APMA 1360 Final Project

JOHN ENSLEY

Brown University

I. Building the Model

It has been observed that roughly 10% of the human population is left-handed. This paper seeks to develop a model that explains this handedness asymmetry we see at the species level [1].

First we define a function $P_{RL}(l)$, which represents the probability that, for the current proportion of left-handers l in a population, a right-handed person is replaced by left-handed offspring over a long period of time. Likewise $P_{LR}(l)$ is the probability a left-handed person is replaced by right-handed offspring. To model the change in l over time, consider the differential equation

$$\frac{\mathrm{d}l}{\mathrm{d}t} = (1-l)P_{RL}(l) - lP_{LR}(l)$$

Observe that the first term is the societal fraction of right-handers multiplied by the probability that right-handers are replaced by left-handers. This gives a growth in the proportion of lefties, and therefore an increase in $\mathrm{d}l/\mathrm{d}t$. Similarly, the second term is the societal fraction of left-handers multiplied by the probability that left-handers are replaced by right-handers. This results in fewer lefties and a decrease in $\mathrm{d}l/\mathrm{d}t$, hence the minus sign.

To simplify the model a bit, we assume that these functions are symmetric, that is $P_{LR}(l) = P_{RL}(1-l)$. The reasoning behind this claim is that there is no inherent difference in fitness between being left-handed and right-handed. The human species could feasibly have ended up at an equilibrium of 90% left-handed instead of 90% right-handed. By making this assumption we can write

$$\frac{\mathrm{d}l}{\mathrm{d}t} = (1-l)P_{RL}(l) - lP_{RL}(1-l)$$

Now we break up the function $P_{RL}(l)$ into two components, one increasing and one decreasing. The increasing part, named $P_{RL}^{coop}(l)$, dominates in cooperative societies. In a totally cooperative society, the minority handedness would die out eventually, since individuals that did not conform would be excluded from group activities and such.

The decreasing part, $P_{RL}^{comp}(l)$, dominates in competitive societies. If left-handed people are rare, they will often fare better in direct competitions with right-handers since righties would not have as much experience against them. This component drives the population toward an equal split between left and right-handers.

Defining a constant $0 \leq c \leq 1$ to be the importance of cooperation in interactions, we can write the probability function as

$$P_{RL}(l) = cP_{RL}^{coop}(l) + (1-c)P_{RL}^{comp}(l)$$

We can show that depending on the shape of the component functions and the value of c, there could be one, three, or five fixed points l^*.

I first chose the following component functions

$$P_{RL}^{coop}(l) = \frac{1}{1+e^{-(10x-5)}}$$

$$P_{RL}^{comp}(l) = 1 - \frac{1}{1+e^{-(10x-5)}}$$

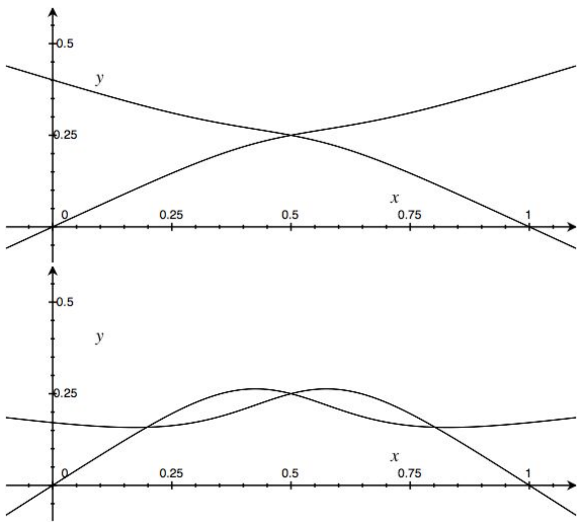

Figure 1: *Plots of $(1-l)P_{RL}(l)$ and $lP_{RL}(1-l)$. Top: c less than critical value. Bottom: c greater than critical value.*

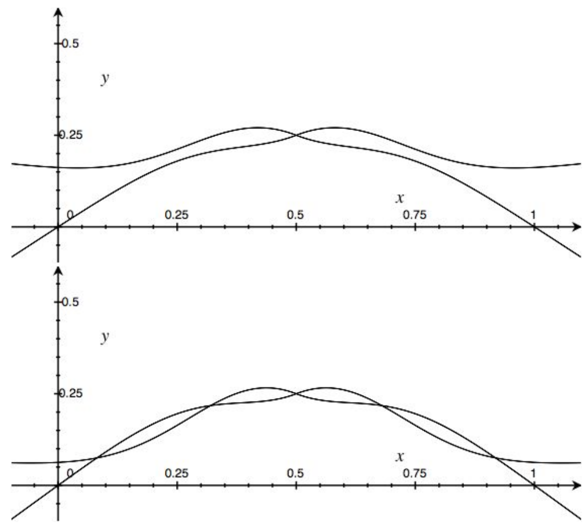

Figure 2: *Plots of $(1-l)P_{RL}(l)$ and $lP_{RL}(1-l)$. Top: c less than critical value. Bottom: c greater than critical value.*

These are sigmoidal in shape, with minima near 0 and maxima near 1 on the interval $[0, 1]$. The bifurcation behavior with these functions will be the same as it would had I chosen other similar functions, the only difference being the exact value of c at which bifurcations occur. Figure 1 shows that as c increases, there exists a supercritical pitchfork bifurcation, going from 1 fixed point to 3 at some critical value of c. This is reflected in the bottom of Figure 3, which was taken from the original article.

It is also possible for the component functions to have a double sigmoidal shape. To see if this would affect the bifurcation behavior, I chose

$$P_{RL}^{coop}(l) = \frac{1}{2}\left(\operatorname{sgn}\left(x - \frac{1}{2}\right)\left(1 - e^{-16(x-\frac{1}{2})^2}\right) + 1\right)$$

$$P_{RL}^{comp}(l) = 1 - P_{RL}^{coop}(l)$$

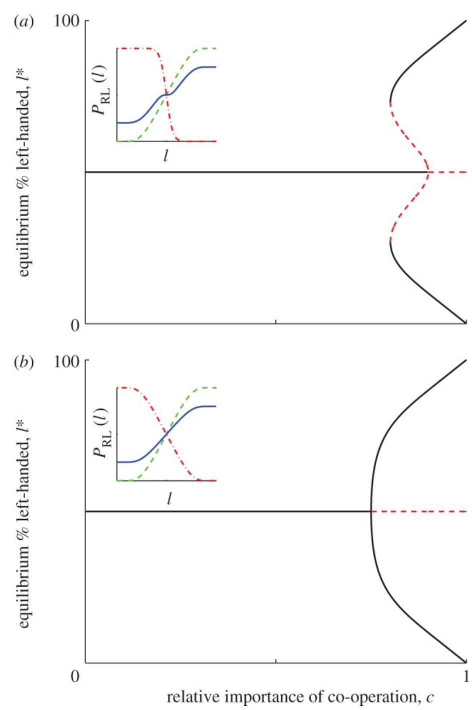

Figure 3: *Equilibrium left-handed percentage as a function of c. Top is for double-sigmoid component functions, bottom is for sigmoid component functions. Solid line represents stable equilibrium, dashed represents unstable equilibrium.*

As before, these have minima near 0 and maxima near 1 on $[0,1]$. Figure 2 shows that as c increases, there exist two saddle point bifurcations at some critical value of c, beyond which there are 5 fixed points instead of 1. This behavior can be seen in the top of Figure 3.

II. Comparing Model to Baseball Data

There is a well-known problem that arises when attempting to model left- or right-handedness in a species population. There is rarely enough consistent empirical data to validate the model, because the constant of cooperation c is hard to quantify. To get around this problem, we focus on athletics. Data on both handedness and cooperation is often much easier to gather in sports than for society as a whole.

The selection process by which athletes become professionals differs from natural selection in that there is a background rate of left-handedness l_{bg} that should remain unaffected. Therefore this process must be modeled carefully.

Let $s \sim N(0,\sigma^2)$ be a random variable that represents a person's skill in a particular sport. The selection process essentially involves choosing the n most skilled athletes from a population of N people who play the sport in question. The derived model gives an expression for l_{pro}, the proportion of professional athletes in the sport who are left-handed.

$$l_{pro} = \frac{1}{2}\frac{l_{bg}\mathrm{erfc}(\hat{s}_c - \Delta\hat{s})}{n/N}$$

Here $\hat{s}_c$ is the skill cutoff. Players with skill $s < \hat{s}_c$ are not selected to be professionals. Also, $\Delta\hat{s}$ is the skill advantage for left-handers. This is proportional to $l^* - l_{pro}$; in other words, the greater the gap between the current lefty fraction and the "optimal" one (the fixed point from our earlier model), the greater the advantage for lefties.

In the original paper, this model was applied to data from various sports such as golf, boxing, baseball, and table tennis. The predicted proportion of left-handers matched the observed proportion fairly well in these cases.

Focusing on baseball, we can further test the selection model by using it to predict the fraction of lefties as a function of rank. Defining l_r as the fraction of left-handers that finished a season ranked in the top r of all hitters, the model can be tweaked to give

$$l_r = \frac{1}{2}\frac{l_{bg}N\mathrm{erfc}(\hat{s}_r - \Delta\hat{s})}{r}$$

This has one free parameter, $\Delta\hat{s}$. The rest come from historical data [2]. We can fit this curve to the observed data quite well, as shown in Figure 4. The best fit occurs at $\Delta\hat{s} = 0.3003$.

In the original paper, the statistic used to determine rank was the number of hits in each season. However, as any baseball fan can attest, hits are not necessarily the best indicator of the quality of a hitter. Suppose in a particular year, Player A has 200 hits in 600 at bats, all of them singles. Player B has 150 hits in 450 at bats, 40 of which were home runs. Clearly the better hitter was Player B. Player A had more hits, but he also had many more opportunities at the plate. The 40 home runs hit by Player B indicate that he likely had a more positive impact on his team's ability to score runs.

I wanted to look at whether more sophisticated measures of hitting ability would still be accurately predicted by the model, and if so, whether the skill advantage for left-handed hitters differs from $\Delta\hat{s} = 0.3003$ in these cases.

I ranked hitters by three different statistics: home runs, batting average (AVG), and on-base plus slugging percentage (OPS).

AVG is simply the fraction of at-bats in which the hitter got a hit.

$$\mathrm{AVG} = \frac{\text{hits}}{\text{at-bats}}$$

OPS attempts to measure two aspects of a hitter's ability: contact and power. It is the sum of on-base percentage (fraction of plate appearances in which the hitter reaches base) and slugging percentage (average number of bases per at-bat).

$$\text{OPS} = \frac{\text{hits + walks + hit by pitch}}{\text{plate appearances}} + \frac{\text{total bases}}{\text{at-bats}}$$

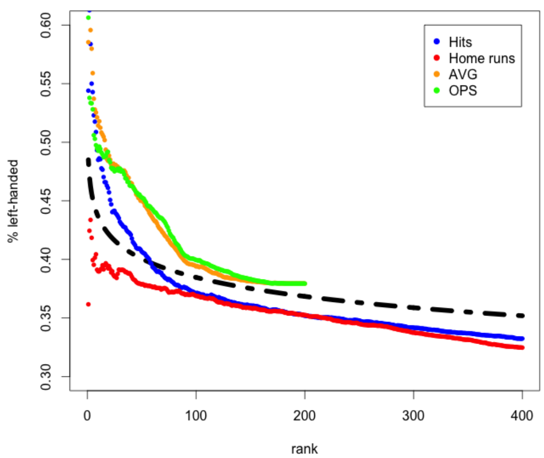

Figure 4: *Percentage of left-handed hitters as a function of season rank by various stats. The black dashed line is the prediction from the model with* $\Delta\hat{s} = 0.3003$.

The results are shown in Figure 4. The black dashed line shows the proportion of left-handed hitters as predicted by the model. Notice that the curve appears to fit the plots of hits and home runs more closely than the plots of batting average and OPS. I believe the reason for this is that the model is designed to treat hitting as a competition between hitter and pitcher on an at-bat by at-bat basis. Hits and home runs are absolute measures of the number of successes achieved by the hitter, whereas AVG and OPS are averages. These averages are more useful for determining the overall quality of a hitter, but this model is better suited for predictions based on totals such as hits.

Finally, I looked at pitching stats as well. Just as with the hitters, I considered two absolute measures and two more sophisticated statistics used to judge pitching quality.

The absolute measures I used were outs recorded and strikeouts. I chose these to be analogous to hits and home runs, respectively. Additionally I used earned run average (ERA) and WHIP.

ERA is simply how many earned runs the pitcher has allowed per nine innings pitched.

$$\text{ERA} = 9 \times \frac{\text{earned runs}}{\text{innings pitched}}$$

WHIP stands for walks plus hits per inning pitched, and is self-explanatory.

$$\text{WHIP} = \frac{\text{walks + hits allowed}}{\text{innings pitched}}$$

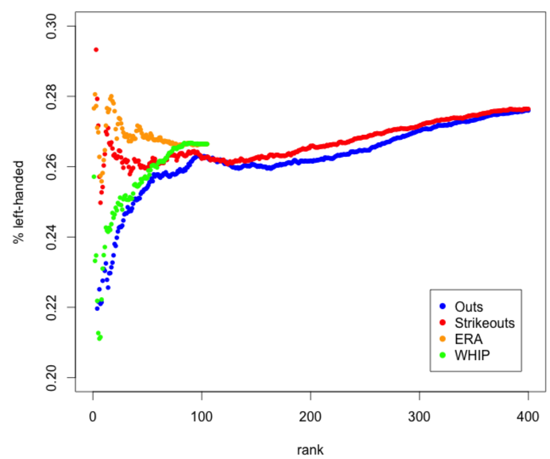

Figure 5: *Percentage of left-handed pitchers as a function of season rank by various stats.*

The results are shown in Figure 5. As we might expect, the plot of outs recorded looks like the mirror image of the model's prediction curve and of the plot for hits in Figure 4. If left-handed hitters have an advantage over right-handed pitchers, as appears to be the case, it makes sense that left-handed pitchers would have a disadvantage against right-handed hitters. Since the majority of hitters are right-handed, we might expect that a smaller proportion of left-handed pitchers rank highly.

Interestingly, this behavior is only reflected for outs and for WHIP. Both strikeouts and ERA show a fairly constant proportion of lefty pitchers over all ranks. It is apparent that the model does not do as well for pitchers as it does for hitters.

References

[1] Daniel M Abrams and Mark J Panaggio. A model balancing cooperation and competition can explain our right-handed world and the dominance of left-handed athletes. *Journal of The Royal Society Interface*, 9(75):2718–2722, 2012.

[2] S Lahman. The lahman baseball database. see http://www.baseball1.com, 2011.