Rory Van Tuyl
2/18/2011

Say it Ain't So, Yogi...

A friend of mine, a Wild Welshman known for his provocative and sometimes controversial opinions, posed the following query: "Do baseball teams playing in the "World" (HA! HA!) Series throw games [?]" This was enough to make any red-blooded American tremble with rage! How dare he even raise this issue? And to do so right in the middle of 2010 World Series with the SF Giants led by the heroic Tim Lincecum trouncing the hated Texans! Insufferable!

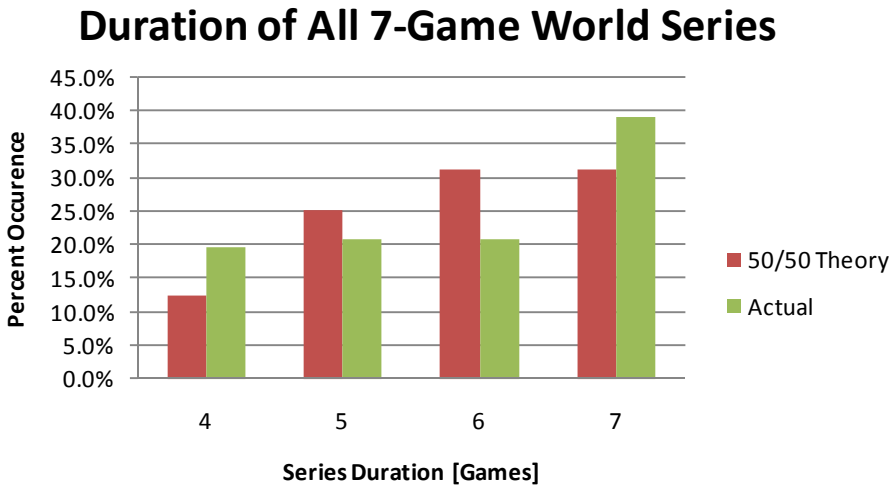
Well, OK, there was the infamous "Black Sox" scandal of 1919 in which the team threw the series in exchange for kickbacks from gamblers. But this was ancient history. The Chicago players of that day made next to nothing and had to pay the infamously tight-fisted Charlie Comiskey for the privilege of doing their own laundry! Today's players are paid so handsomely that it's hard to imagine what incentive they might have for doing such a dirty deed.

But still…what about the past? Baseball being replete with statistics of every description, and information about past World Series being readily available online, I decided to look into the matter.

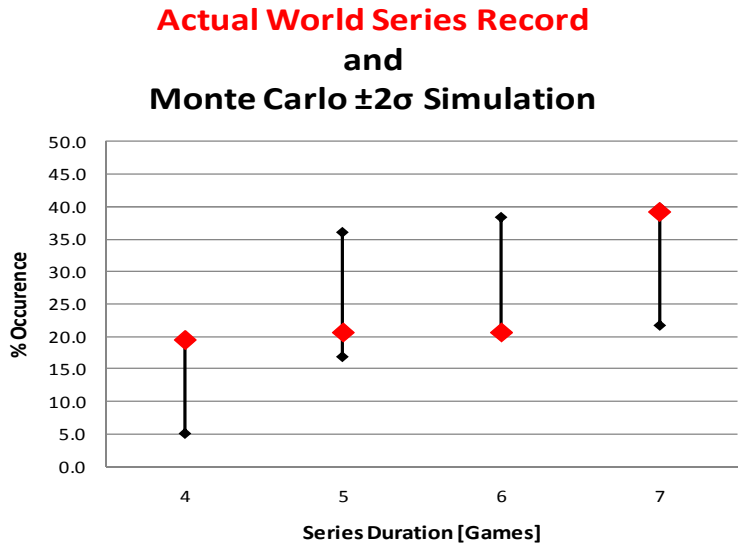
The World Series as it was played in 1909 and has been played every year since 1924, consists of a best four-out-of-seven games tournament. The statistics for how long a World Series should last if -and it's a **big if**- teams were evenly matched and the outcome of each game were as random as the toss of a coin, should be, according the Wild Welshman (he was right in this case):

4 Games: 12.5%
5 Games: 25%
6 Games: 31.25%
7 Games: 31.25%

So the first data to be mined from the World Series treasure trove of statistics were simply this: how long have Series lasted, historically? If the results matched the "pure chance" 50/50 theory, there could be no case made for skullduggery, at least not continued and repeated skullduggery. But if there were, by some remote chance, a deviation from this theory, a new theory would be required, and who knows where that might lead?



As can be seen from the above chart, the "50/50" random occurrence theory appears to be all wet. More Series conclude after just 4 games, and far fewer after 6 games than would be expected by pure chance. But to make sure that the actual outcome wasn't just a statistically reasonable instance of the myriad possibilities available in a random world I ran what's called a "Monte Carlo" simulation. In this simulation the computer "plays" 100 World Series assuming each game is statistically independent of the others, and shows the range of outcomes we would expect 95% of the time.

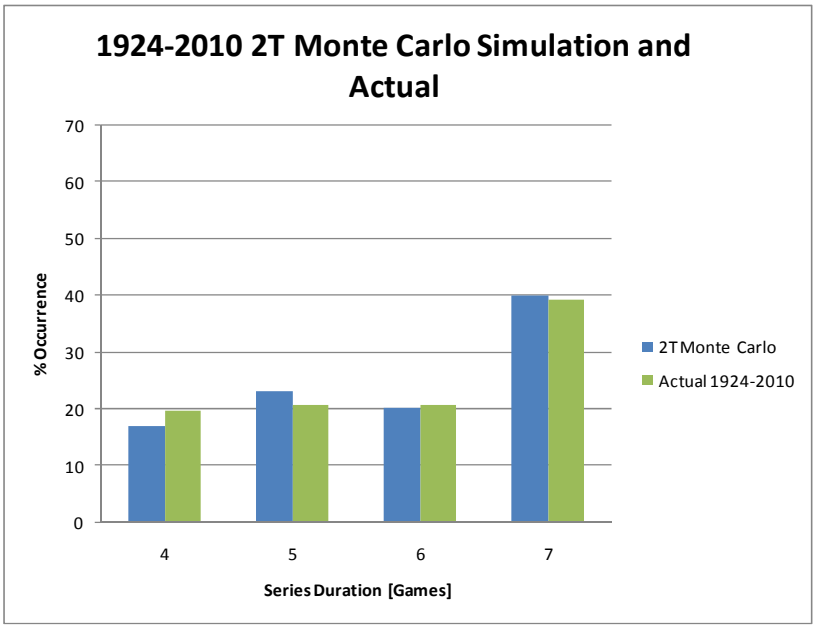


In this chart the red dots represent the actual record, and the black bars represent the range of outcomes for ninety five out of one hundred World Series. So yes, it is just barely possible that the actual record could be explained by random chance. But since there have been less than 100 seven-game World Series played, the chances that we would have seen the actual historical record merely by chance are very, very slim.

So I asked the question, "What really goes on in a World Series?" Baseball is a game of skill as well as chance. Even the best batters succeed in hitting the ball only about one third of the time, and the best pitchers are often so below par that they must be yanked from the game. But leaving those details aside, I looked at the two types of World Series you could imagine from a simple point of view: Series between well-matched teams, and Series where one team is just more powerful than the other. Given that these teams have never played each other in the regular season, there can be no *a priori* knowledge of their relative strengths. That picture should emerge after the first four or five games. If one team is much better than the other, the Series will likely end after four or five outings. But if the Series goes to a sixth or seventh game, the teams must be well-matched. The Monte Carlo simulation could be enhanced to reflect this notion of two types of Series: well-matched (50/50 theory); mismatched (new theory).

The new theory I tried postulated a chance of winning for each team in the first 4 to 5 games. After that, it examined which team had the necessary 3-to-2 game lead at the end of five games played, and assigned that team different odds of success.

This approach hit a time run first time at bat. By assuming that all teams in the simulated 100-game series started out evenly matched, and that the survivors of five games must be slightly mismatched, with a 3-to-2 advantage going to the team that had won 3 of the first 5 games, I got a good fit between the Monte Carlo simulation and the actual record.



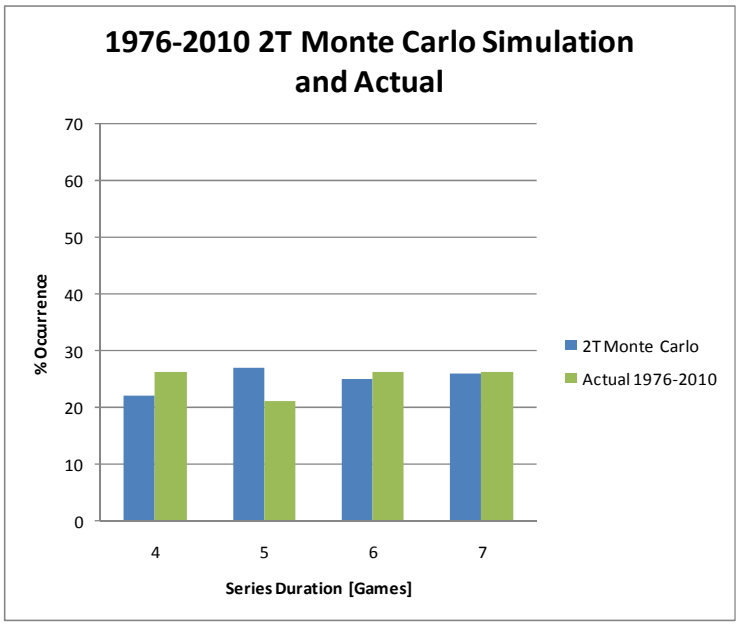
End of story, right? Not so fast! A good investigator doesn't stop looking when he has just one suspect, does he? A scientist doesn't rush to publication without considering all the alternatives, right? So let's take a look at some really intriguing detail imbedded in the historical record.

By examining the World Series outcomes from 1924 to 2010, I noticed two very interesting eras: 1955 – 1975 and 1976-2010. The statistics for these two eras were completely different, and the results could both be matched fairly well with the new theory in Monte Carlo simulation.

First, the Modern Era. The years 1976-2010 were the era in which the "Reserve Clause" that held players in thrall to their original teams had vanished, letting players become free agents with their own hired barracudas to milk the teams – and thus the fans – for top dollar. During the years 1967-2000 for which I could find data, the ***average*** major league player went from earning $19,000 per season to taking home $1,895,630! This was a 100-fold increase brought about by a relentless compound annual growth rate of 14.9% over the period, far greater than ***my*** salary increased over those same years! Clearly, players earning a couple of mill per year would have little incentive to do anything but bust their butts in each and every game, even to the point where they would take performance-enhancing drugs to increase their market value and satisfy

their competitive instincts. Gamblers or other meddlers just could not compete with the rewards these player received above-board. But what about TV revenues? The wild Welshman had proposed that nefarious owners might pull some dirty trick to stretch out the Series, thereby filling their already-bursting coffers with yet more filthy lucre. But in fact, during the modern era of huge TV revenues, exactly the opposite has occurred: more Series end in just four games than would be predicted by the 50/50 model, and there are fewer seven-game series than called for by pure chance. My British friend would be advised to sniff around the football stadia or cricket pitches of his native island than to further impugn or noble ninesomes!

Indeed, the 1976-2010 World Series record looks a lot like the 1924-2010 record, not adhering to the 50/50 theory, but easily replicated by the new Monte Carlo simulation.



To match the 1976-2010 record, I assumed that the teams were mismatched in the first 4-5 games, with 2:1 advantage to one team. Naturally, this boosted the number of Series that would retire in 4 or 5 outings. This makes lots of sense. As mentioned earlier, quick wins should most often occur when teams are mismatched. The simulation for 1976-2010 then assumed that teams playing beyond 5 games were evenly matched, with 50-50 odds of winning each of the remaining games. Once again, this makes sense: well-matched teams play longer series.

So the Monte Carlo simulation seemed to explain all. Or did it? I had made very different assumptions in order to fit the 1924-2010 record and the 1976-2010 record. I felt a bit queasy about the 1924-2010 fit, but the 1976-2010 fit seemed to be consistent with common sense.

And then there was the "Golden Age" of Major League Baseball, 1955-1975. (At least it was "Golden" from my point of view.) This was the era of the heroes of my youth: Casey Stengel and his Yankees, Walter Alston and the Brooklyn Dodgers, Red Schoendienst of the St. Louis

Cardinals, and – after Casey's retirement, the all-time baseball icon Yogi Berra with the Yankees. Slugger Ted Williams of the BoSox, Jackie Robinson and the 1955 World Champion Dodgers, Whitey Ford with his 1,956 career strikeouts. These were the Gods of Baseball. No Welshman would dare to besmirch these guys' honor!

There was only one problem with the "Golden Age." In 20 years, out of 15 Series played beyond 5 games, only ***once*** did the series stop at 6 games! This means that year after year, the team that led the Series by 3-to-2 after 5 games took the day off, statistically speaking, for game six. Here, if anywhere in the history of the World Series of Baseball, was the possibility of skullduggery!

In the years after WWII, major league players didn't make much money. As late as 1969, the year I joined HP as a young engineer at a base salary of $12,600, a ball player talented enough to make the majors would start at $10,000 per annum. The pay in 1955 must have been paltry indeed. Resentment toward the league owners, who had been handed an exemption to the US antitrust laws back in 1915 by Judge Kennesaw Mountain Landis and had squeezed the players ever since, was building. Starting in 1969 the talented center fielder Curt Flood challenged major League Baseball's reserve clause all the way to the Supreme Court. He lost, but the "injustice" of the reserve clause had been exposed and by 1975 it died a well-deserved death.

So, did the politics of "Golden Age" baseball affect the outcome of those Sixth Games, games that should have been won by the leading team half the time, but were instead won only once in fifteen Series? Here, the detailed description of, and stats for, the sixth games give little insight. If there was hanky-panky, it would take a real insider to know it.

Nonetheless, I feel compelled to switch here from the guise of an objective investigator to that of an unapologetic defender. I simply cannot let "Lies, damned lies and statistics" do a number on my heroes of yore. Here is the case for the defense:

1. **Statistics cut both ways.** While it is true that the game 6 statistics (14 out of 15 Series that went more than 5 games went for the full 7 games) are skewed, the Series outcomes are not. Simple 50/50 theory predicts that the team lagging behind with only 2 wins after 5 games should reverse their fortunes 25% of the time to win the Series after 7 games. In the "Golden Age" this reversal occurred exactly 26.5% of the time. Right on the money: no fraud there!
2. **Home field advantage can explain it all.** Of the 14 game sixes in question, the ones where 50/50 theory predicts that the leading team should have won seven but actually won none, ten of the games were won by the team playing at the home field. Interestingly, these same teams were at home in game one of the same series, and in those games, the home field advantage was hard at work: the game six winners won eleven out of fourteen of their first games because…(wait for it)…they were on their ***home field!*** The hometown fans, cheering on their boys, caused these laggers to stay alive to game 7.

And so, my wild Welsh friend: **I rest my case!!!**