

Empirical Bayes Estimation of Field Goal Percentage

Why Did I Choose This Project?

Minimum Sports!

Maximum Statistics!

Question!

Which of these two players has a better performance: the one that has made **3** goals out of **5** attempts ($\text{FG}\% = \frac{3}{5} = 0.6$) or the one that has made **100** goals out of **200** attempts? ($\text{FG}\% = 100/200 = 0.5$)

Naive suggestion:

Goals made so far/all attempts

Drawback: very poor measure at the beginning of the season

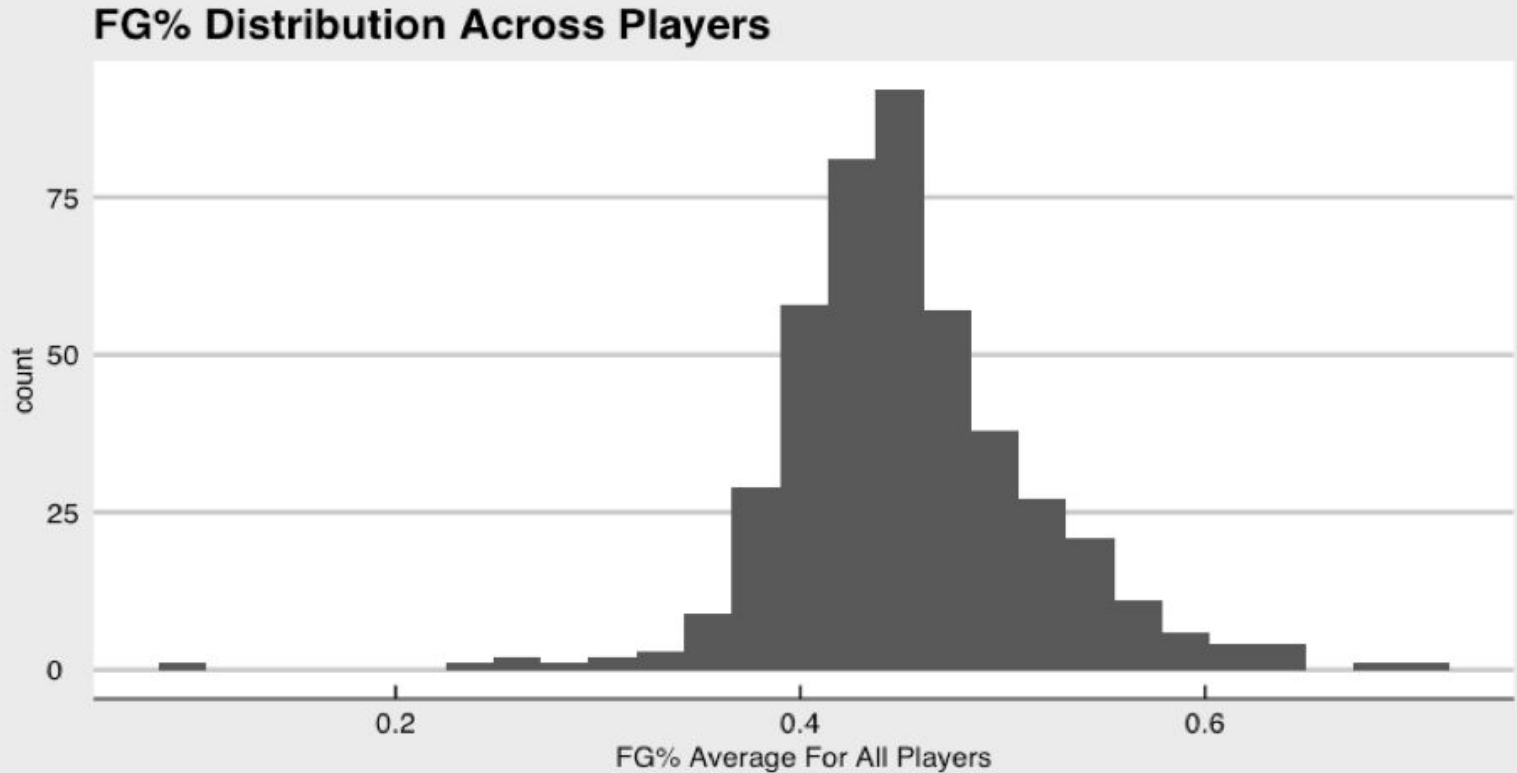
Example: $1/1=1$

$0/1=0$

Objective: Predicting season-long Field Goal Percentage for a player.

- Why don't we predict that he won't get a goal during the whole season?
- Because we're going in with **prior expectations**.
- The best way to present these **prior expectations** is the beta distribution.

Looking at the distribution of FG% across players.



Step 1: Estimate a beta prior using the data.

So, we want to fit the following model:

$$X \sim \text{Beta}(a_0, b_0)$$

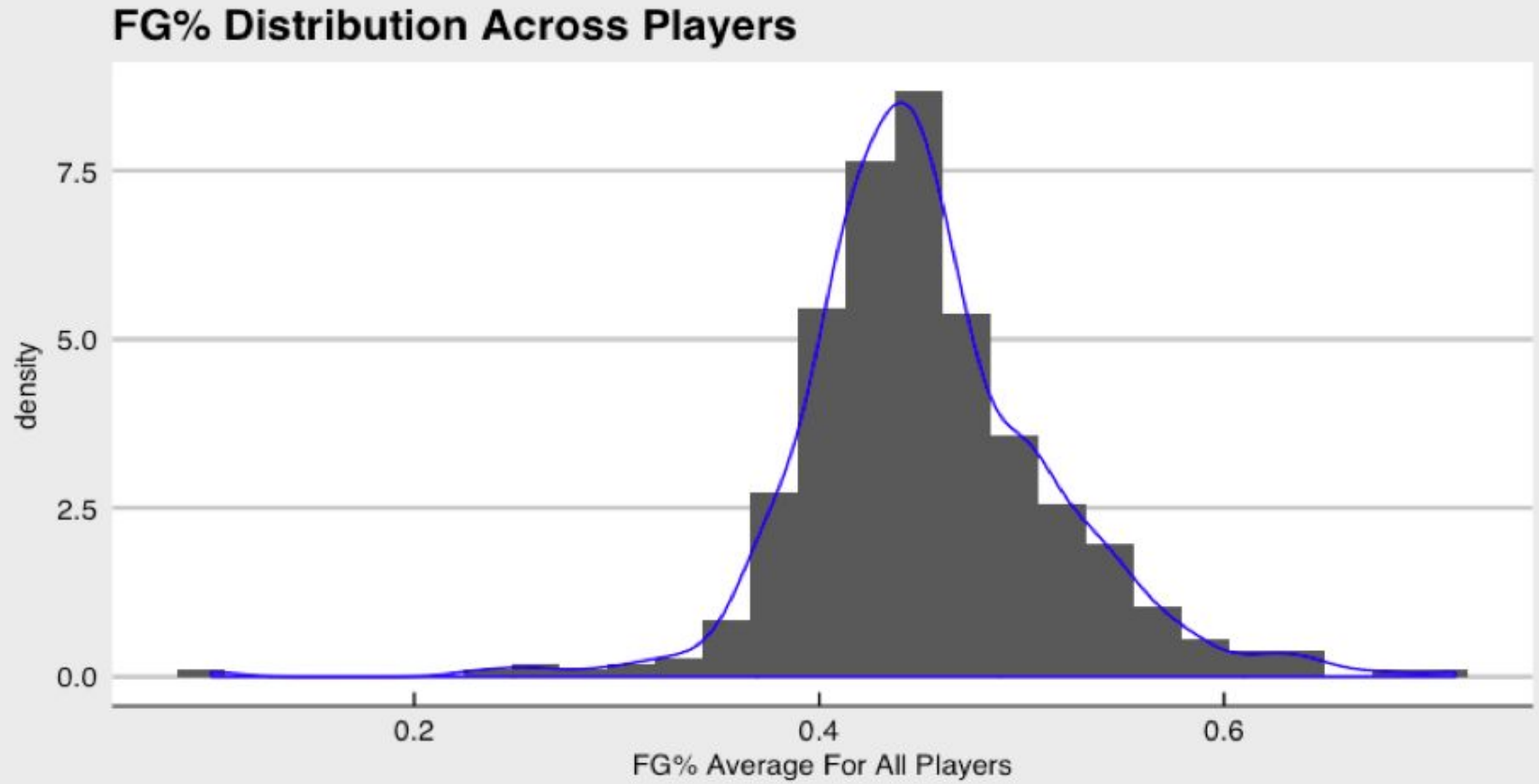
Recall that the beta distribution is given by the following pdf:

$$\frac{x^{(a-1)}(1-x)^{b-1}}{B(a, b)}$$

where

$$B(a, b) = \frac{\gamma(a)\gamma(b)}{\gamma(a+b)}$$

Step 2: Pick (a_0, b_0) Parameters



Step 2: Pick (a_0, b_0) Parameters

$$a_0 = 26.5$$

$$b_0 = 32$$

Step 3: Use that distribution for each individual estimate:

$$\frac{a_0 + \text{goals}}{a_0 + b_0 + \text{attempts}}$$

Back to the question: 3/5 or 100/200?

A player made 200 attempts and got 100 goals:

$$\frac{26.5 + 100}{26.4 + 32 + 200} = 0.4893129$$

Another player made 5 attempts and got 3 goals:

$$\frac{26.5 + 3}{26.5 + 32 + 5} = 0.4643616$$

Results

Best Players

player	fgp	attempts	made	eb_estimate
DeAndre Jordan	0.6771488	3816	2584	0.6737677
Montrezl Harrell	0.6488706	487	316	0.6278768
Clint Capela	0.6215072	1181	734	0.6135578
Salah Mejri	0.6456693	254	164	0.6096141

Worst Players

player	fgp	attempts	made	eb_estimate
Elijah Millsap	0.3274648	284	93	0.3488187
Andrew Harrison	0.3378685	441	149	0.3512928
Marcus Smart	0.3535660	1977	699	0.3564094
Jake Layman	0.3047619	105	32	0.3576242

Pros:

- Didn't pick the players who had very low number of goals but inspected across the long career.
- Very, very easy to implement.
- But when anyone asks what you did, remember to call it “Empirical Bayesian shrinkage towards a Beta prior.” ;)

Questions?

Thank you!

Reference:

- <https://squared2020.com/2017/08/20/basics-in-negative-binomial-regression-predicting-three-point-field-goal-percentages/>
- http://varianceexplained.org/r/empirical_bayes_baseball/