A Markov Chain Analysis of NCAA Football Overtime

Will, Rebecca, Eljo, Jackson

01 Introduction

- NCAA overtime rules
- Modelling of college football

Introduction

NCAA OT Rules Before 2021

- Drive from the 25-yard line until the fourth overtime
- Alternating 2-point conversion plays starting fifth overtime

NCAA OT Rules After 2021

- Mandatory 2-point conversion after touchdowns starting in the second overtime
- Alternating 2-point conversion plays starting third overtime

How does this affect college football teams' strategy?

- Strategy: the team that wins the coin flip defers possession
  - Knowing what the other team did when it is their turn to drive
  - Advantage?

Current OT Rules (since 2021)

NCAA Overtime

01
First OT

Starting team drives from 25 yards, opposing team responds

02
Second OT

Same as First OT, but teams are required to go for 2 on touchdown

03
Third OT Onwards

Alternating two-point conversions

1 — 2 — 3+

- Conventional strategy: Go second!

Goldner (2012)

- Markov Chain with nine absorbing states
- 349 states based on down, distance, and yardline
- Expected points for any field position using absorption probabilities
- Full football drives vs overtime rules

Wilson (2020)

- Logistic regression and decision trees
- Includes new OT rules
- **No significant advantage to the conventional ”defense first” strategy in overtime**
- Overtime as a single process vs modeling via states

Wilson, Rick L. ”College Football Overtime Outcomes: Implications for In-Game Decision-Making.” *Frontiers in Artificial Intelligence, vol. 3,* 2020, doi:10.3389/frai.2020.00061.

Goldner, Keith. “A Markov Model of Football: Using Stochastic Processes to Model a Football Drive.” *Journal of Quantitative Analysis in Sports 8* (2012): n. pag.

02

The Model

- ➢ Necessary Assumptions
- ➢ States and Transition Probabilities
- ➢ The Matrix

Assumptions (Data from 2021-2023)

1. The probability of a defensive touchdown or *safety* is negligible

2. No team will attempt a two point conversion in the first overtime

3. Teams will not miss extra points

4. A team with the opportunity to win the game with any score will simply attempt a field goal from inside 25 yards

Markov States

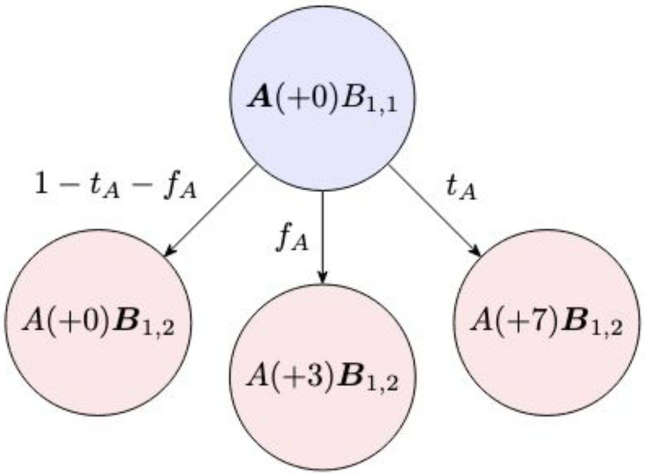
- States capture 4 essential characteristics
  1. Team with Possession (A or B)
  2. Score Differential (many possibilities)
  3. Overtime Period (1, 2, or 3)
  4. Period Frame (1 or 2)

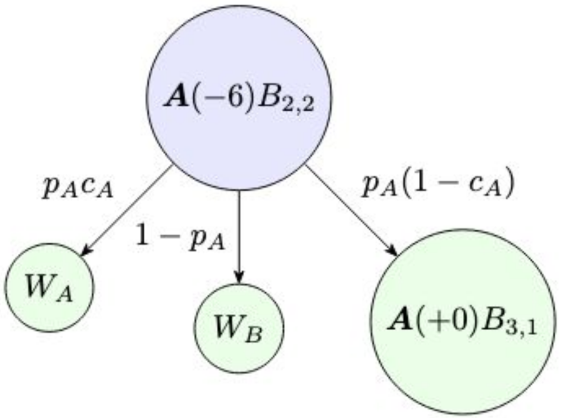
Team with Possession	Score Differential(for A)	Overtime Period	Overtime Frame	Notation
A	+0	1	1	$\boldsymbol{A}(+0)B_{1,1}$
B	+0	1	2	$A(+0)\boldsymbol{B}_{1,2}$
B	3	1	2	$A(+3)\boldsymbol{B}_{1,2}$
B	7	1	2	$A(+7)\boldsymbol{B}_{1,2}$
B	+0	2	1	$A(+0)\boldsymbol{B}_{2,1}$
A	+0	2	2	$\boldsymbol{A}(+0)B_{2,2}$
A	−3	2	2	$\boldsymbol{A}(-3)B_{2,2}$
A	−6	2	2	$\boldsymbol{A}(-6)B_{2,2}$
A	−8	2	2	$\boldsymbol{A}(-8)B_{2,2}$
A	+0	3	1	$\boldsymbol{A}(+0)B_{2,2}$
B	+0	3	2	$A(+0)\boldsymbol{B}_{3,2}$
B	+2	3	2	$A(+2)\boldsymbol{B}_{3,2}$
B	+0	3	1	$A(+0)\boldsymbol{B}_{3,1}$
A	−2	3	2	$\boldsymbol{A}(-2)B_{3,2}$
B	+0	3	2	$\boldsymbol{A}(+0)B_{3,2}$
A Victory	N/A	N/A	N/A	W_A
B Victory	N/A	N/A	N/A	W_B

Transition Probabilities

- ***t*:** team scores a **touchdown** given a first down from the 25 yard line.
- ***f*:** team settles for a **field goal** given a first down from the 25 yard line.
- ***p*:** team scores a **touchdown** given a first down from the 25 yard line, and that they **must score** a touchdown.
- ***c*:** probability of a successful **two-point conversion** attempt.
- **k:** **field goal** success rate between 20-25 yard line.

State Transition Examples





Markov Matrix

	$\boldsymbol{A}(+0)B_{1,1}$	$A(+0)\boldsymbol{B}_{1,2}$	$A(+3)\boldsymbol{B}_{1,2}$	$A(+7)\boldsymbol{B}_{1,2}$	$A(+0)\boldsymbol{B}_{2,1}$	$\boldsymbol{A}(+0)B_{2,2}$	$\boldsymbol{A}(-3)B_{2,2}$	$\boldsymbol{A}(-6)B_{2,2}$	$\boldsymbol{A}(-8)B_{2,2}$	$\boldsymbol{A}(+0)B_{3,1}$	$A(+2)\boldsymbol{B}_{3,2}$	$A(+0)\boldsymbol{B}_{3,2}$	$A(+0)\boldsymbol{B}_{3,1}$	$\boldsymbol{A}(-2)B_{3,2}$	$\boldsymbol{A}(+0)B_{3,2}$	$A_{\mathbf{W}}$	$B_{\mathbf{W}}$
$\boldsymbol{A}(+0)B_{1,1}$	0	$1-t_A-f_A$	f_A	t_A	0	0	0	0	0	0	0	0	0	0	0	0	0
$A(+0)\boldsymbol{B}_{1,2}$	0	0	0	0	$1-f_B$	0	0	0	0	0	0	0	0	0	0	0	f_B
$A(+3)\boldsymbol{B}_{1,2}$	0	0	0	0	f_B	0	0	0	0	0	0	0	0	0	0	$1-t_B-f_B$	t_B
$A(+7)\boldsymbol{B}_{1,2}$	0	0	0	0	p_B	0	0	0	0	0	0	0	0	0	0	$1-p_B$	0
$A(+0)\boldsymbol{B}_{2,1}$	0	0	0	0	0	$1-f_B-t_B$	f_B	$(1-c_B)t_B$	c_Bt_B	0	0	0	0	0	0	0	0
$\boldsymbol{A}(+0)B_{2,2}$	0	0	0	0	0	0	0	0	0	$1-f_A$	0	0	0	0	0	f_A	0
$\boldsymbol{A}(-3)B_{2,2}$	0	0	0	0	0	0	0	0	0	f_A	0	0	0	0	0	t_A	$1-t_A-f_A$
$\boldsymbol{A}(-6)B_{2,2}$	0	0	0	0	0	0	0	0	0	$p_A(1-c_A)$	0	0	0	0	0	p_Ac_A	$1-p_A$
$\boldsymbol{A}(-8)B_{2,2}$	0	0	0	0	0	0	0	0	0	p_Ac_A	0	0	0	0	0	0	$1-p_Ac_A$
$\boldsymbol{A}(+0)B_{3,1}$	0	0	0	0	0	0	0	0	0	0	c_A	$1-c_A$	0	0	0	0	0
$A(+2)\boldsymbol{B}_{3,2}$	0	0	0	0	0	0	0	0	0	0	0	0	c_B	0	0	$1-c_B$	0
$A(+0)\boldsymbol{B}_{3,2}$	0	0	0	0	0	0	0	0	0	0	0	0	$1-c_B$	0	0	0	c_B
$A(+0)\boldsymbol{B}_{3,1}$	0	0	0	0	0	0	0	0	0	0	0	0	0	c_B	$1-c_B$	0	0
$\boldsymbol{A}(-2)B_{3,2}$	0	0	0	0	0	0	0	0	0	c_A	0	0	0	0	0	0	$1-c_A$
$\boldsymbol{A}(+0)B_{3,2}$	0	0	0	0	0	0	0	0	0	$1-c_A$	0	0	0	0	0	c_A	0
$A_{\mathbf{W}}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
$B_{\mathbf{W}}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Absorption Probabilities

$$P(\text{A Win}) = -\frac{1}{2}(-f^4 - f^3 - f^2k^2 + f^2kt + f^2t^2 + f^2t + 2f^2 - fk^2t + fk^2 - 2fkpt + fkt^2 + fkt$$
$$- 2fk + fpt - ft^2 + 2ft - f - kpt^2 + kpt + kt^3 - kt^2 - kt + k - p^2t^2 + pt^3 + pt^2 - t^3 + t^2 - t - 1)$$

$$P(\text{B Win}) = \frac{1}{2}(-f^4 - 2f^3t + f^3 + f^2k^2 - 3f^2kt - f^2t^2 + 3f^2t - 2f^2 + fk^2t - fk^2 + 2fkpt - 3fkt^2 + fkt$$
$$+ 2fk - 2fpt^2 - fpt + 3ft^2 - 4ft + f + kpt^2 - kpt - kt^3 + kt^2 + kt - k + p^2t^2 - pt^3 - pt^2 + t^3 - t^2 + t - 1)$$

- Assumes equal rates by team.
- No c!

Absorption Probabilities

$$
\begin{aligned}
P(\text{A win}) = \frac{1}{c_A + c_B - 2c_Ac_B}&(2c_Ac_Bf_A^2f_Bt_B - 2c_Ac_Bf_A^2f_B + c_Ac_Bf_A^2t_At_B - c_Ac_Bf_A^2t_A \\
&- c_Ac_Bf_A^2t_B + c_Ac_Bf_A^2 - 2c_Ac_Bf_Af_B^2t_A + 2c_Ac_Bf_Af_B^2 - 2c_Ac_Bf_Af_Bp_At_B \\
&- c_Ac_Bf_Af_Bt_A^2 + 3c_Ac_Bf_Af_Bt_At_B - c_Ac_Bf_Af_Bt_A + c_Ac_Bf_Af_Bt_B - c_Ac_Bf_Ap_At_At_B \\
&+ c_Ac_Bf_Ap_At_B + c_Ac_Bf_Ap_Bt_At_B - c_Ac_Bf_Ap_Bt_A + 2c_Ac_Bf_At_A^2t_B - 2c_Ac_Bf_At_A^2 \\
&- 2c_Ac_Bf_At_At_B + 2c_Ac_Bf_At_A + 2c_Ac_Bf_At_B - 2c_Ac_Bf_A - c_Ac_Bf_B^2t_A^2 \\
&+ 2c_Ac_Bf_B^2t_A - c_Ac_Bf_B^2 - c_Ac_Bf_Bp_At_At_B \\
&+ c_Ac_Bf_Bp_At_B - c_Ac_Bf_Bp_Bt_A^2 + c_Ac_Bf_Bp_Bt_A - c_Ac_Bf_Bt_A^3 + c_Ac_Bf_Bt_A^2t_B + 2c_Ac_Bf_Bt_A^2 - 3c_Ac_Bf_Bt_A \\
&- c_Ac_Bf_Bt_B + 2c_Ac_Bf_B - c_Ac_Bp_Ap_Bt_At_B - c_Ac_Bp_At_A^2t_B + 2c_Ac_Bp_At_At_B - c_Ac_Bp_At_B \\
&+ c_Ac_Bp_Bt_A^2t_B - c_Ac_Bp_Bt_A^2 + c_Ac_Bp_Bt_At_B + c_Ac_Bp_Bt_A + c_Ac_Bt_A^3t_B - c_Ac_Bt_A^3 - c_Ac_Bt_A^2t_B + c_Ac_Bt_A^2 \\
&- c_Ac_Bt_At_B - c_Ac_Bt_A + c_Ac_Bt_B - c_Ac_B + 2c_Af_A^2f_B^2 + c_Af_A^2f_Bt_A - c_Af_A^2f_B + 3c_Af_Af_B^2t_A - 3c_Af_Af_B^2 \\
&+ 2c_Af_Af_Bp_At_B + c_Af_Af_Bp_Bt_A + 2c_Af_Af_Bt_A^2 - 4c_Af_Af_Bt_A - 2c_Af_Af_Bt_B + 3c_Af_Af_B \\
&+ c_Af_Ap_At_At_B - c_Af_Ap_At_B - c_Af_At_At_B + c_Af_At_A + c_Af_B^2t_A^2 - 2c_Af_B^2t_A + c_Af_B^2 \\
&+ c_Af_Bp_At_At_B - c_Af_Bp_At_B + c_Af_Bp_Bt_A^2 - c_Af_Bp_Bt_A + c_Af_Bt_A^3 - 3c_Af_Bt_A^2 \\
&- c_Af_Bt_At_B + 4c_Af_Bt_A + c_Af_Bt_B - 2c_Af_B + c_Ap_Ap_Bt_At_B + c_Ap_At_A^2t_B \\
&- 2c_Ap_At_At_B + c_Ap_At_B - c_Ap_Bt_At_B - c_At_A^2t_B + c_At_A^2 + 2c_At_At_B - c_At_A - c_At_B \\
&+ c_A - 2c_Bf_A^2f_B^2 - c_Bf_A^2f_Bt_A - 2c_Bf_A^2f_Bt_B \\
&+ 3c_Bf_A^2f_B - c_Bf_A^2t_At_B + c_Bf_A^2t_A + c_Bf_A^2t_B - c_Bf_A^2 - c_Bf_Af_B^2t_A \\
&+ c_Bf_Af_B^2 - c_Bf_Af_Bp_Bt_A - c_Bf_Af_Bt_A^2 - 3c_Bf_Af_Bt_At_B + 5c_Bf_Af_Bt_A \\
&+ c_Bf_Af_Bt_B - 3c_Bf_Af_B - c_Bf_Ap_Bt_At_B + c_Bf_Ap_Bt_A - 2c_Bf_At_A^2t_B + 2c_Bf_At_A^2 \\
&+ 3c_Bf_At_At_B - 3c_Bf_At_A - 2c_Bf_At_B + 2c_Bf_A - c_Bf_Bt_A^2t_B + c_Bf_Bt_A^2 \\
&+ c_Bf_Bt_At_B - c_Bf_Bt_A - c_Bp_Bt_A^2t_B + c_Bp_Bt_A^2 \\
&- c_Bp_Bt_A - c_Bt_A^3t_B + c_Bt_A^3 + 2c_Bt_A^2t_B - 2c_Bt_A^2 - c_Bt_At_B + 2c_Bt_A) \\
P(\text{B win}) = 1 - P(\text{A win})&
\end{aligned}
$$

- Team-specific rates.

Data Collection/Rate Estimation

- **Data Source:** R library “cfbfastR”, play-by-play
  - 2021-2023 seasons of Power Five play (since OT rule change)
- **R Analysis:**
  - Evaluated assumptions
  - Calculated average success rates across league of touchdown, field goal, etc. to use in Markov Chain

Average Transition Probabilities

- **Sample Space:** { touchdown, field goal, neither }
    - Drives mapping to sample space include a first down from 23-27 yards to goal.
- **Periods:**
    - **t** and **f** are from 1st-3rd periods only to avoid late game confounds.
- **Estimating p: touchdowns to tie the game**
    - Two approaches: using historical OT data only vs 4th period data

Approaches to *p*: touchdowns to tie the game

- **Question:**
    - How do we define "under-pressure" touchdowns to find the probability that a team scores a touchdown when one is needed to have a chance at winning the game?
- **First Approach:**
    - Using all historical overtime data (2014-2023), we found the probability that a team with the second possession in the 1st or 2nd OT, down by a touchdown (6-8 points) scored a touchdown
- **Second approach:**
    - Using fourth period data (2021-2023), we found the probability that a team down by 4-8 points, or more than 12 points, in a drive with a first down on the 23-27 yard line, with 1-3 minutes left in the game, scored the touchdown

03 Analysis

- ➢ Analysis of Model
- ➢ Discussion

Possession Advantage (2 approaches)

- **Approach 1:** p calculated using historical OT data

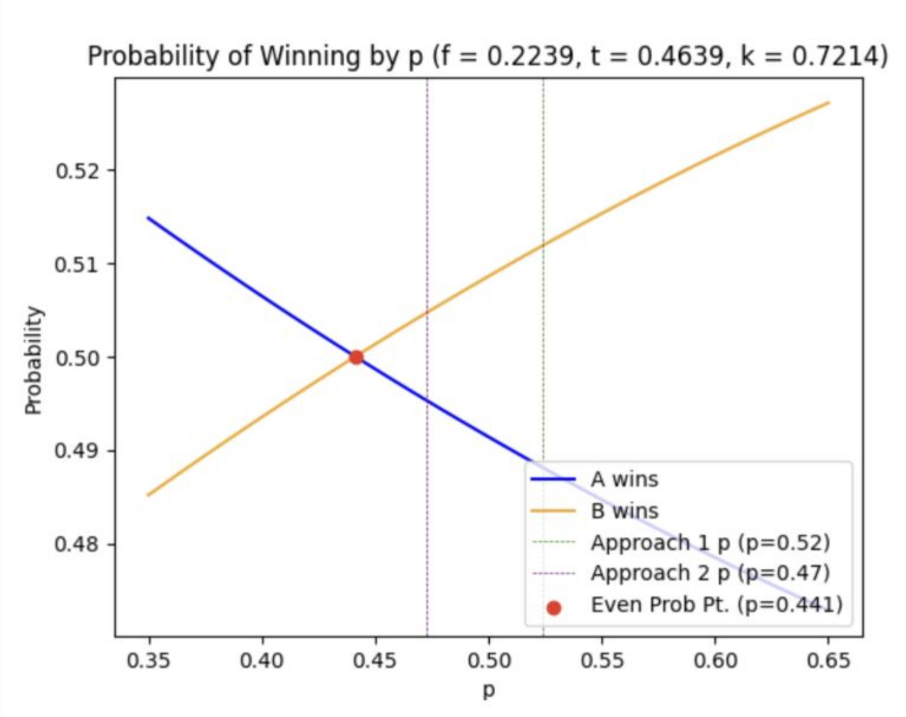
$$P(\text{A wins}|t = 0.4639;\ f = 0.2239;\ \underline{p = 0.5238;}\ k = 0.7214) = 48.81\% \quad \implies P(\text{B wins}) \quad = 51.19\%$$

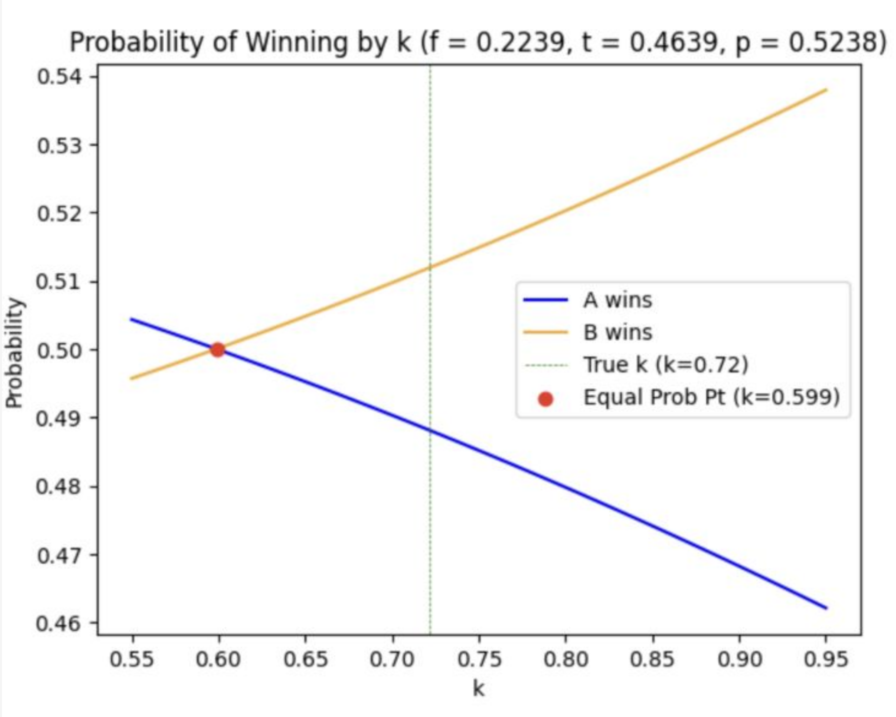
- **Approach 2:** p calculated using recent 4th quarter data

$$P(\text{A wins}|t = 0.4629\ ;\ f = 0.2229\ \underline{;p = 0.4727;}\ k = 0.7214) = 49.53\% \quad \implies P(\text{B wins}) \quad = 50.47\%$$

- Team B benefits in both cases!

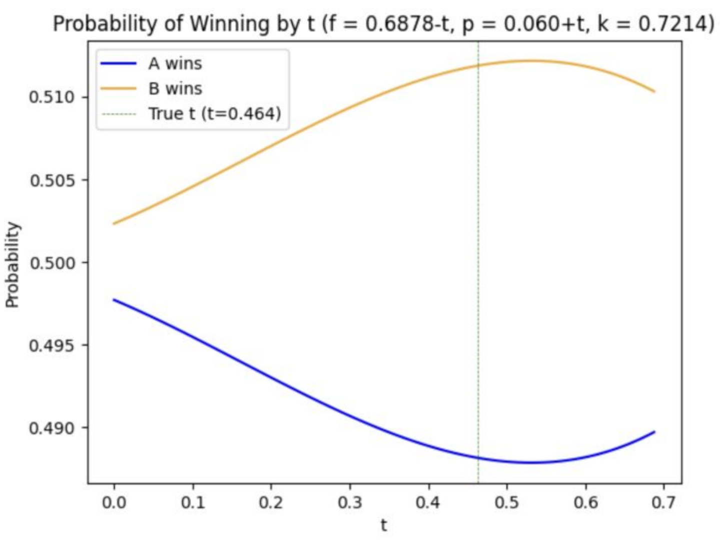
Marginal Effects of “Response Parameters”



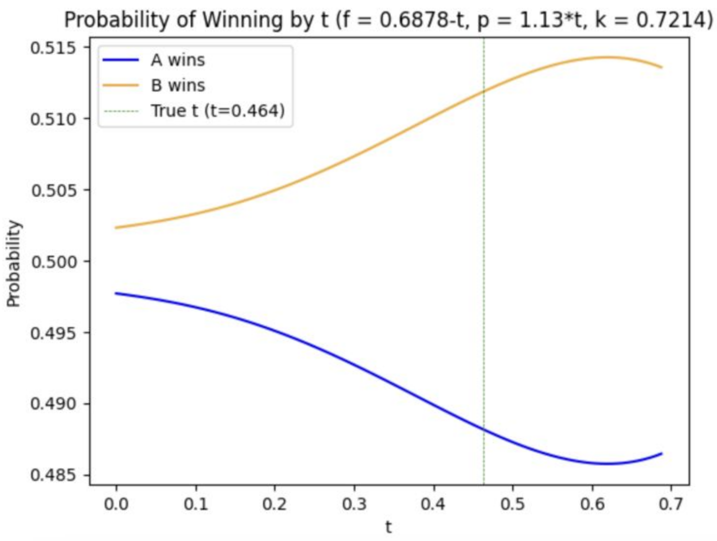


How does TD/FG Percentage affect win rates?

- Assuming constant scoring rate: ie, f + t is fixed.
- 2 methods for approximating p: **constant p - t** difference, **constant p/t ratio**. Similar results.



(a) Constant Difference $p - t$ Approach



(b) Constant Ratio p/t Approach

The 2-pt Conversion Question

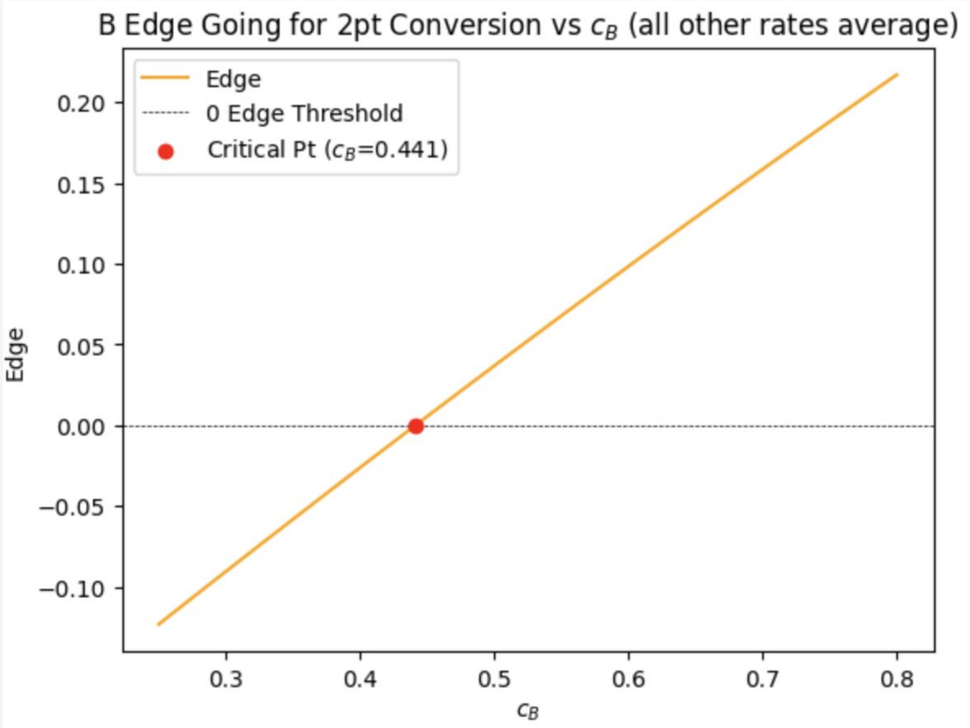
- **Scenario**: Team B down 7 in the second frame of 1st OT. Should they go for 2?
- Probability in question (assumes an average Team A):

$$P_B = P\left(W_B | A(+0)\boldsymbol{B}_{2,1} \textbf{ and } t_A = 0.4639;\ f_A = 0.2239;\ p_A = 0.5238;\ k_A = 0.7214\right)$$

$$\begin{aligned}\textbf{Team B Edge} &= c_B - P_B \\ &= .236 c_B t_B + \\ &\frac{c_B\left(-.0220 c_B t_B + .0248 f_B + .0186 t_B - .126\right)}{.009 c_B^2 + .0842 c_B - .794} + .312 f_B + .476 t_B \\ &+ \frac{.454\left(-.006 c_B^3 t_B + .005 c_B^2 f_B - .018 c_B^2 t_B - .026 c_B^2 + .025 c_B f_B + .019 c_B t_B - .126 c_B\right)}{.009 c_B^2 + .084 c_B - .794}\end{aligned}$$

The Edge Threshold

- Analysis assumed only varying B's conversion rate.
- Could be replicated for specific teams if exact rates are known.



Conclusions

- The deferring team does have a probabilistic advantage (~51.19%).
    - Mainly due to ‘response’ parameters, p and k.
- Responding team should generally go for 2 in the 1st overtime if down 7.
- Our model can be applied to further strategic questions (eg. tight fourth downs).

References

- https://www.espn.com/college-football/story/_/id/39111711/what-ncaa-college-football-rules
- Goldner, Keith. 2012. "A Markov Model of Football: Using Stochastic Processes to Model a Football Drive." *Journal of Quantitative Analysis in Sports* 8: n. pag. https://doi.org/10.1515/1559-0410.1400
- Wilson, Rick L. 2020. "College Football Overtime Outcomes: Implications for In-Game Decision-Making." *Frontiers in Artificial Intelligence* 3. https://doi.org/10.3389/frai.2020.00061.
- Gilani, Saiem, Akshay Easwaran, Jared Lee, and Eric Hess. 2021. "cfbfastR: The SportsDataverse's R Package for College Football Data." https://cfbfastR.sportsdataverse.org/.