In Baseball, You Don't Know Nothing: The Computational Mechanics of Pitching

Matthew Lawson

For Physics 250

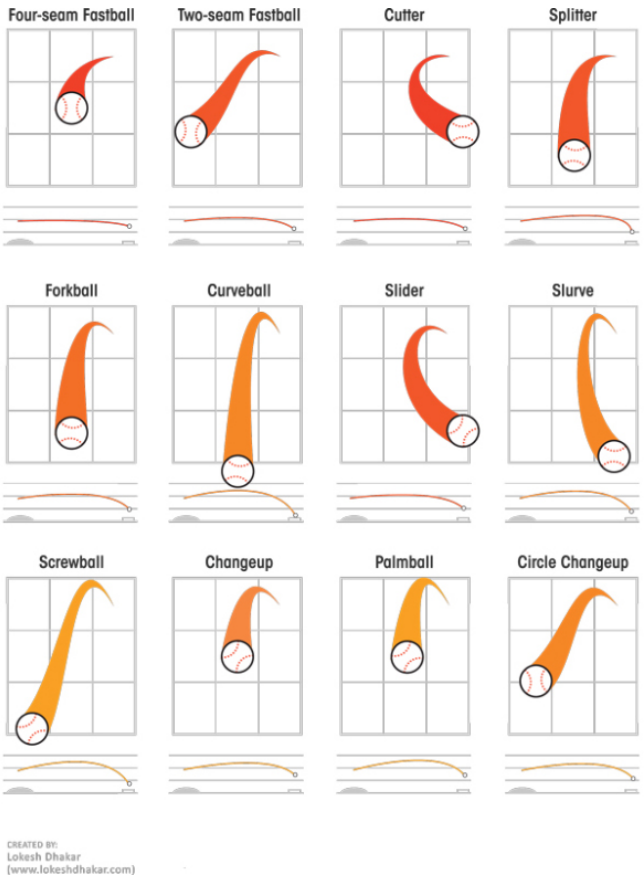
University of California at Davis

June 4, 2015

The Pitch

High-speed delivery of baseball to batter, who wants to hit it or let it pass (ball vs strike).

Pitch Types

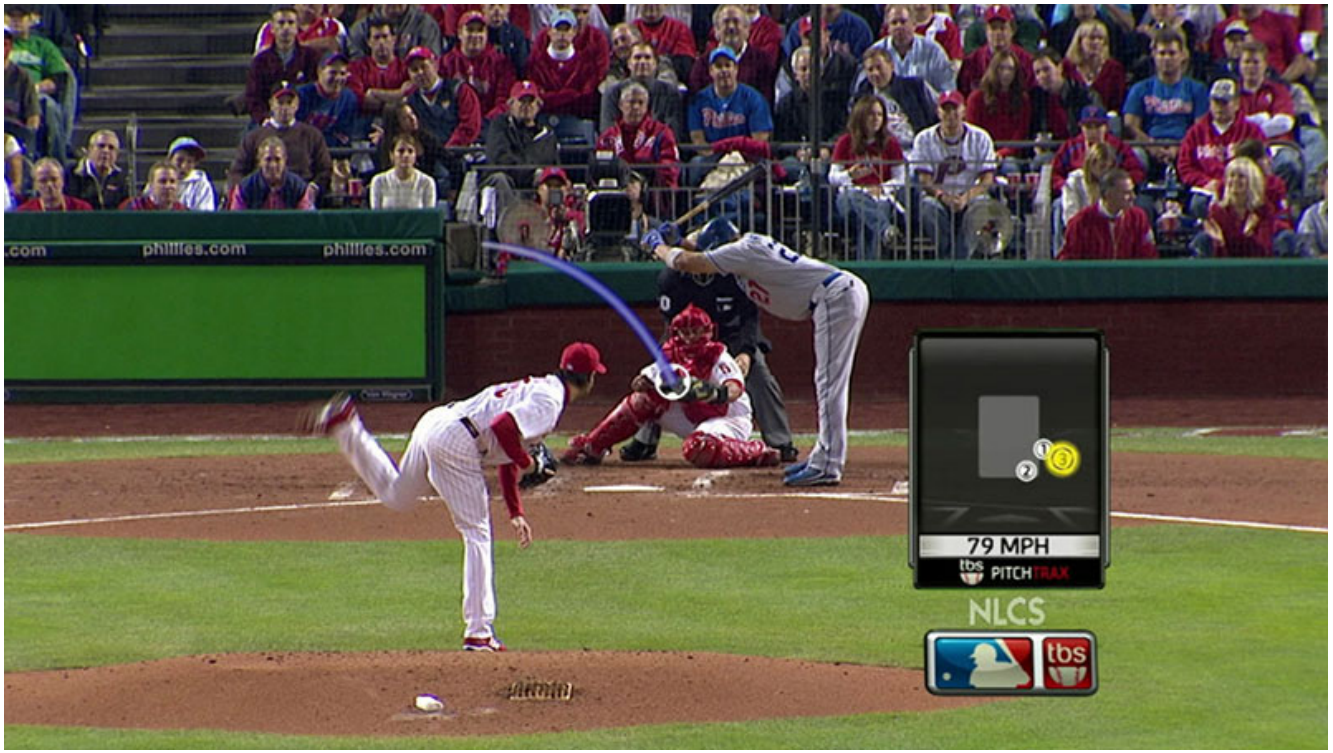


Many wind-ups look similar (change-up identical to fastball, four-seam vs two-seam look identical). Prediction important!

existing prediction methods

Hitters use a wet-ware recurrent neural net (brain) to do real-time motion recognition. This works better for some pitchers than others...Mostly it's guesswork.

pitchf/x



"All pitchers are liars or cry-babies" –Yogi Berra

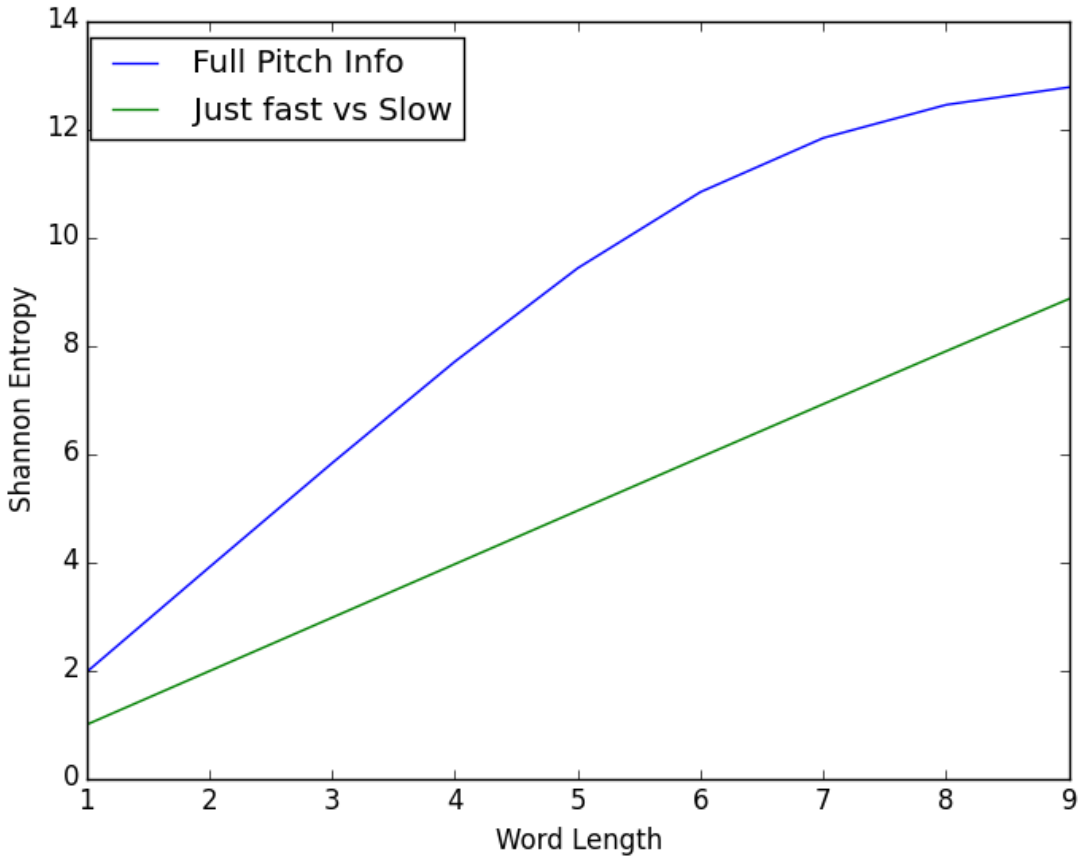
But data doesn't lie.

Pitchrx software package, baseballheatmaps.com, the MLB, and mariadb result in a queriable relational database of every pitch since 2008.

My Dataset

Only Matt Cain
Only right handed hitters
Only well-ordered data (could use more work)
Classify pitches into 5 types (Thank you, pitchf/x)

Entropy



Entropy Results go here

Subtree Merging

Given the large alphabet size Subtree merging seems natural.
Hypothesis: Start state will emerge naturally from data, no need to break up data into at-bats.

partition dataset further

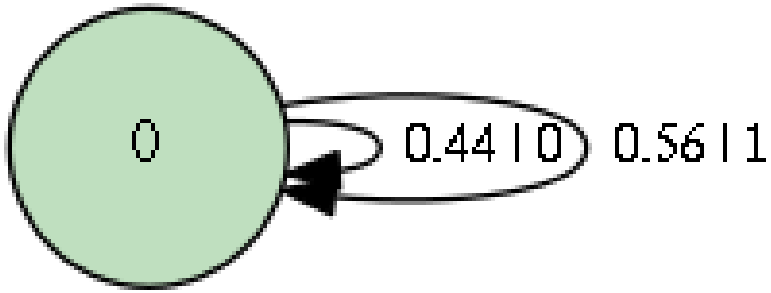
Just use 1 for fastballs and 0 for off-speed pitches, then Bayesian inference becomes at least feasible

Bayesian Inference

A few memory-intensive compute hours later...

Bayesian Inference

A few memory-intensive compute hours later...



In the next week or so...

I'm working on ways to refine the inference methods to work on short samples, like per at-bat, per inning, or per-game

Thanks!

Any Questions?