Nick Foster

February 23, 2010

Prof. Little

Montserrat

Joe DiMaggio’s Streak

Considered one of the greatest feats of hitting in baseball by baseball fans, players, and writers, Joe DiMaggio’s record 56 game hitting streak may never be broken. Started in mid-May during the 1941 season, Joltin’ Joe got a hit in every game he played in for the next two months. This accomplishment has stood the test of time since then, as now one has even come close to it since. Only one player has even gotten more than a 40 game hitting streak since, which Pete Rose surpassed in 1978, finally having his streak end at 44 games.

Now most people who are familiar with baseball will assert that DiMaggio’s streak will never be broken simply because to maintain that sort of consistency over strand of five or ten games is hard enough, but to last 56 games in a row is near impossible. To achieve a streak like a player must elevate his game to a near superhuman level, especially in a sport where getting a hit 3 times out of ten is considered great.

Others, like Leonard Mlodinow in his *Wall Street Journal* article “The Triumph of the Random, From banking to baseball, winning streaks owe much to the laws of chance”, will say that while DiMaggio’s high skill-level (DiMaggio is in the Baseball Hall of Fame) played a part in his miraculous streak, the laws of probability and chance may have been the driving force behind that two month hot streak in the summer of 1941.

Mlodinow in his article argues that DiMaggio just happened to be in the right place at the right time and had just the right amount of talent to succeed. As a result, Lady Luck shined upon him, and led to the streak. Mlodinow uses other real world examples to help prove his point, the most basic of which is the coin toss situation. He says that a coin, when tossed, has a 50% chance of coming up heads as well as a 50% chance for tails. So if you tossed that coin 100 times you would expect each to appear 50 times randomly. However there is a 75% chance in this scenario that you would see a streak of 6 or more heads or tails in a row.

This equates to the DiMaggio streak in that, while DiMaggio was a .325 career hitter, for him to have a sudden hit streak that would cause him to hit in 56 games where his batting average would jump up .083 to .408, is more due to the fact that a streak of this length was inevitable due to probability. The chance that a streak of this length was to probable not to happen, therefore it had to happen eventually; DiMaggio just happened to be the man to do it.

He backs this up with a study done by Samuel Arbesman and Stephen H. Strogatz of Cornell University, who used all the baseball statistics from 1871 to 2005 to simulate 10,000 alternative baseball universes; basically rewriting baseball history 10,000 times. They found that 42% of these histories had a streak that either matched or surpassed DiMaggio’s. The hitting streaks ranged from 109 games to 39 games. Ty Cobb, not DiMaggio, broke the record over 300 times throughout the 10,000 repetitions.

The baseball history of real life would just be one of those simulated histories that had a hitting streak of 56 games or more. Based on this, DiMaggio streak seems far less miraculous and instead seems to be almost a normal event. The main problem with this argument is that completely eliminates the human aspect of baseball, which is a sport played by humans, not a computer program in a lab at Cornell.

There is no doubt in my mind that this computer simulation by no means accounted for every variable in baseball. In fact, accounting for these variables would in fact be impossible due to the fact that baseball is a game of infinite variables. The fact that human beings play the game means that the game is affected in a way that a computer program cannot account for, even with all the statistics they wanted to add to it. Mlodinow even confirms this for us, stating “They are just asking whether a similar streak would have happened sometime in the history of baseball even if each player hit with the unheroic and unmiraculous – but steady – ability of a robot.”

That in and of itself should disprove his point; he contradicts himself from the beginning. The players playing baseball don’t hit with the steady ability of a robot. Players wear down over a season. DiMaggio was no exception; he suffered a knee injury during his minor league career that would have been a factor while playing a long baseball season. Even without the knee injury, players aren’t always fresh. Players can’t be at 100% all the time. During his streak, DiMaggio played seven separate double-headers. In today’s world, players won’t even start both ends of a double-header, never mind get hits in both games. DiMaggio went a step beyond and hit .410 in double header games.

Even in the computer scenario, looking at the results more closely, the statistics actually support the fact that the DiMaggio streak is an anomaly. It says that a streak of DiMaggio’s magnitude happened in 42% of those 10,000 scenarios. This means more often than not, meaning less than half the time, did a streak surpass DiMaggio. It mentions that the record streaks ranged from 109 games to 39 games. A 109 game streak is at the top of that range, meaning it in and of itself is a onetime occurrence, meaning that it happened once in 10,000 tries. It happened .01% of the time, making it seem far less as outrageous as one might think.

Not only did DiMaggio play tired, but he played against top talent. During his streak he batted against four Hall of Fame pitchers: Lefty Grove, Bob Feller (twice), Hal Newhouser (twice), and Ted Lyons. These pitchers were dominant in their era and were extremely hard to hit. Anyone of these 6 games could have ended the streak for DiMaggio but he still hit. We cite that DiMaggio had a high batting average which increases his chances of getting it hit, but we never consider that the pitchers DiMaggio faced would cause his chances to diminish. Bob Feller gave up the least hits in the American League in 1941 with 284, which certainly diminished DiMaggio’s odds of getting a hit in either of those games. Feller in fact gave up only a little more than 6 hits per game hit pitched in, meaning that DiMaggio essentially had to have one of approximately six hits that would be allotted to his team in which he played against Feller.

The pitchers didn’t even have to be Hall of Famers; it could have just been a pitcher who knew how to get DiMaggio out. Anyone who has played baseball or even follows the sport knows that sometimes a batter just doesn’t read a pitcher well. Maybe the batter can’t see the ball coming out of the pitchers hand that well due to his delivery or maybe he can’t pick up the spin on the pitcher’s pitches, which are both vital when as a batter you have less than half of a second to react in order to hit a pitch. DiMaggio faced over 30 pitchers during his streak. Chances are he would have faced at least one that would have given him a hard time, yet he still rose above these difficulties and continued the hit streak. There is no way that the computer program could have accounted for either of these factors.

It’s not just DiMaggio’s streak that in sports that cannot be attributed to mostly probability. Mladinow cites the World Series, saying that in the current format, the superior team will not win anywhere near as often as you would expect them to.

In baseball, even if one assumes that the better team has a lopsided 55/45 edge over the inferior one, the lesser team will win the seven-game World Series 40% of the time… If you play a best-of-one game series, then, by our assumption, the lesser team will win 45% of the time. Playing a longer series will cut down that probability. The problem is that playing a seven-game series only cuts it down to 40%, which isn’t cutting it down by much. What if you demand that the lesser team win no more than 5% of the time—a constraint called statistical significance? The World Series would have to be the best of 269 games

The World Series is supposed to show the superior team and crown them champion, but it is not supposed to make it inevitable for them. The games still need to be played, and that is what is great about sports. Nobody watches the NCAA basketball tournament knowing that the system will work out so all the number one seeds make the final four. In fact it’s quite the opposite. People love to see upsets. That’s what is great about the World Series. Yes, one team may be heavily favored, but to see the underdog team play far above the ability and rise above expectations is a wonderful sight. The superior team should be able to deal with the pressure and perform to their level and beat and inferior team four times in a seven game series. If not, then they do not deserve to win that series, and should not be crowned champions.

People love sports because they go against everything we expect. When an underdog beats a clearly superior team, it’s amazing to watch. When a player displays abilities far beyond what we expect from them we can’t look away. Sports are great purely because they defy probability. Every so often an athlete stands up in the face of pressure, injuries, and fatigue and accomplishes something far beyond what anyone believe they can accomplish. It happens over and over again, and every time it does it inspires the people who watch it. The USA-Soviet Union hockey game from 1980 Olympics. The 11-seed George Mason making the Final Four by defeating the 1-seed UConn. Joe Namath guaranteeing victory and then beating the Colts in Super Bowl III. In each case the probability of all these teams winning was slim to none, and yet in each case they defied the odds and created great sports moments. That’s why probability in sports have far less of an effect than any computer program or mathematical formula can ever truly capture. This applies directly to Joe DiMaggio’s famed hitting streak. It is a prime example of a sports scenario that probability cannot explain, and should not attempt to.