Vamshi Jandhyala

How Lucky Can a Baseball Team Get?

August 2026

A puzzle from Fiddler on the Proof, about the records that a season manufactures out of nothing.

Problem

Two teams of equal skill, the Algebraists and the Geometers, play each other 162 times over a season. Each game is won by either team with probability $\frac{1}{2}$, independently of every other game. On average, how many games would you expect the team with the *better* record to have won?

Solution

The teams play nobody but each other, so there is only one random quantity in the whole problem. Let X be the number of games the Algebraists win. Every game has a winner, so the Geometers win the other $162 - X$, and the better record is

$$\max(X, 162 - X) = 81 + |X - 81|.$$

The two records are not two separate pieces of luck. They are one piece of luck, read forwards and backwards. Whatever the season does, the better record sits as far above 81 as the worse one sits below it, and the question is simply how far that is.

So the answer is 81 plus the average distance from X to its own mean, with $X \sim \mathrm{Bin}(162, \frac{1}{2})$. That average distance is the mean absolute deviation, and for a symmetric binomial it has a closed form worth knowing. For $X \sim \mathrm{Bin}(2m, \frac{1}{2})$,

$$\mathbb{E}\,|X - m| \;=\; \frac{m}{4^m}\binom{2m}{m}.$$

The identity is easy to believe from the smallest case: with $2m = 2$ the distance is $1, 0, 1$ with probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$, giving $\frac{1}{2}$, and the formula gives $1 \cdot \binom{2}{1}/4 = \frac{1}{2}$ as well.

Taking $m = 81$,

$$\boxed{\mathbb{E}[\text{better record}] \;=\; 81 + \frac{81}{2^{162}}\binom{162}{81} \;=\; 86.0699 \text{ wins.}}$$

What the number is saying

For large m the central binomial coefficient satisfies $\binom{2m}{m}/4^m \sim 1/\sqrt{\pi m}$, so the better record sits about

$$\sqrt{m/\pi} \;=\; 9/\sqrt{\pi} \;\approx\; 5.08$$

games above .500. Two teams that are exactly as good as each other finish five games apart, and there is nothing to explain: the gap is what 162 coin flips look like. A season is long enough to produce a standings table and not nearly long enough to make it mean anything.

The computation

Playing the season is a better check than evaluating the formula again, since it tests the model rather than the algebra.

```
import numpy as np

rng = np.random.default_rng(20260811)
N = 4_000_000
wins = rng.binomial(162, 0.5, N)     # the Algebraists' wins
better = np.maximum(wins, 162 - wins) # the Geometers win the rest
print(f"{better.mean():.6f}")
# 86.070587     exact value 86.069876
```

Four million seasons give 86.0706 against the exact 86.0699.

Extra Credit

Now there are 30 *teams of equal skill, and each team plays each other team five times, so each plays* 145 *games in all. On average, how many games would you expect the team with the* best *record to have won?*

Solution

Two facts about a single team survive the move to a league, and one does not.

Each team's 145 games are still 145 independent fair coin flips from its own point of view, so the marginal law is exactly $W_i \sim \mathrm{Bin}(145, \frac{1}{2})$, with mean 72.5 and variance 36.25. That much is unchanged.

What breaks is independence. The 435 pairings supply 2175 games, every game gives out exactly one win, and therefore

$$W_1 + W_2 + \cdots + W_{30} \;=\; 2175 \quad \text{always.}$$

Two teams meet five times, and in those five games one team's wins are the other's losses, while all their other games are independent. So

$$\mathrm{Cov}(W_i, W_j) \;=\; -\tfrac{5}{4}, \qquad \mathrm{Corr}(W_i, W_j) \;=\; -\frac{5/4}{145/4} \;=\; -\tfrac{1}{29}.$$

That value is forced, and the conserved total shows why: a sum with no randomness in it has no variance, so $30 \cdot \frac{145}{4} + 30 \cdot 29 \cdot \mathrm{Cov} = 0$, which gives $\mathrm{Cov} = -\frac{5}{4}$ with no calculation about shared games at all.

The maximum of thirty variables depends on their joint law, not on their margins, and this joint law does not factorise. There is no closed form to find, and pretending otherwise would be the only real error available here. The value below comes from playing the league out:

$$\boxed{\mathbb{E}[\text{best record}] \;\approx\; 84.98 \text{ wins.}}$$

Independence is not a harmless simplification

The tempting shortcut is to treat the thirty teams as thirty separate 145-game seasons. Under that model the answer is exactly computable, since the maximum of independent variables has distribution function $F(k)^{30}$:

$$\mathbb{E}[\text{best}] \;=\; \sum_{k\ge 0}\left(1 - F(k)^{30}\right) \;=\; 84.7711, \qquad F = \text{c.d.f. of Bin}(145, \tfrac{1}{2}).$$

That is 0.21 wins short of the truth. The shortfall is small, but its sign is the interesting part, and it is the opposite of what most people guess. Negative correlation sounds like it should hold the leader back, since one team can only climb by pushing others down. It does the reverse, and the reason is that the pushing down is the point.

Three facts settle it, and none of them needs the joint law written out. Every team still has variance $145/4 = 36.25$, so no single team is more erratic in one model than the other. But within a season the thirty records are spread further apart in the round robin: for exchangeable variables with correlation ρ, the expected sample variance is $\sigma^2(1-\rho)$, and with $\rho = -1/29$ that is

$$36.25 \times \tfrac{30}{29} \;=\; 37.5 \qquad \text{against} \qquad 36.25.$$

And the centre those records are spread around does not move: the mean record is exactly 72.5 in every round-robin season, where under independent schedules it wanders with a standard deviation of about 1.1 wins. So the round robin fixes the middle of the table and pushes the ends apart, which lifts the top of it. The same pinning is why the best record is also *less* variable in the round robin, standard deviation 2.81 against 2.98: the leader is measured from a centre that never moves.

The two-team case proves the direction, because there the coupling is total and both answers are exact. The puzzle itself gives 86.0699. Had the Algebraists and Geometers instead played 162 games each against the rest of the world, the better of their two records would average

$$81 + \tfrac{1}{2}\,\mathbb{E}\,|X - Y| \;=\; 81 + \frac{81}{2^{324}}\binom{324}{162} \;=\; 84.5877,$$

using $X - Y + 162 \sim \text{Bin}(324, \tfrac{1}{2})$ to turn a difference of two binomials back into one. The excess above .500 is 5.0699 against 3.5877, a ratio of 1.4131, and the asymptotics say where that is heading: $\sqrt{m/\pi}$ against $\sqrt{m/2\pi}$, a factor of exactly $\sqrt{2}$.

With two teams the correlation is -1 rather than $-1/29$, and the mechanism above runs at full strength: the mean record is pinned at exactly 81 and the whole of the season's randomness goes into separating the two. With thirty teams each pairing supplies only five of a team's 145 games, the correlation is diluted to $-1/29$, and the same effect survives at a fifth of a win.

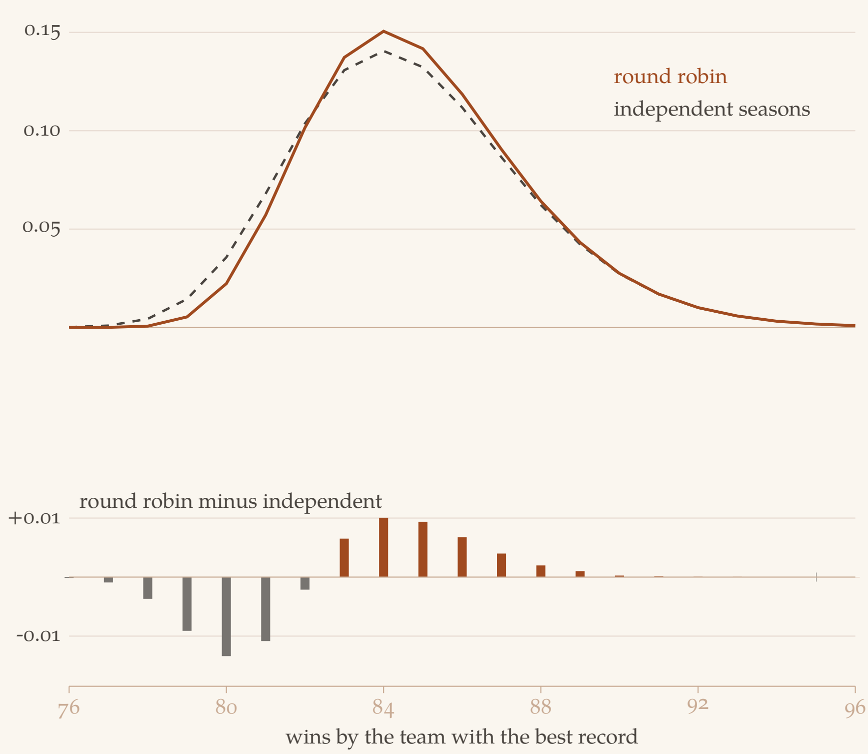

Figure 1: *The best record in a* 30*-team round robin, against the same thirty teams playing independent* 145*-game seasons. Same margins, same variance per team, different joint law. At the top the two curves are almost indistinguishable, which is the honest impression: the effect is a fifth of a win. The lower panel is their difference, and there the structure is plain. The round robin takes probability off the low records, below about* 82*, and puts it on the middle and upper ones, from* 83 *to about* 89*. It is a rightward shift rather than a fattening of the tails, and it comes with a slight tightening, standard deviation* 2.81 *against* 2.98.

The computation

Two encodings, written separately, to guard against a plausible bug agreeing with itself. The first draws a $\mathrm{Bin}(5, \frac{1}{2})$ for each pairing; the second flips all 2175 games one at a time. Both check that every game has been accounted for before taking the maximum.

```
import numpy as np

T, PER = 30, 5
pairs = [(i, j) for i in range(T) for j in range(i + 1, T)]
rng = np.random.default_rng(20260811)

M = 4_000_000
W = np.zeros((M, T), dtype=np.int16)
for i, j in pairs:
    d = rng.binomial(PER, 0.5, M).astype(np.int16)
    W[:, i] += d
    W[:, j] += PER - d
assert (W.sum(axis=1) == len(pairs) * PER).all()

best = W.max(axis=1)
```

```
print(f"{best.mean():.4f}")
# 84.9816   standard error 0.0014
```

Four million leagues give 84.9816 with a standard error of 0.0014. The independent game-by-game encoding gives 84.9796 over four hundred thousand leagues, agreeing to within its own noise.