support* $\{1, 2, \ldots, n\}$.

Proof. Suppose there are random variables L_A and L_B such that $P(L_A = j) = a_j$ and $P(L_B = j) = b_j$, for $j \in \{1, 2, \ldots, n\}$, where $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i = 1$. Furthermore, consider $L_A \leq_{hr} L_B$. For purposes of contradiction suppose there can be a Simpson's paradox. So $d_{L_A} < d_{L_B}$, which is equivalent to $E[L_A] > E[L_B]$. Notice for $k \in \{2, \ldots, n-1\}$

$$\frac{a_k}{a_k + \cdots + a_n} = P(L_A = k | L_A \geq k) = \lambda_{L_A}(k) \geq \lambda_{L_B}(k) = P(L_B = k | L_B \geq k) = \frac{b_k}{b_k + \cdots + b_n}$$

$$\iff a_k(b_k + \cdots + b_n) \geq b_k(a_k + \cdots + a_n)$$

$$\iff a_k b_k + a_k(b_{k+1} + \cdots + b_n) \geq a_k b_k + b_k(a_{k+1} + \cdots + a_n)$$

$$\iff a_k(b_{k+1} + \cdots + b_n) \geq b_k(a_{k+1} + \cdots + a_n)$$

Therefore,

$$a_1 \geq b_1 \tag{3}$$

$$a_2(b_3 + ... + b_n) \geq b_2(a_3 + ... + a_n) \tag{4}$$

$$a_3(b_4 + ... + b_n) \geq b_3(a_4 + ... + a_n) \tag{5}$$

$$\vdots$$

$$a_{n-2}(b_{n-1} + b_n) \geq b_{n-2}(a_{n-1} + a_n) \tag{6}$$

$$a_{n-1}b_n \geq b_{n-1}a_n \tag{7}$$

$$\sum_{i=1}^{n} i a_i > \sum_{i=1}^{n} i b_i \tag{8}$$

$$a_i > 0, b_i > 0, \forall i \in \{1, \ldots, n\} \tag{9}$$

$$\sum_{i=1}^{n} a_i = 1 = \sum_{i=1}^{n} b_i \tag{10}$$

Notice, either $a_n \geq b_n$ or $a_n < b_n$.

Case 1: Suppose $a_n \geq b_n$. Then by (7), $a_{n-1} \geq b_{n-1}$. Since $a_n \geq b_n$ and $a_{n-1} \geq b_{n-1}$, then by (6), $a_{n-2} \geq b_{n-2}$. Continuing this process we see that $a_i \geq b_i, \forall i \in \{1, \ldots, n\}$. By (10), we then obtain $a_i = b_i, \forall i \in \{1, \ldots, n\}$. But, this is a contradiction by (8).

Case 2: Suppose $a_n < b_n$. Notice, either $a_{n-1} + a_n \geq b_{n-1} + b_n$ or $a_{n-1} + a_n < b_{n-1} + b_n$.

Subcase 1: Suppose $a_{n-1} + a_n \geq b_{n-1} + b_n$. By $a_n < b_n$, we have $a_{n-1} > b_{n-1}$. Also by (6), $a_{n-2} \geq b_{n-2}$. Similarly to the process from Case 1, we see that $a_i \geq b_i, \forall i \in \{1, \ldots, n-2\}$. Also, since $a_{n-1} + a_n \geq b_{n-1} + b_n$, then by (10) we obtain $a_i = b_i, \forall i \in \{1, \ldots, n\}$. But, this is a contradiction by (8).

Subcase 2: Suppose $a_{n-1} + a_n < b_{n-1} + b_n$. Notice, $a_{n-2} + a_{n-1} + a_n \geq b_{n-2} + b_{n-1} + b_n$ or $a_{n-2} + a_{n-1} + a_n < b_{n-2} + b_{n-1} + b_n$.

Subsubcase 3: Suppose $a_{n-2} + a_{n-1} + a_n \geq b_{n-2} + b_{n-1} + b_n$. Similarly to Subcase 1, we can show that $a_i \geq b_i, \forall i \in \{1, \ldots, n-3\}$ and obtain a similar contradiction.

So, continuing this process we would then obtain that the only remaining case that has not been contradicted is when

$$a_n < b_n$$

$$a_{n-1} + a_n < b_{n-1} + b_n$$

$$\vdots$$

$$a_4 + \cdots + a_n < b_4 + \cdots + b_n$$

$$a_3 + \cdots + a_n < b_3 + \cdots + b_n$$

Summing the left sides of each inequality and then the right sides of each inequality the following is obtained

$$\sum_{j=3}^{n} (j-2)a_i < \sum_{j=3}^{n} (j-2)b_j \tag{11}$$

We can see by (8), $a_2 = 1 - (a_1 + a_3 + \cdots + a_n)$ and $b_2 = 1 - (b_1 + b_3 + \cdots + b_n)$, so

$$-a_1 + \sum_{j=3}^{n} (j-2)a_i > -b_1 + \sum_{j=3}^{n} (j-2)b_j$$

But, by (3) we then obtain

$$\sum_{j=3}^{n} (j-2)a_i > \sum_{j=3}^{n} (j-2)b_j$$

which is a contradiction by (11). Therefore, a Simpson's paradox cannot occur when lifetimes have support $\{1, 2, \ldots, n\}$.

□

With the process used to prove the proposition above, there was a dependency on the largest value of the support. So, this process cannot be utilized to show that discrete lifetimes with infinite support doesn't have a Simpson's paradox. It is evident that the conclusion whether Simpson's paradox may occur is determined by the magnitude of the expectation of each lifetime distribution. Since what is known from the beginning is the behavior of the hazard rates, then it would be convenient to express the probabilities of the lifetime distributions in terms of the hazard rates.

Proposition 2. *Let L be a discrete random variable with a support contained in $\mathbb{N}$. Then $\forall n \in \mathbb{N}$*

$$P(L \geq n) = \prod_{j=1}^{n-1} (1 - \lambda_L(j))$$

Proof. Suppose that L is a discrete random variable with a support contained in $\mathbb{N}$. Then $\forall n \in \mathbb{N}$

$$1 - \lambda_L(n-1) = 1 - \frac{P(L = n-1)}{P(L \geq n-1)} = \frac{P(L \geq n)}{P(L \geq n-1)}$$

$$\begin{aligned} P(L \geq n) &= P(L \geq n-1)(1 - \lambda_L(n-1)) \\ &= P(L \geq n-2)(1 - \lambda_L(n-2))(1 - \lambda_L(n-1)) \\ &\vdots \\ &= P(L \geq 1) \prod_{j=1}^{n-1} (1 - \lambda_L(j)) \\ &= \prod_{j=1}^{n-1} (1 - \lambda_L(j)) \end{aligned}$$

□

Now that the probabilities of the lifetime distributions can be expressed in terms of the hazard rates, this tool can be utilized to prove the general theorem for this section.

Theorem 5. *A Simpson's paradox cannot occur when lifetimes have support* $\mathbb{N}$.

Proof. Suppose there are random variables L_A and L_B such that $P(L_A = j) = a_j$ and $P(L_B = j) = b_j$, for $\forall j \in \mathbb{N}$, where $\sum_{i=1}^{\infty} a_i = \sum_{i=1}^{\infty} b_i = 1$. Furthermore, suppose $L_A \leq_{hr} L_B$, which is equivalent to $1 - \lambda_{L_A}(n) \leq 1 - \lambda_{L_B}(n)$, $\forall n \in \mathbb{N}$. Then by Proposition 2, $\forall n \in \mathbb{N}$

$$P(L_A \geq n) = \prod_{j=1}^{n-1}(1 - \lambda_{L_A}(j)) \leq \prod_{j=1}^{n-1}(1 - \lambda_{L_B}(j)) = P(L_B \geq n)$$

Therefore, $\forall n \in \mathbb{N}$

$$\sum_{j=n}^{\infty} a_j \leq \sum_{j=n}^{\infty} b_j$$

which then implies

$$E[L_A] = \sum_{k=1}^{\infty}\left[\sum_{i=k}^{\infty} a_i\right] \leq \sum_{k=1}^{\infty}\left[\sum_{i=k}^{\infty} b_i\right] = E[L_B]$$

which depends on $\sum_{i=1}^{\infty} a_i = \sum_{i=1}^{\infty} b_i$. Therefore, concluding that $d_{L_A} \geq d_{L_B}$. Therefore, a Simpson's paradox cannot occur when lifetimes have support $\mathbb{N}$.

$\square$

It can be easily shown that Proposition 2 can be utilized to prove Example 7, Example 8, and Proposition 1. Furthermore, Proposition 2 will also be used to show that the lifetimes of any pair of random variables with identical support can never have a Simpson's paradox as seen in the following section.

5 Simpson's Paradox for Unequally Spaced Discrete Lifetime Distributions

At this point we have only considered modeling lifetimes to discrete distributions, where the supports are equally spaced. With this support structure we have shown that a Simpson's paradox is not possible. The next objective is to consider modeling the lifetimes to discrete distributions, where the supports are unequally spaced. When the supports are unequally spaced, it is possible to compare random variables where the supports are not identical. For the meantime, we will consider random variables with unequally spaced supports, which are identical.

Proposition 3. *A Simpson's paradox cannot occur when lifetimes have identical finite support.*

Proof. Suppose there are random variables L_A and L_B such that $P(L_A = n_j) = a_j$, $P(L_B = n_j) = b_j$, and $n_j \in \mathbb{N}$, $\forall j \in \{1, 2, \ldots, N\}$, where $n_j < n_{j+1}$ $\forall j \in \{1, 2, \ldots, N-1\}$ and $\sum_{i=1}^{N} a_i = \sum_{i=1}^{N} b_i = 1$.

Furthermore, suppose $L_A \leq_{hr} L_B$, which is equivalent to $1 - \lambda_{L_A}(n) \leq 1 - \lambda_{L_B}(n)$, $\forall n \in \mathbb{N}$. Then by Property 1

$$P(L_A \geq n) = \prod_{j=1}^{n-1}(1 - \lambda_{L_A}(j)) \leq \prod_{j=1}^{n-1}(1 - \lambda_{L_B}(j)) = P(L_B \geq n)$$

$\forall n \in \mathbb{N}$. Therefore, $\forall j \in \{1, 2, \ldots, N\}$

$$\sum_{k=j}^{N} a_{n_k} \leq \sum_{k=j}^{N} b_{n_k}$$

which then implies

$$E[L_A] = \sum_{k=1}^{N}\left[(n_k - n_{k-1})\sum_{i=k}^{N} a_{n_i}\right] \leq \sum_{k=1}^{N}\left[(n_k - n_{k-1})\sum_{i=k}^{N} b_{n_i}\right] = E[L_B]$$

where $n_0 = 0$, and depends on the fact $n_j < n_{j+1}$ $\forall j \in \{1, 2, \ldots, N-1\}$ and $\sum_{i=1}^{N} a_i = \sum_{i=1}^{N} b_i$. Hence, $d_{L_A} \geq d_{L_B}$. Therefore, a Simpson's paradox cannot occur when lifetimes have identical finite support.

□

As mentioned before, Proposition 2 will be used to show that unequally spaced lifetime distributions with identical support cannot have a Simpson's paradox, as concluded in the next proof.

Theorem 6. *A Simpson's paradox cannot occur when lifetimes have identical support.*

Proof. Suppose there are random variables L_A and L_B such that $P(L_A = n_j) = a_j$, $P(L_B = n_j) = b_j$, and $n_j \in \mathbb{N}$, $\forall j \in \mathbb{N}$, where $n_j < n_{j+1}$ $\forall j \in \mathbb{N}$ and $\sum_{i=1}^{\infty} a_i = \sum_{i=1}^{\infty} b_i = 1$. Furthermore, suppose $L_A \leq_{hr} L_B$, which is equivalent to $1 - \lambda_{L_A}(n) \leq 1 - \lambda_{L_B}(n)$, $\forall n \in \mathbb{N}$. Then by Property 1

$$P(L_A \geq n) = \prod_{j=1}^{n-1}(1 - \lambda_{L_A}(j)) \leq \prod_{j=1}^{n-1}(1 - \lambda_{L_B}(j)) = P(L_B \geq n)$$

$\forall n \in \mathbb{N}$. Therefore, $\forall j \in \mathbb{N}$

$$\sum_{k=j}^{\infty} a_{n_k} \leq \sum_{k=j}^{\infty} b_{n_k}$$

which then implies that

$$E[L_A] = \sum_{k=1}^{\infty}\left[(n_k - n_{k-1})\sum_{i=k}^{\infty} a_{n_i}\right] \leq \sum_{k=1}^{\infty}\left[(n_k - n_{k-1})\sum_{i=k}^{\infty} b_{n_i}\right] = E[L_B]$$

where $n_0 = 0$, and depends on the fact $n_j < n_{j+1}$ $\forall j \in \mathbb{N}$ and $\sum_{i=1}^{\infty} a_i = \sum_{i=1}^{\infty} b_i$. Hence, $d_{L_A} \geq d_{L_B}$. Therefore, a Simpson's paradox cannot occur when lifetimes have identical support.

□

Since it has been shown that a Simpson's paradox cannot occur when lifetimes have identical support, the next lifetime distribution to consider would be when the supports are not identical. When discrete random variables do not have identical support, this allows for the flexibility to compare a broader range of lifetime distributions. Notice that this freedom will allow a Simpson's paradox to occur, but there are restrictions to the the lifetime distributions (L_A and L_B) so that $\lambda_{L_A}(n) \geq \lambda_{L_B}(n)$ for all $n \in \mathbb{N}$ holds.

Lemma 1. *Let L_A and L_B be discrete lifetime distributions. If $\lambda_{L_A}(n) \geq \lambda_{L_B}(n)$ for all $n \in \mathbb{N}$, then $supp(L_B) \subseteq supp(L_A)$.*

Proof. Suppose that $\lambda_{L_A}(n) \geq \lambda_{L_B}(n)$ for all $n \in \mathbb{N}$. For purposes of contradiction, consider there exists a value $n_i \in supp(L_B)$ and $n_i \notin supp(L_A)$, where n_i is the ith possible lifetime of L_B. But we can see that $\lambda_{L_A}(n_i) = 0 < \frac{P[L_B=n_i|L_B \geq n_i]}{n_i - n_{i-1}} = \lambda_{L_B}(n_i)$, which is a contradiction. Therefore, Lemma 1 is true.

□

Considering Lemma 1, the most trivial case to observe where lifetimes do no have identical support is when one random variable may obtain the values $1, 2, 3$ and the other random variable may either obtain the values (a) $1, 3$, or (b) $2, 3$. It can be shown that a Simpson's paradox cannot occur when the other random variable can obtain the value $1, 3$, but it will be shown for a more general case. That is, when $n_1 < n_2 < n_3$, a Simpson's paradox cannot occur when one random variable may obtain n_1, n_2, n_3 and the other random variable may obtain n_1, n_3.

Example 9. Let populations A and B be random individuals with lifetimes L_A and L_B, respectively. If $supp(L_A) = \{n_1, n_2, n_3\}$ and $supp(L_B) = \{n_1, n_3\}$, where $n_1 < n_2 < n_3$, then a Simpson's paradox cannot occur between the lifetimes of populations A and B.

Proof. Suppose there is a random variable L_A and L_B, where $supp(L_A) = \{n_1, n_2, n_3\}$ and $supp(L_B) = \{n_1, n_3\}$, where $n_1 < n_2 < n_3$. By Lemma 1, we know that $\lambda_{L_A}(n) \leq \lambda_{L_B}(n)$ for all $n \in \mathbb{N}$ cannot occur. Suppose $P(L_A = n_1) = a_1$, $P(L_A = n_2) = a_2$, $P(L_A = n_3) = a_3 = 1 - a_1 - a_2$, and let $P(L_B = n_1) = b_1$ and $P(L_B = n_3) = b_3 = 1 - b_1$. Suppose $\lambda_{L_A}(n) \geq \lambda_{L_B}(n)$ for all $n \in \mathbb{N}$. For purposes of contradiction, consider that $d_{L_A} < d_{L_B}$, which is equivalent to $E[L_A] > E[L_B]$. Notice, we obtain the following inequalities

$$a_1 \geq b_1 \tag{12}$$

$$a_2 \geq 0 \tag{13}$$

$$\frac{1}{n_3 - n_2} \geq \frac{1}{n_3 - n_1}$$

$$n_1 a_1 + n_2 a_2 + n_3 a_3 > n_1 b_1 + n_3 b_3 \tag{14}$$

Notice, from (14), $a_3 = 1 - a_1 - a_2$, and $b_3 = 1 - b_1$, we obtain $(n_3 - n_1)a_1 + (n_3 - n_2)a_2 < (n_3 - n_1)b_1$. But, we see from (13) and then (12) that $(n_3 - n_1)a_1 + (n_3 - n_2)a_2 \geq (n_3 - n_1)a_1 \geq (n_3 - n_1)b_1$, which is a contradiction. Therefore, a Simpson's paradox cannot occur for this example.

□

Example 10. Suppose populations A and B have random individuals with lifetimes L_A and L_B, respectively. Let $P(L_A = 1) = 0.1$, $P(L_A = 2) = 0.4$, $P(L_A = 3) = 0.5$, and let $P(L_B = 2) = 0.75$ and $P(L_B = 3) = 0.25$. Notice, when comparing the hazard rates

$$\lambda_{L_A}(1) = a_1 = 0.1 > 0 = \lambda_{L_B}(1)$$

$$\lambda_{L_A}(2) = a_2/(a_2 + a_3) = 4/9 > 3/8 = a_2/2 = \lambda_{L_B}(2)$$

$$\lambda_{L_A}(3) = \lambda_{L_B}(3)$$

Therefore, $\lambda_{L_A}(n) \geq \lambda_{L_B}(n)$ for all $n \in \mathbb{N}$. But, we can see when comparing the crude death rates

$$d_{L_A} = 1/E[L_A] = 1/2.4 < 1/2.25 = 1/E[L_B] = d_{L_B}$$

Therefore, a Simpson's paradox is possible!

To summarize, it had already been established by Cohen (1986) that continuously distributed lifetimes can never have a Simpson's paradox. Furthermore, it has been shown that discrete lifetimes with identical supports cannot have a Simpson's paradox, but when the discrete lifetimes do not have identical supports a Simpson's paradox may occur. Simpson's paradox is an arithmetic effect when it occurs. However it requires more than just discreteness of the lifetime distributions to appear. The reason Simpson's paradox was able to occur is because the different distances between the random variables applied a weighted factor to the hazard rates, which did not have any effect on the crude death rate. Notice, a weighted factor can never exist when the discrete lifetimes have identical supports. If considering the reason why a Simpson's paradox was able to occur in Examples 3 and 4, we would arrive to a similar conclusion of why one may occur with discrete lifetimes that do not have identical supports. It appears that the only way a Simpson's paradox may occur is if there is a weighted factor between the groups (or hazard rates when considering lifetime data).

References

David R. Brillinger. A Justification of Some Common Laws of Mortality. *Transactions of Society of Actuaries*, 13(36AB):116–118, 1961.

James E. Ciecka. Edmond Halley's Life Table and Its Uses. *Journal of Legal Economics*, 15:65–74, 2008.

Joel E. Cohen. An Uncertainty Principle in Demography and the Unisex Issue. *The American Statistician*, 40 (1):7, 1986.

S. Grimshaw, J. McDonald, G. McQueen, and S. Thorley. Estimating Hazard Functions for Discrete Lifetimes. *Communications is Statistics - Simulation and Computation*, 34(2):451–463, 2005.

Edmund Halley. An Estimate of the Degrees of Mortality of Mankind, Drawn from the Curious Tables of the Births and Funerals at the City Breslaw, with an Attempt to Ascertain the Price of Annuities upon Lives. *Philosophical Transactions*, 17:596–610, 1693.

H. L. Rietz. An Uncertainty Principle in Demography and the Unisex Issue. *The American Mathematical Monthly*, 28(4):158–165, 1986.

Elizabeth L. Turner. Cultural imagery and statistical models of the force of mortality: Addison, Gompertz and Pearson. *Journal of the Royal Statistical Society*, pages 1–17, 2010.

E. Waymire. Halley: A Name that Lives on in Mortality. *Physics Today*, 41:157, 1988. doi: 10.1063/1.2811654.

Ward Whitt. Uniform Conditional Variability Ordering of Probability Distributions. *Journal of Applied Probability*, 22(3):619–633, 1985.

Yaming Yu. Stochastic Ordering of Exponential Family Distributions and Their Mixtures. *Journal of Applied Probability*, pages 1–4, 2009.

Yaming Yu. Relative log-concavity and a pair of triangle inequalities. *Bernoulli*, 16:1, 2010.