

Which Teams Would Have Won the 2020 NCAA Men's and Women's Basketball Tournaments?

Chancellor Johnstone, Dan Nettleton

October 24, 2021

Background

- In 2020, the NCAA Division 1 Basketball tournaments (along with several conference tournaments) were cancelled due to COVID-19.
- We want to determine tournament win probabilities through the estimation of team strength.

Question of Interest

Given a collection of game-by-game win probabilities, can we generate tournament win probabilities for each team?

Knockout Tournament Win Probability

- The probability for a team i winning in round J is,

$$q_{iJ} = q_{iJ-1} \left[\sum_{s \in \mathcal{O}_{iJ}} p_{is} q_{sJ-1} \right],$$

- “the probability of team i winning in round J ”
- We can also determine the probability that a team makes the tournament based on their conference and estimated strength.

8-team Knockout Tournament Example

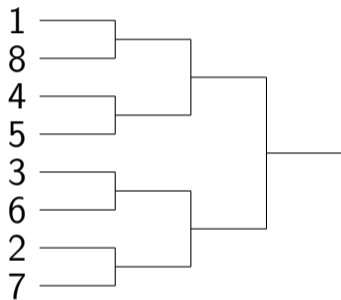


Figure 1: Bracket for eight team single-elimination tournament.

8-team Knockout Tournament Example Probabilities

Table 1: Example win probability matrix for eight team knockout tournament

	1	2	3	4	5	6	7	8
1		0.35	0.23	0.14	0.08	0.05	0.03	0.01
2	0.65		0.35	0.23	0.14	0.08	0.05	0.03
3	0.77	0.65		0.35	0.23	0.14	0.08	0.05
4	0.86	0.77	0.65		0.35	0.23	0.14	0.08
5	0.92	0.86	0.77	0.65		0.35	0.23	0.14
6	0.95	0.92	0.86	0.77	0.65		0.35	0.23
7	0.97	0.95	0.92	0.86	0.77	0.65		0.35
8	0.99	0.97	0.95	0.92	0.86	0.77	0.65	

8-team Knockout Tournament Example

Table 2: Probability of winning each round

	Round 1	Round 2	Final
1	0.99	0.87	0.60
8	0.01	0.00	0.00
4	0.65	0.10	0.03
5	0.35	0.03	0.01
3	0.86	0.33	0.10
6	0.14	0.02	0.00
2	0.95	0.65	0.27
7	0.05	0.01	0.00

Question of Interest

Given the two teams playing, can we estimate margin of victory?

The Harville Method

$$M_{ij} = H + S_i - S_j + e_{ij}$$

- M_{ij} = observed margin of victory (MOV) for team i playing at home against team j (home score – away score).
- H = home field advantage parameter.
- S_i, S_j = strengths of team i and team j , respectively.
- e_{ij} = unobserved, random error associated with team i playing at home against team j .
- H, S_i, S_j estimated from game-outcome data.

Four Team Example

Week	Home Team	Away Team	Home Score	Away Score	MOV
1	3	1	11	16	5
1	4	2	14	7	-7
2	4	1	10	19	9
2	3	2	21	24	3
3	1	4	14	13	-7
3	2	3	3	28	25

$$5 \approx \hat{H} + \hat{S}_3 - \hat{S}_1$$

$$-7 \approx \hat{H} + \hat{S}_4 - \hat{S}_2$$

$$9 \approx \hat{H} + \hat{S}_4 - \hat{S}_1$$

$$3 \approx \hat{H} + \hat{S}_3 - \hat{S}_2$$

$$-7 \approx \hat{H} + \hat{S}_3 - \hat{S}_4$$

$$25 \approx \hat{H} + \hat{S}_2 - \hat{S}_1$$

Example Data Team Strengths

- Team strengths for example data are,

$$\hat{H} = 5.75, \hat{S}_1 = -6.125, \hat{S}_2 = 9.375, \hat{S}_3 = -3.375, \hat{S}_4 = 0.125.$$

2019-2020 Regular Season Team Strengths

Table 3: Ranks for top 10 NCAA men's teams in 2020

Team	$\hat{\theta}$	Rank	AP	NET	KP
Kansas	25.25	1	1	2	1
Gonzaga	22.79	2	2	1	2
Duke	22.31	3	11	6	5
Michigan State	20.54	4	9	7	7
Baylor	20.44	5	5	5	3
Arizona	19.39	6	-	14	19
San Diego State	18.65	7	6	4	6
West Virginia	18.43	8	24	17	10
Ohio State	18.22	9	19	16	8
Dayton	18.07	10	3	3	4

Table 4: Ranks for top 10 NCAA women's teams in 2020

Team	$\hat{\theta}$	Rank	AP	RPI	CSM
Connecticut	27.35	1	5	4	3
South Carolina	23.32	2	1	1	1
Stanford	22.54	3	7	6	7
Oregon	22.41	4	2	2	2
Princeton	21.18	5	22	9	20
Oregon State	20.95	6	14	20	17
Louisville	20.95	7	6	7	6
Baylor	20.64	8	3	4	4
Kentucky	20.13	9	16	24	-
Arizona	18.44	10	12	28	16

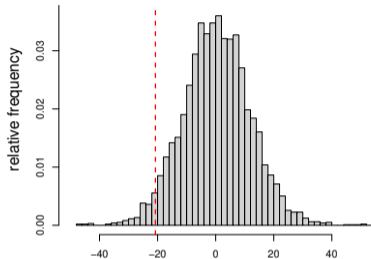
Question of Interest

How do we go from estimating margin of victory to win probabilities?

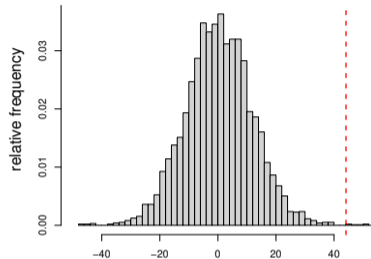
Win Probabilities Through Conformal Prediction

- *Conformal prediction* is an inference methodology based on how well incoming observations conform to previously seen observations.
- The steps are as follows:
 1. For a new observation, specify a hypothesized response value, e.g., a MOV of -25.
 2. Retrain prediction model with observed training data as well as the new observation with hypothesized response.
 3. Calculate a *conformity score* for each observation.
 4. Compare conformity score for new observation to conformity scores for all other observations; higher magnitude conformity scores correspond to hypothesized responses that are less likely.
- Useful for generating prediction intervals around a predicted response value.

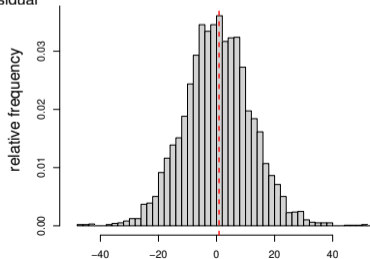
Win Probabilities Through Conformal Prediction



residual

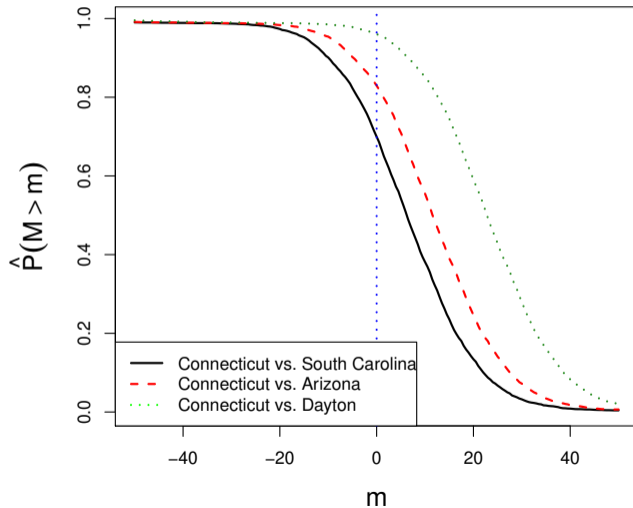


residual



residual

Example with 2019-2020 Regular Season Data



Other Win Probability Methods

We can generate win probability estimates in other ways

- Linear regression under normality assumptions

$$\begin{aligned}\hat{P}(M_{ij} > 0) &= \hat{P}\left(\frac{M_{ij} - \hat{M}_{ij}}{\text{prediction error s.d.}} > \frac{-\hat{M}_{ij}}{\text{prediction error s.d.}}\right) \\ &= 1 - F_{t,n-p}\left(\frac{-\hat{M}_{ij}}{\text{prediction error s.d.}}\right)\end{aligned}$$

- Logistic regression

$$\frac{e^{\text{estimated log odds}}}{1 + e^{\text{estimated log odds}}}$$

March Madness 2020

- Following cancellation, there were 20 unfinished men's conference tournaments; 18 unfinished women's conference tournaments.
- Each team falls into one of five different situations.
- We cannot calculate tournament win probabilities without generating probabilities for making the tournament.
- We simplify bracket generation to make the task of constructing probabilities of making the tournament tractable (S-curve).

S-curve Bracket Simplification

Table 5: March Madness example bracket using S-curve method

Seed	Region			
	1	2	3	4
1	1	2	3	4
2	8	7	6	5
3	9	10	11	12
4	16	15	14	13
5	17	18	19	20
6	24	23	22	21
7	25	26	27	28
8	32	31	30	29
9	33	34	35	36
10	40	39	38	37
11	41	42	43	44
12	48	47	46	45
13	49	50	51	52
14	56	55	54	53
15	57	58	59	60
16	64	63	62	61

Situations for NCAA Teams

1. A team has already made the tournament.
2. A team must win their conference tournament or relies on a small number of teams ranked below them winning their respective conference tournament to make the tournament.
3. A team relies on a small number of teams ranked below them winning their respective conference tournament to make the tournament.
4. A team must win their conference tournament to make the tournament.
5. A team cannot make the tournament.

Probabilities of Making Tournament

Table 6: Situation 3 tournament probabilities

Team	Overall Rank	Probability
Yale	40	0.995
Michigan	41	0.917
North Carolina	42	0.618
Central Florida	43	0.232
St. John's	44	0.031

Table 7: Situation 4 tournament probabilities

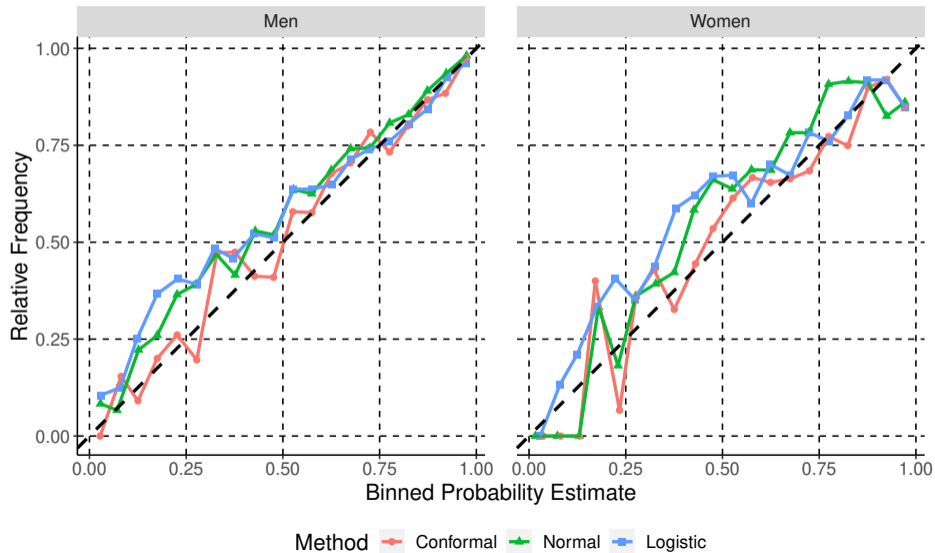
Team	Overall Rank	Probability
Old Dominion	48	0.239
Northern Iowa	49	0.173
Oklahoma	52	0.011
Western Kentucky	54	0.199
West Virginia	57	0.038
Rice	58	0.248
Middle Tennessee	63	0.226

Tournament Win Probabilities

Table 8: Win probabilities given example bracket for top ranked women's teams

Team	Probability
Connecticut	0.289
South Carolina	0.127
Stanford	0.104
Oregon	0.094
Princeton	0.064
Oregon State	0.061
Louisville	0.058
Baylor	0.043
Kentucky	0.032
Arizona	0.021

Win Probability Calibration



Win Probability Calibration by Season

	ECE			MCE		
Season	Conformal	Linear	Logistic	Conformal	Linear	Logistic
2014-2015	0.077	0.111	0.104	0.824	0.412	0.423
2015-2016	0.103	0.101	0.082	0.711	0.392	0.817
2016-2017	0.071	0.115	0.110	0.826	0.817	0.890
2017-2018	0.072	0.132	0.101	0.717	0.465	0.581
2018-2019	0.067	0.118	0.090	0.329	0.245	0.405
2019-2020	0.121	0.147	0.134	0.375	0.331	0.325

	ECE			MCE		
Season	Conformal	Linear	Logistic	Conformal	Linear	Logistic
2014-2015	0.086	0.097	0.088	0.920	0.305	0.672
2015-2016	0.083	0.094	0.071	0.314	0.276	0.242
2016-2017	0.058	0.079	0.105	0.302	0.325	0.434
2017-2018	0.064	0.071	0.057	0.819	0.793	0.828
2018-2019	0.057	0.061	0.068	0.126	0.142	0.195
2019-2020	0.052	0.098	0.114	0.189	0.247	0.289

Table 9: Calibration comparison for all methods within each women's season (top) and men's season (bottom)




- Data
 - ▶ Inconsistencies with naming convention.
 - ▶ Women's data is much more difficult to obtain.
- Model
 - ▶ Does not account for: momentum, injuries, match-ups, etc.
 - ▶ We (over)simplify the bracket generation process.
 - ▶ Double-dip on overall team strength.

Future Work

- Utilize more complex modeling approaches to margin of victory prediction, e.g., fused lasso approach.
- Apply to betting scenarios.
- Incorporate more accurate portrayal of existing selection process.
- Extend to other professional and amateur leagues.

Questions?

References

-  Edwards, C. T. (1991).
The combinatorial theory of single-elimination tournaments.
PhD thesis, Montana State University-Bozeman, College of Letters & Science.
-  Harville, D. (1977).
The use of linear-model methodology to rate high school or college football teams.
Journal of the American Statistical Association, 72(358):278–289.
-  Vovk, V., Shen, J., Manokhin, V., and Min-ge, X. (2019).
Nonparametric predictive distributions based on conformal prediction.
Machine Learning, 108(3):445–474.

Conformal Predictive Distributions

We formalize this using conformal predictive distributions to generate estimates of win probability for the home team in an upcoming match-up as,

$$\pi(y_c, \tau) = \begin{cases} \frac{i+\tau}{n+1} & R_{n+1}(y_c) \in (R_{(i)}(y_c), R_{(i+1)}(y_c)) \text{ for } i \in 0, \dots, n \\ \frac{i-1+2\tau}{n+1} & R_{n+1}(y_c) = R_{(i)}(y_c) \text{ for } i \in 0, \dots, n \end{cases},$$

where $R_i(y_c) = y_i - \hat{y}_i(y_c)$ and $R_{n+1}(y_c) = y_c - \hat{y}_{n+1}(y_c)$, the index (i) represents that i -th order statistic of the collection of conformity scores and τ is a $U(0, 1)$ random variable.

CPD Four Team Example

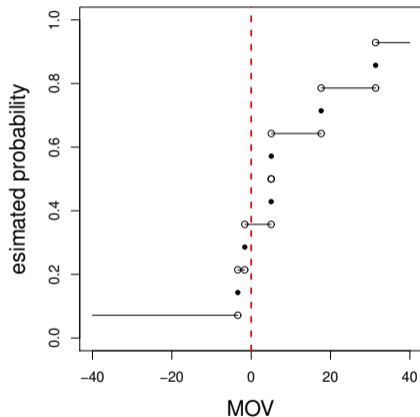
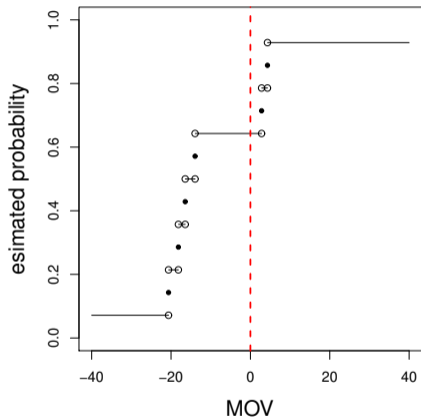


Figure 3: MOV conformal predictive distributions for team 1 vs. team 2 (left) and team 3 vs. team 4 (right).

2020 Conference Champions

Table 10: Winners of completed conference tournaments for NCAA women's basketball

Conference	Winner
Atlantic-10	Dayton
ACC	North Carolina St.
American	Connecticut
Big East	DePaul
Big Ten	Maryland
Horizon	IUPUI
Ivy League	Princeton
Mountain West	Boise St.
Ohio Valley	Southeast Missouri St.
Pac-12	Oregon
SEC	South Carolina
Southern	Samford
Summit	South Dakota
WCC	Portland

Table 11: Winners of completed conference tournaments for NCAA men's basketball

Conference	Winner
ASUN	Liberty
Big South	Winthrop
CAA	Hofstra
Horizon League	Northern Kentucky
Missouri Valley	Bradley
Mountain West	Utah St.
Northeast	Robert Morris
Ohio Valley	Belmont
Patriot League	Boston
Southern	E. Tennessee St.
The Summit	North Dakota St.
WCC	Gonzaga

Princeton was awarded the automatic bid to the tournament based on regular season performance, not by winning the Ivy League conference tournament.

NCAA Men's Situations

Table 12: Situations for men's teams ranked from thirty-three to sixty-four

Situation	Teams
1	Utah St., Florida, Auburn
2	Alabama, Providence, Syracuse, Mississippi St., Memphis, NC St., Arizona St., Rhode Island, Virginia, USC, Oklahoma St., Tennessee, Notre Dame, Richmond, Yale, Clemson, Connecticut
3	Indiana, LSU, Arkansas, Oklahoma, Wichita St., Cincinnati
4	Stanford
5	all other teams

Men's Probabilities of Making Tournament

Table 13: Probability of winning conference tournament

Team	Overall Rank	Probability
Alabama	45	0.055
Providence	46	0.078
Syracuse	47	0.063
Mississippi St.	48	0.101
Memphis	49	0.102
NC St.	50	0.046
Arizona St.	51	0.158
Rhode Island	52	0.157
Virginia	53	0.068
USC	56	0.069
Oklahoma St.	57	0.025
Tennessee	58	0.038
Notre Dame	60	0.058
Richmond	61	0.128
Yale	62	0.464
Clemson	63	0.040
Connecticut	64	0.070

Table 14: Probability of making tournament for men's teams in Situations 3 and 4

Team	Overall Rank	Probability
Indiana	36	0.999
LSU	37	0.994
Arkansas	38	0.956
Stanford*	39	0.801
Oklahoma	40	0.522
Wichita St.	41	0.350
Cincinnati	42	0.224

Men's Basketball Example Bracket

Table 15: Bracket 1 for men's tournament and respective seeding

Seed	Region 1	Region 2	Region 3	Region 4
1	Kansas	Gonzaga	Duke	Michigan St.
2	West Virginia	San Diego St.	Arizona	Baylor
3	Ohio St.	Dayton	Maryland	Michigan
4	Creighton	Texas Tech	Florida St.	Louisville
5	BYU	Oregon	Seton Hall	Villanova
6	Marquette	Iowa	Houston	Penn St.
7	Colorado	Purdue	Kentucky	Wisconsin
8	Butler	Rutgers	Minnesota	Illinois
9	Utah St.	Florida	Auburn	Indiana
10	Oklahoma	Stanford	Arkansas	LSU
11	Wichita St.	Cincinnati	Xavier/Providence	Saint Mary's (CA)/Alabama
12	Liberty	North Texas	E. Tennessee St.	Yale
13	Vermont	Akron	No. Colorado	Belmont
14	Bradley	UC Irvine	Stephen F. Austin	Texas St.
15	New Mexico St.	Hofstra	Winthrop	North Dakota St.
16	Prairie View/Robert Morris	Siena/Norfolk St.	Boston	Northern Kentucky

Men's Basketball Tournament Win Probabilities

Table 16: Win probabilities given exemplar brackets for top ranked men's teams

Team	Probability
Kansas	0.269
Gonzaga	0.144
Duke	0.123
Michigan State	0.067
Baylor	0.063
Arizona	0.043
San Diego State	0.031
West Virginia	0.027
Ohio State	0.023
Dayton	0.023

Probability of Being At-Large Bid

- For team ranked i to make the tournament as an at large bid we have to look at every team below them and enumerate all possible combinations of conference outcomes
- Results in the following probability distribution,

$$P(L_i \leq l) = \sum_{m=0}^l \left\{ \sum_{A \in \mathcal{F}_m} \prod_{s \in A} p_s \prod_{s \in A^c} (1 - p_s) \right\},$$