Effects of pitch selection randomization on performance

Cristian Gamboa

January 3, 2016

Abstract

The importance of randomizing pitch selection is fairly obvious, if a pitcher is utterly predictable, the batter is at an enormous advantage.

The main concern of the article is the magnitude of the effect of pitch selection randomization in pitchers performance. Secondarily, the question if human generated sequences can be studied under the standard randomization test doesn't have an obvious answer, and will need one to complete the study.

Randomness has been tested many ways. Two of the main one is by studying frequencies and sequences. The first two test tackle the former and the third one the latter. Each approach is then tested under a fairly simple statistical model. We restrict the dataset to qualifying starting pitchers from both leagues to ensure sufficiently long pitch sequences.

We get promising results from the first and third tests, and the second shows some interesting further avenues of research.

1 Effects of pitch selection randomization on performance

1.1 Motivation

The importance of randomizing pitch selection is fairly obvious, if a pitcher is utterly predictable, the batter is at an enormous advantage. For example, a pitcher that only throws breaking balls outside the strike zone on a 0-2 count would never get an out on that pitch, as every batter would take the pitch. Equally, a pitcher that throws a fastball in the middle of the zone to ensure a strike every time he's on a 3-0 count would get rocked.

The concern of this article is not if pitch selection randomization affects pitcher performance, but how much. A secondary concern is if human randomization can be captured with the standard randomization tests.

To try to answer this we're using three methods. Randomness can be tested in many ways, but two of the main ones are frequency and sequence testing. The first two methods try the frequency approach, while the third does sequence.

The rest of the article is as follows. The basic model for statistical testing is described in section 1.2. In section 1.3 there is a brief description of each randomness test applied and their results. Conclusions and some further avenues of research are presented in section 1.4. Finally, after that, bibliographical references and an appendix with tables with full results from the tests for each player.

1.2 Model

In order to study this question we're using the simplest model possible:

$$Y_n = \alpha + \beta X_n + \epsilon_n \tag{1}$$

Where Y_n is the FIP- of the player, α is a constant, β the corresponding coefficient for X_n, the statistic used in each test, and ϵ_n an error term.

Using FIP as the independent variable gives us a picture of the pitcher performance clean of confounding factors like defense, luck and event sequencing. Furthermore, using FIP-, instead of FIP, allows us to control for park and league effects.

Given that FIP- is expressed as a percentage variation, it only makes sense to use a logarithmic transformation of the regressor in each case.

It's important to note that in both sides of the equation a smaller number is better[1], so, instead of the intuitive result, we're expecting positive values for β.

1.3 Testing

Using the dataset provided we constructed two main new variables. First, using `pitchType` we classified pitches either as fastballs or as off-speed pitches. `FA`, `FT`, `FF`, `FC` and `FS` were classified as fastballs. `CH`, `CU`, `SL`, `SI`, `FO`, `KN`, `KC`, `SC`, `GY` and `EP` were classified as off-speed. All pitches classified as `PO`, `IN`, `AB`, `AS` and `UN` were dropped.

Second, using `px`, `pz`, `szt`, `szb`, we classified each pitch as a strike or a ball.

We limit the study to qualified starting pitchers from both leagues to ensure the pitch sequences are long enough to satisfy the test requirements.

1.3.1 χ^2 test

The χ^2 test is the first stop when testing for randomness. The basic procedure[2] of this test is the comparison between the frequency of realized observations and the expected frequency. It does it by adding the squared difference in the realized observations and their expectation in each category. Let's call the realized observation that fall in category s Y_s, total number of observations, n, number of categories, k, the expectation of each category p_s, and the statistic we're constructing, V. V is then constructed according to:

$$V = \sum_{1 \leq s \leq k} \frac{(Y_s - np_s)^2}{np_s}$$

The statistic V behaves, under some assumptions, like the χ^2 distribution. Thus we can check against its critical values table to test our hypothesis.

As stated, we will use fastball and off-speed pitches as our categories. In order to avoid using the same data to compute the expectancy and run the test, we obtain the league wide fastball frequencies from the year 2013 and 2014, i.e., 0.55 and 0.54, respectively, with weighted average, 0.54.

[1] This is not actually true in every case for the right side, but it only comes up once in the dataset.
[2] Knuth (1981)

After constructing the statistic, we can see that all but five of the pitchers tested fail to approach a random selection. Most notable is Chris Archer, from the Tampa Bay Rays, who is the only one to fail because he's too close to the expected numbers.

As an example of an actionable item derived from this, we can say that David Price, from the Boston Red Sox, has failed in this measure because he's over representing his fastballs and should throw them around 20% less.

When analyzed under the model stated in section 1.2 we see that the relation is statistically significant[3] at the 99% level. A one percent change in this measure is expected to bring down the FIP- of the pitcher almost two points. To continue with the Price example, the implementation of the change, in theory, would bring down his FIP- almost 13 points.

1.3.2 Expanded χ^2 test

We can expand our analysis space if we also consider the decision of throwing for a strike. In this fashion we now have four contingent categories, i.e. fastball/strike, fastball/ball, off-speed/strike and off-speed/ball. As in the previous test, the expected frequencies for strikes were computed from the years 2013 and 2014, their weighted average is 0.69.

Under this test every pitcher fails the test. This time, we'll use as an example Madison Bumgarner, from the San Francisco Giants, he mainly fails the test because he over represented the off-speed pitches outside the zone. The test suggest a drop of 31% that goes mainly to fastballs outside the zone.

Unfortunately, this model is not statistically significant at any relevant level. Still, it's may be valuable as showing a model for expansion in this line of research. We'll revisit this later.

1.3.3 Maurer test

Let's, once again, consider only the fastball/off-speed dimension. Aside from frequencies, another way of evaluating randomness is sequences. If we call fastballs 1 and off-speed, 0, the series of pitches thrown during the year can be taken as the product of a random bit generator. As such, we can examine it using the test introduced by Ueli M. Maurer in 1992[4].

The test functions as the sum of the distance between a block and the last time it appeared in the series. Formally, let the sequence s^N be divided in non overlapping blocks of length L, K is the number of steps of the test, Q the number of initialization steps and the statistic to be constructed $fT_U\left(s^N\right)$. $fT_U\left(s^N\right)$ is then constructed according to:

$$fT_U\left(s^N\right) = \frac{1}{K}\sum_{n=Q+1}^{Q+K} \log_2 A_n\left(s^N\right)$$

The distribution of the test and its critical values can be found in the referenced paper, but of special note, as a transformation of a normal distribution, it's symmetrical. Thus, instead of the statistic, we're using the distance from the mean of the distribution as the regressor in the model. The block size selected was 3, given that the sequences are a little shorter than optimal, for example, in average NL qualified pitchers threw 3048 pitches last year.

We can see that all but two pitchers pass the test[5]. On the other hand, Chris Archer sequencing is

[3] See table 1

[4] Maurer (1992)

[5] The exceptions are Jake Odorizzi, from the Tampa Bay Rays, and Chris Heston, from the San Francisco Giants.

virtually perfect, followed by Francisco Liriano, Chris Sale and Clayton Kewshaw.

Admittedly, it's harder to find actionable items under this measure.

The model is statistically significant at the 99% level, and percentual changes are almost twice as effective in magnitude as frequency adjustments.

Table 1: Test analysis results

	χ^2	Exp. χ^2	Maurer
Coef.	1.967***	1.639	3.574***
R. Std. Err.	0.662	1.560	1.150
Pval	0.004	0.297	0.002
Const_	85.113	84.644	99.718
r^2	0.09	0.01	0.09

1.4 Conclusion

We started with two main questions: Can the standard randomness tests work (or, more properly, pass) a human generated random sequence? and how big is the effect of pitch selection randomness in pitcher performance?

Both get satisfactory answers. For the first one, in particular in Maurer Test, almost every pitcher passes the test. As for the second, we find an effect and the numbers don't jump out of the page as ridiculous. A 1% in the measure of randomness has an effect between $1,9\%$ to 3.5%, the R^2 statistics are around 0.1 which makes sense because pitch speed, location and movement taken together are most certainly more important than sequencing.

Further research avenues are plentiful. Probably the most important one is finding a way to take information from the third test and make it actionable. Probably the easiest way is computing the sum per sequence to see which ones repeat more often or, maybe looking for runs of sequences, something along those lines. Also on the third test, it would be interesting to try with longer multi-year sequences and up the block length.

The second test teases the possibility of a more granular analysis. Instead of binary categories, it could be tested, for example, using 8 or 17 pitch zones (4 for strikes and 4 for balls, and 9 and 8, respectively) or testing each pitch type instead of combining them.

An obvious expansion is to implement different randomness tests, either that examine the same randomness properties and check if the results hold, or that check different properties and derive new actionable information.

Another interesting possibility is to consider this as a catcher's game calling problem. Making this analysis is trivially easy with the tools developed but too time consuming to include in this article.

References

Knuth, Donald E. 1981. *The Art of Computer Programming*. Vol. 2. Addison-Wesley Publishing Company.

Maurer, Ueli M. 1992. A universal statistical test for random bit generators. *Journal of Cryptology*, **5**(2), 89–105.

2 Appendix

Table 2: Complete pitcher tests results

Name	Tm	IP	FIP-	χ^2	Pval	Exp.χ^2	Pval	fT_u	adj fT_u	pval
Mark Buehrle	TOR	198.2	104	189.99	0.00	400.32	0.00	1.99	0.41	0.31
CC Sabathia	NYY	167.1	112	777.78	0.00	795.81	0.00	2.06	0.35	0.34
R.A. Dickey	TOR	214.1	109	2395.71	0.00	2405.83	0.00	1.27	1.13	0.08
Colby Lewis	TEX	204.2	100	105.13	0.00	161.65	0.00	2.26	0.14	0.43
Alfredo Simon	DET	187	118	1521.97	0.00	1559.58	0.00	1.04	1.37	0.05
Scott Kazmir	2TM	183	98	113.56	0.00	256.48	0.00	2.16	0.25	0.38
John Danks	CHW	177.2	108	24.54	0.00	118.78	0.00	2.33	0.07	0.47
Felix Hernandez	SEA	201.2	94	1887.16	0.00	2032.77	0.00	1.44	0.96	0.12
Ubaldo Jimenez	BAL	184	96	828.77	0.00	851.75	0.00	1.71	0.69	0.20
Corey Kluber	CLE	222	73	1852.36	0.00	1858.78	0.00	1.50	0.90	0.14
Edinson Volquez	KCR	200.1	96	2758.51	0.00	2774.78	0.00	1.18	1.23	0.07
Yovani Gallardo	TEX	184.1	96	14.98	0.00	78.53	0.00	2.33	0.07	0.47
David Price	2TM	220.1	68	329.93	0.00	492.23	0.00	2.10	0.31	0.36
Mike Pelfrey	MIN	164.2	99	801.06	0.00	928.06	0.00	1.90	0.50	0.27
Marco Estrada	TOR	181	107	53.54	0.00	98.09	0.00	2.31	0.10	0.45
Carlos Carrasco	CLE	183.2	70	4.04	0.04	174.60	0.00	2.35	0.06	0.47
Wade Miley	BOS	193.2	96	7.47	0.01	244.58	0.00	2.29	0.11	0.45
Jose Quintana	CHW	206.1	79	34.53	0.00	210.29	0.00	2.29	0.11	0.45
Chris Tillman	BAL	173	106	123.36	0.00	194.15	0.00	2.29	0.11	0.45
Chris Archer	TBR	212	73	0.01	0.92	83.12	0.00	2.40	0.00	0.50
Kyle Gibson	MIN	194.2	98	19.34	0.00	350.06	0.00	2.32	0.09	0.46
Jeff Samardzija	CHW	214	101	600.03	0.00	616.45	0.00	1.83	0.57	0.24
Hector Santiago	LAA	180.2	122	600.03	0.00	616.45	0.00	1.18	1.22	0.07
Danny Salazar	CLE	185	89	281.09	0.00	371.18	0.00	2.11	0.29	0.36
Rick Porcello	BOS	172	104	128.47	0.00	175.18	0.00	2.15	0.25	0.38
Chris Sale	CHW	208.2	65	0.30	0.58	48.88	0.00	2.39	0.01	0.49
Erasmo Ramirez	TBR	163.1	95	1.10	0.30	121.72	0.00	2.33	0.07	0.47
Sonny Gray	OAK	208	87	94.30	0.00	213.59	0.00	2.17	0.23	0.39
Collin McHugh	HOU	203.2	87	13.11	0.00	55.80	0.00	2.07	0.33	0.34
Jake Odorizzi	TBR	169.1	97	1720.16	0.00	1755.36	0.00	0.78	1.63	0.02
Trevor Bauer	CLE	176	107	259.98	0.00	394.01	0.00	2.08	0.32	0.35
Yordano Ventura	KCR	163.1	89	118.28	0.00	258.70	0.00	2.24	0.16	0.42
Garrett Richards	LAA	207.1	99	53.18	0.00	194.59	0.00	2.31	0.09	0.46
Dallas Keuchel	HOU	232	71	196.31	0.00	413.25	0.00	2.18	0.22	0.39

Name	Tm	IP	FIP-	χ^2	Pval	Exp.χ^2	Pval	fT_u	adj fT_u	pval
Taijuan Walker	SEA	169.2	103	1508.35	0.00	1514.74	0.00	0.86	1.54	0.03
Wei-Yin Chen	BAL	191.1	99	116.07	0.00	241.38	0.00	2.27	0.13	0.44
Bartolo Colon	NYM	194.2	99	976.30	0.00	1273.26	0.00	1.49	0.91	0.14
A.J. Burnett	PIT	164	89	1869.84	0.00	1889.80	0.00	1.10	1.30	0.06
John Lackey	STL	218	93	1495.39	0.00	1640.31	0.00	1.22	1.19	0.07
Aaron Harang	PHI	172.1	122	213.41	0.00	306.33	0.00	2.09	0.31	0.35
Zack Greinke	LAD	222.2	72	2.92	0.09	149.44	0.00	2.30	0.10	0.45
Dan Haren	2TM	187.1	121	1139.33	0.00	1166.95	0.00	1.44	0.96	0.12
Francisco Liriano	PIT	186.2	84	70.93	0.00	262.04	0.00	2.39	0.01	0.49
Jason Hammel	CHC	170.2	94	0.60	0.44	129.10	0.00	2.34	0.06	0.47
James Shields	SDP	202.1	116	38.98	0.00	349.94	0.00	2.33	0.08	0.46
Jon Lester	CHC	205	75	283.79	0.00	495.16	0.00	2.12	0.28	0.37
Ian Kennedy	SDP	168.1	117	70.72	0.00	249.76	0.00	2.29	0.11	0.45
Max Scherzer	WSN	228.2	72	52.47	0.00	275.97	0.00	2.34	0.07	0.47
Jake Arrieta	CHC	229	60	2229.34	0.00	2237.92	0.00	1.49	0.91	0.13
Lance Lynn	STL	175.1	90	947.67	0.00	1183.61	0.00	1.51	0.90	0.14
Gio Gonzalez	WSN	175.2	80	149.46	0.00	294.73	0.00	2.19	0.21	0.40
Brett Anderson	LAD	180.1	102	3.12	0.08	11.10	0.00	2.33	0.07	0.47
Tyson Ross	SDP	196	78	0.19	0.66	96.42	0.00	2.36	0.04	0.48
Jon Niese	NYM	176.2	114	237.71	0.00	356.62	0.00	2.04	0.36	0.33
Clayton Kershaw	LAD	232.2	52	0.87	0.35	224.48	0.00	2.43	0.03	0.51
Andrew Cashner	SDP	184.2	100	193.34	0.00	375.33	0.00	2.11	0.29	0.36
Jeff Locke	PIT	168.1	104	61.87	0.00	247.85	0.00	2.24	0.16	0.42
Mike Leake	2TM	192	108	1019.66	0.00	1046.89	0.00	1.92	0.49	0.28
Madison Bumgarner	SFG	218.1	79	38.85	0.00	138.69	0.00	2.35	0.05	0.48
Matt Harvey	NYM	189.1	79	109.68	0.00	154.24	0.00	2.20	0.20	0.40
Chris Heston	SFG	177.2	110	3119.06	0.00	3120.79	0.00	0.27	2.13	0.00
Jimmy Nelson	MIL	177.1	101	1017.97	0.00	1041.14	0.00	1.90	0.50	0.27
Jordan Zimmermann	WSN	201.2	98	74.91	0.00	90.12	0.00	2.27	0.13	0.44
Rubby De La Rosa	ARI	188.2	124	79.93	0.00	144.23	0.00	2.26	0.14	0.43
Julio Teheran	ATL	200.2	116	107.72	0.00	240.55	0.00	2.22	0.19	0.41
Gerrit Cole	PIT	208	70	251.05	0.00	414.78	0.00	2.11	0.30	0.36
Anthony DeSclafani	CIN	184.2	92	82.95	0.00	175.29	0.00	2.25	0.15	0.43
Kyle Hendricks	CHC	180	86	2137.52	0.00	2141.30	0.00	1.20	1.20	0.07
Tom Koehler	MIA	187.1	120	14.06	0.00	102.65	0.00	2.35	0.05	0.47
Shelby Miller	ATL	205.1	91	1459.61	0.00	1513.77	0.00	1.22	1.18	0.08
Carlos Martinez	STL	179.2	84	22.69	0.00	66.86	0.00	2.27	0.13	0.44
Jacob deGrom	NYM	191	70	78.70	0.00	180.69	0.00	2.31	0.09	0.45
Michael Wacha	STL	181.1	101	397.35	0.00	454.65	0.00	1.97	0.43	0.30
Alex Wood	2TM	189.2	97	10.37	0.00	181.47	0.00	2.11	0.29	0.36