**STATS 50: Mathematics of Sport** **Spring 2019**

# Week 2 – Overview of Stochastic Processes

*Lecturer: Maxime Cauchois*

***Warning:*** *these notes may contain factual errors*

Most of these notes are adapted from a paper from D.Percy [1].

## 1 Introduction to stochastic processes

Stochastic processes are natural tools to model the evolution of a game or a team's performance, as they encompass both the progression of the studied event as well as its inherent randomness. Additionally, they allow practitioners and gamblers to leverage all the information present up to a given time for their predictions and bets, creating a perfect setting for online learning and real-time decision making.

There are several approaches to deal with historical data gathered from past years results. The first and most simple one is to consider that all previous events $X_i$, $i = 1, ..., n$ are drawn independently from the same distribution $\{\mathcal{P}_\theta, \theta \in \Theta\}$, where $\theta$ is a parameter which lies in some space $\Theta$: one usually refers to this as the i.i.d. assumption (for independent and identically distributed), where all observations are assumed to have an identical importance, and are supposed to bring the same amount of information for future inference. Future results will then be predicted using the distribution $\mathcal{P}_{\hat{\theta}}$, where $\hat{\theta}$ is the estimator of $\theta$ inferred from past data.

However, when your data have a spatial, or even better a temporal component, it seems more accurate to keep this trait of your data, and to find a model which accurately reflects this additional information. Stochastic processes are especially well equipped for this task, and will enable us to allow more weight to more recent data than historical data, for instance. This leads to potentially more optimal and adaptive decisions related to the strategy to adopt or the final outcome of a game.

## 2 Discrete-time stochastic processes

### 2.1 General Presentation

Mathematically, a discrete-time stochastic process is a sequence $\{X_n\}_{n\geq 0}$ of random variables lying in the same space $E$, where $n = 0, 1, 2, ...$ represents the time of the observation, and introduces an order inside our variables. $E$ can take several forms, depending of the nature of your data.

- $E = \mathbb{R}^d$ if each observation at time $n$ consists of $d$ different real numbers, each of those coming from a continuous distribution.
- $E = \{1, ..., m_1\} \times ... \times \{1, ..., m_d\}$ if an observation is has $d$ categorical components, each one of them taking $m_i$ different possible values
- $E$ can be a mixture of both cases above, containing both categorical and continuous elements.

Contrary to a sequence of i.i.d observations (which can also be represented as a stochastic process), the main interest that resides within the study of stochastic processes comes from the joint distribution of the process $\{X_n\}$. In other terms, our goal is to determine how our variables correlate from one another, so as to infer the likelihood of the next observation (i.e the one you are interested in predicting or studying).

## 2.2 Examples of stochastic processes

It is very easy to think of sports instances in which stochastic processes appear quite naturally.

- One might for instance be interested in the evolution of a score, in which case $n$ then represents the number of scoring events up to this point, and $X_n$ encodes the actual score at this moment
- If one wants to look at a bigger picture, $n$ can become the number of games already played at a given moment of the season, while $X_n$ will represent the current ranking or evaluation of one (or several) teams.

## 2.3 Markov Chains

In this section, we will especially focus on stochastic processes which take their values in a discrete space taking at most a countable number of values (but very often finite), and for which a very nice property of "independence from the past given the present" holds.

More precisely, a Markov chain is a discrete-time stochastic process such that for any $n \geq 0$ and any sequence of states $(x_0, ..., x_n, x_{n+1}) \in E^{n+2}$, the following holds:

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, ..., X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n) \tag{1}$$

This last property, also known as the Markov property, states that the future of our stochastic process does not depend on its past evolution (i.e what happened at steps $0, ..., n-1$) from the moment you know in which state it has arrived at time $n$.

This property is a relaxation of the i.i.d property, where we were requiring all $X_i$ to be independent, since here $X_{n+1}$ is allowed to depend on the past, but with the restriction of depending only on it through $X_n$.

The distribution of a finite Markov chain is very easy to specify. Indeed, suppose that $E = \{1, ..., m\}$, i.e there are $m$ different states in which our Markov chain is bound to evolve. Then the evolution of our Markov chain can be described with a transition matrix $P \in \mathbb{R}^{m \times m}$ such that for any $i, j \in [m]$:

$$P_{i,j} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

This matrix, alongside with the initial distribution of $X_0$, suffices to describe the entire distribution of $(X_n)_{n \geq 0}$, thanks to the Markov property (1). Indeed, for any sequence of states $i_0, ..., i_n \in E$, it is straightforward to show that:

$$\mathbb{P}(X_n = i_n, ..., X_0 = i_0) = \pi_{i_0} P_{i_0,i_1} \ldots P_{i_{n-1},i_n}$$

where $\pi_{i_0}$ stands for the probability that $X_0 = i_0$.

In particular, if $X_0 = i$, then $P^n(i, j)$ is precisely the probability that $X_n = j$, i.e $P^n$ yields the distribution of $X_n$. To prove this fact, one has to look at all the different possible paths leading from $i$ to $j$ in $n$ steps:

$$\begin{aligned}\mathbb{P}(X_n = j \mid X_0 = i) &= \sum_{i_1,\dots,i_{n-1}} \mathbb{P}(X_n = j, X_{n-1} = i_{n-1}, \dots, X_1 = i_1 \mid X_0 = i) \\ &= \sum_{i_1,\dots,i_{n-1}} P_{i,i_1} \dots P_{i_{n-1},j} \\ &= P^n(i,j)\end{aligned}$$

The easiness to specify the distribution of a Markov chain is certainly one of the main reasons why they are so widely used, and have so many applications in lots of different fields.

### 2.4 Examples of Markov Chains

There is a variety of models used in sports which leverage the power of Markov chains to enhance their analyses and predictions, hopefully some of which you will study in your projects, but it is still worth mentioning straightforward instances of those.

- Any random walk $(S_n)_{n\geq 0}$ is an instance of a Markov chain, and possibly its most simple one. A random walk can be obtained from a sequence of independent variables $\{X_n\}_{n\geq 1}$ by setting $S_0 = 0$ and $S_n = S_{n-1} + X_n$ for any $n \geq 1$. The goal here is to add at each step a new component to our current "portfolio", independent from the past, and you can imagine the variety of different environments to which they can be applied: portfolio variations in gambling, share prices in finance, movements of animals in biology, etc. In sports, this is a very simple but effective way of dealing with a sequence of games played throughout a season. Suppose that all the game outcomes are independent (which is obviously a strong assumption), and that we denote their results by 1 for a win and 0 for a loss. In this case, the total number of wins will be a random walk, and hence a Markov chain (with very simple transition matrix).
- Lots of articles in sports aiming to describe the dynamics of a game or its evolution across time use models based on Markov chains, especially in tennis (see for a non exhaustive list [2], [3], [4]).

More generally, as long as you assume that the event which affects the quantities of interest arrives independently from the past, you will have to deal with a form of Markov chain.

## 3 Decision Analysis

### 3.1 Bayesian Setting

Very often, our problem at hand can be reduced to a simple outcome prediction: we have some data $x$, and want to predict some result $y$, for instance the winner of a game.

Usually, our model is parameterized by some $\theta \in \Theta \subset \mathbb{R}^d$, where $d$ is some number (reflecting the number of degrees of freedom), and we have several examples or a sequence $\mathcal{D}_n = ((x_1, y_1), \dots, (x_n, y_n))$ of examples which should help us decide where $\theta$ should lie. Indeed, depending on our model and on the data $\mathcal{D}_n$ we get to observe, we can attribute for each $\theta$ a value $\mathcal{L}(\theta, \mathcal{D}_n)$ reflecting the likelihood of observing a specific $\theta$. For instance, suppose we want to predict the probability $\theta \in [0,1]$ that Djokovic beats Nadal during their match: given their previous matches, we can infer the likelihood of having the actual probability equal to some $\theta_0$ for each $\theta_0 \in [0,1]$. Typically, since we hold a reasonable amount of samples and the results are well balanced (28-25 Djokovic), we see that a value around 0.5 should be way more likely than 0. Note that in this specific case, the predictive variable $x$ could be the tennis court of the match (likely to influence the outcome), and the response $y$ will simply be the outcome of the game (who won?).

In the Bayesian setting, the parameter $\theta$ is not fixed but on the contrary also comes from a distribution $\theta \sim g(\theta)$, called a prior distribution. The latter reflects the knowledge that you have about your problem before looking at our data. For instance, if you do not know anything about tennis, it may be safer to assume at first that $\theta$ comes from a uniform distribution over $[0, 1]$, which in this case, is as uninformative as possible. On the other hand, if you are a long time fan, and thus very knowledgeable about your problem, you might want to reflect your belief in your prior.

Each time that you get to observe the outcome of a new game $((x_j, y_j)$, you can update your information about $\theta$ and make it reflect what you witnessed. Obviously, if you were to observe that Djokovic had beaten Nadal 50 times in a row, your distribution of $\theta$ should derive towards 1 and get more and more peaked, as you grow more and more confident about the actual location of his actual probability of winning. Given $\mathcal{D}_n$, you will therefore compute your final posterior distribution $g(\theta \mid \mathcal{D}_n)$ (the one "up-to-date").

With this posterior distribution at hand, we can then move to the prediction part. Classically, our model is such that given some predictor $x$ and parameter $\theta$, our outcome $y$ has a certain distribution $p(y \mid x, \theta)$ (such as in linear regression, where $y = x^T\theta + \epsilon$). Since we do not have the real $\theta$, we are going to replace it by its posterior distribution, so as to get our posterior predictive distribution $\hat{p}(y \mid x, \mathcal{D}_n)$:

$$\hat{p}(y \mid x, \mathcal{D}_n) = \int_{\theta \in \Theta} p(y \mid x, \theta) g(\theta \mid \mathcal{D}_n) d\theta \qquad (2)$$

The posterior likelihood of observing $y$ given $x$ and the data $\mathcal{D}_n$ we observe is just the average on all $\theta$ of the likelihood of observing $y$ given $x$ and a specific $\theta$ weighted by our posterior distribution of $\theta$, accordingly updated to reflect our new data. In other words, you can think of $g(\theta \mid \mathcal{D}_n)$ as the probability that our parameter of interest is effectively equal to such $\theta$, and of $p(y \mid x, \theta)$ as the probability of actually observing $y$ given the $x$ you have at hand, and some $\theta$. Because you have some beliefs on $\theta$, you reflect this by taking the average over $\theta$ of all those probabilities, which leads to the formula (2).

We see that this procedure also creates the possibility to give more influence to more recent information, since it just suffices to weigh more recent information when computing the updated distribution.

### 3.2 A simple tennis example

Suppose that we want to make real-time tennis prediction for a betting website, and we want to predict the winner of the next game. Our model is very basic: each point is played independently of each other, with the same probability $\theta \in [0, 1]$ for the server of winning the point. We want to answer the following question: given the current score, what is the probability that the server wins the game? For the sake of argument, we will refer to the server as She and the returner as He.

We are precisely in a situation which can be modeled as a Markov chain, with 17 different states representing the different possible scores (assuming that 40-40 is the same as 30-30 and so on). Now we can compute the transition matrix very easily, and it will be quite sparse, since from one given, we can only move to two different other states. For instance, if $\theta = 0.51$, our matrix will look something like Figure 1, where GS stands for "Game Server", and GR for "Game Returner".

| states | 0-0 | 0-15 | 0-30 | 0-40 | 15-0 | 15-15 | 15-30 | 15-40 | 30-0 | 30-15 | 30-30 | 30-40 | 40-0 | 40-15 | 40-30 | GS | GR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| String | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 |
| 0-0 | 0.0 | 0.49 | 0.0 | 0.0 | 0.51 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 0-15 | 0.0 | 0.0 | 0.49 | 0.0 | 0.0 | 0.51 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 0-30 | 0.0 | 0.0 | 0.0 | 0.49 | 0.0 | 0.0 | 0.51 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 0-40 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.51 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.49 |
| 15-0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.49 | 0.0 | 0.0 | 0.51 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 15-15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.49 | 0.0 | 0.0 | 0.51 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 15-30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.49 | 0.0 | 0.0 | 0.51 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 15-40 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.51 | 0.0 | 0.0 | 0.0 | 0.0 | 0.49 |
| 30-0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.49 | 0.0 | 0.0 | 0.51 | 0.0 | 0.0 | 0.0 | 0.0 |
| 30-15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.49 | 0.0 | 0.0 | 0.51 | 0.0 | 0.0 | 0.0 |
| 30-30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.49 | 0.0 | 0.0 | 0.51 | 0.0 | 0.0 |
| 30-40 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.51 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.49 |
| 40-0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.49 | 0.0 | 0.51 | 0.0 |
| 40-15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.49 | 0.51 | 0.0 |
| 40-30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.49 | 0.0 | 0.0 | 0.0 | 0.0 | 0.51 | 0.0 |
| GS | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 |
| GR | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |

**Figure 1:** Transition matrix for a tennis game

Now, we saw in the section on Markov chains that we only had to compute the iterates of the transition matrices of know what was our distribution of states after a few points. We can for instance compute $P^{20}$ to get the probability of winning the game after 20 points, under any initial score, which is summed up in Figure 2.

We learn that if the game starts, then after 20 points there is a probability of 52.4% that She has won the game, 47.38% that He has won the game, and some 0.2% chance that the game is still at play. If She is leaded by 0-40, on the contrary, then She has a probability 93.1% of losing the game within the next 20 points, and only 6.9% chance of getting a come-back and win it.

| | states | 0-0 | 0-15 | 0-30 | 0-40 | 15-0 | 15-15 | 15-30 | 15-40 | 30-0 | 30-15 | 30-30 | 30-40 | 40-0 | 40-15 | 40-30 | GS | GR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | String | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 | Float64 |
| 1 | 0-0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0024 | 0.0 | 0.0 | 0.0 | 0.0 | 0.5237 | 0.4738 |
| 2 | 0-15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0012 | 0.0 | 0.0 | 0.0013 | 0.3646 | 0.6329 |
| 3 | 0-30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.001 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2023 | 0.7967 |
| 4 | 0-40 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0003 | 0.0 | 0.0 | 0.0003 | 0.0687 | 0.9308 |
| 5 | 15-0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0012 | 0.0 | 0.0 | 0.0012 | 0.6766 | 0.321 |
| 6 | 15-15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0015 | 0.0 | 0.0 | 0.0 | 0.0 | 0.5217 | 0.4768 |
| 7 | 15-30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0007 | 0.0 | 0.0 | 0.0008 | 0.3307 | 0.6678 |
| 8 | 15-40 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0005 | 0.0 | 0.0 | 0.0 | 0.0 | 0.135 | 0.8645 |
| 9 | 30-0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0009 | 0.0 | 0.0 | 0.0 | 0.0 | 0.8267 | 0.1724 |
| 10 | 30-15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0007 | 0.0 | 0.0 | 0.0007 | 0.7053 | 0.2933 |
| 11 | 30-30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.001 | 0.0 | 0.0 | 0.0 | 0.0 | 0.5195 | 0.4795 |
| 12 | 30-40 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0005 | 0.0 | 0.0 | 0.0005 | 0.2647 | 0.7343 |
| 13 | 40-0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0002 | 0.0 | 0.0 | 0.0002 | 0.9433 | 0.0563 |
| 14 | 40-15 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0005 | 0.0 | 0.0 | 0.0 | 0.0 | 0.8845 | 0.115 |
| 15 | 40-30 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0005 | 0.0 | 0.0 | 0.0005 | 0.7643 | 0.2347 |
| 16 | GS | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 |
| 17 | GR | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |

**Figure 2:** Probabilities after 20 points played

The last thing we can compute is the probability that She wins the game as a function of $\theta$.

Indeed, we see that, even though She wins only 51% of the points, She still has a greater probability of eventually winning the game.

Given some $\theta \in [0,1]$, the probability $p_G(\theta)$ that She wins the game is simply $\lim_{t\to\infty} P^t(0-0, GS)$, i.e the limit when the number of points becomes very large that She has won the game starting from 0-0. The real function $p_G(\theta)$ is shown in figure 3.

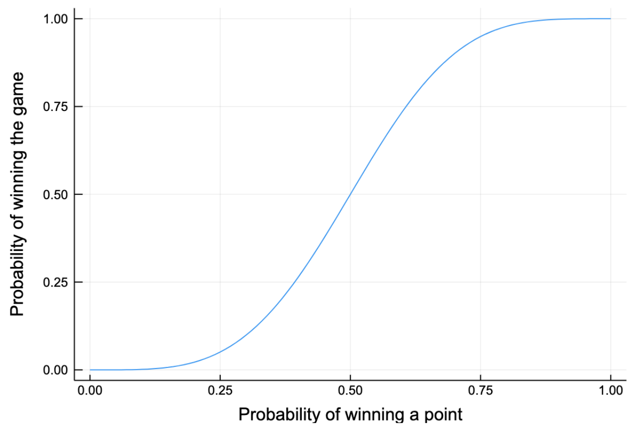


**Figure 3:** Probability for the server to win the game

Finally, this tennis example can be easily related to our Bayesian setting discussion. Indeed, suppose that the actual probability $\theta \in [0,1]$ of winning a point is unknown, and is thought to be originally distributed from a uniform distribution. During the match, we can record every point player on her serve and update our belief about $\theta$. Indeed if given that She played $n$ points on her serve, and won $i$ of them:

$$\theta \sim Beta(1+i, 1+n-i)$$

For instance if She played 100 points on her serve, and won 52 of them, figure 4 gives you the posterior distribution of $\theta$. Note that the more points have been played, the more peaked will the distribution be.

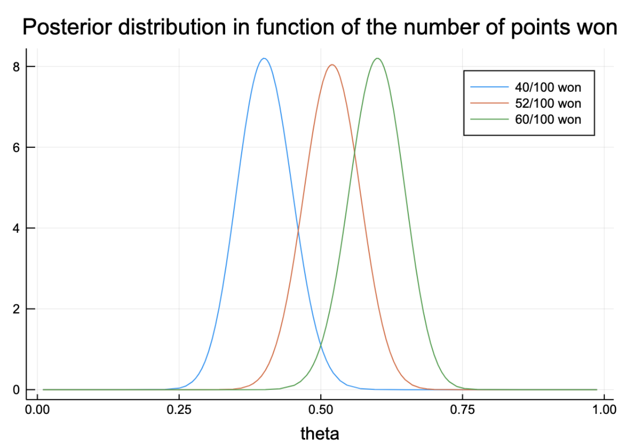


**Figure 4:** Posterior distributions for several number of points won

# 4 Predicting the Big Game Outcome

Our goal in this section is to give a prediction of the winner of the 2019 Big Game, using only its list of former winners (see Figure ). Looking at the Wikipedia page of the event, we learn that Stanford leads by 64 wins against 46 for Berkeley and 11 draws. Because draws are not allowed anymore we will remove them of our analysis, hoping that it will not undermine our prediction too much (which is a reasonable assumption, given that a tie was quite unlikely even then, and the last tie occurred in 1988).

California victories | Stanford victories | Ties

| No. | Date | Location | Winner | Score |
|---|---|---|---|---|
| 1 | 1892 | San Francisco | Stanford | 14–10 |
| 2 | 1892 | San Francisco | Tie | 10–10 |
| 3 | 1893 | San Francisco | Tie | 6–6 |
| 4 | 1894 | San Francisco | Stanford | 6–0 |
| 5 | 1895 | San Francisco | Tie | 6–6 |
| 6 | 1896 | San Francisco | Stanford | 20–0 |
| 7 | 1897 | San Francisco | Stanford | 28–0 |
| 8 | 1898 | San Francisco | California | 22–0 |
| 9 | 1899 | San Francisco | California | 30–0 |
| 10 | 1900 | San Francisco | Stanford | 5–0 |
| 11 | 1901 | San Francisco | California | 2–0 |
| 12 | 1902 | San Francisco | California | 16–0 |
| 13 | 1903 | San Francisco | Tie | 6–6 |
| 14 | 1904 | Berkeley | Stanford | 18–0 |
| 15 | 1905 | Stanford | Stanford | 12–5 |
| 16 | 1906 | Berkeley | Stanford | 6–3 |
| 17 | 1907 | Stanford | Stanford | 21–11 |
| 18 | 1908 | Berkeley | Stanford | 12–3 |
| 19 | 1909 | Stanford | California | 19–13 |
| 20 | 1910 | Berkeley | California | 25–6 |
| 21 | 1911 | Stanford | California | 21–3 |
| 22 | 1912 | Berkeley | Tie | 3–3 |
| 23 | 1913 | Stanford | Stanford | 13–8 |
| 24 | 1914 | Berkeley | Stanford | 26–8 |
| 25 | 1919[a][b] | Stanford | California | 14–10 |
| 26 | 1920 | Berkeley | California | 38–0 |
| 27 | 1921 | Stanford | California | 42–7 |
| 28 | 1922 | Stanford | California | 28–0 |
| 29 | 1923 | Berkeley | California | 9–0 |
| 30 | 1924 | Berkeley | Tie | 20–20 |
| 31 | 1925 | Stanford | Stanford | 27–14 |
| 32 | 1926 | Berkeley | Stanford | 41–6 |
| 33 | 1927 | Stanford | Stanford | 13–6 |
| 34 | 1928 | Berkeley | Tie | 13–13 |
| 35 | 1929 | Stanford | Stanford | 21–6 |
| 36 | 1930 | Berkeley | Stanford | 41–0 |
| 37 | 1931 | Stanford | California | 6–0 |
| 38 | 1932 | Berkeley | Tie | 0–0 |
| 39 | 1933 | Stanford | Stanford | 7–3 |
| 40 | 1934 | Berkeley | Stanford | 9–7 |
| 41 | 1935 | Stanford | Stanford | 13–0 |
| 42 | 1936 | Berkeley | California | 20–0 |
| 43 | 1937 | Stanford | #2 California | 13–0 |
| 44 | 1938 | Berkeley | #9 California | 6–0 |
| 45 | 1939 | Stanford | California | 32–14 |
| 46 | 1940 | Berkeley | #3 Stanford | 13–7 |
| 47 | 1941 | Stanford | California | 16–0 |
| 48 | 1942 | Berkeley | Stanford | 26–7 |
| 49 | 1946[c] | Berkeley | Stanford | 25–6 |
| 50 | 1947 | Stanford | #9 California | 21–18 |
| 51 | 1948 | Berkeley | #4 California | 7–6 |
| 52 | 1949 | Stanford | #3 California | 33–14 |
| 53 | 1950 | Berkeley | Tie | 7–7 |
| 54 | 1951 | Stanford | #19 California | 20–7 |
| 55 | 1952 | Berkeley | California | 26–0 |
| 56 | 1953 | Stanford | Tie | 21–21 |
| 57 | 1954 | Berkeley | California | 28–20 |
| 58 | 1955 | Stanford | #18 Stanford | 19–0 |
| 59 | 1956 | Berkeley | California | 20–18 |
| 60 | 1957 | Stanford | Stanford | 14–12 |
| 61 | 1958 | Berkeley | California | 16–15 |
| 62 | 1959 | Stanford | #19 California | 20–17 |
| 63 | 1960 | Berkeley | California | 21–10 |
| 64 | 1961 | Stanford | Stanford | 20–7 |
| 65 | 1962 | Berkeley | Stanford | 30–13 |
| 66 | 1963 | Stanford | Stanford | 28–17 |
| 67 | 1964 | Berkeley | Stanford | 21–3 |
| 68 | 1965 | Stanford | Stanford | 9–7 |
| 69 | 1966 | Berkeley | Stanford | 13–7 |
| 70 | 1967 | Stanford | California | 26–3 |
| 71 | 1968 | Berkeley | Stanford | 20–0 |
| 72 | 1969 | Stanford | #14 Stanford | 29–28 |
| 73 | 1970 | Berkeley | California | 22–14 |
| 74 | 1971 | Stanford | #18 Stanford | 14–0 |
| 75 | 1972 | Berkeley | California | 24–21 |
| 76 | 1973 | Stanford | Stanford | 26–17 |
| 77 | 1974 | Berkeley | Stanford | 22–20 |
| 78 | 1975 | Stanford | #13 California | 48–15 |
| 79 | 1976 | Berkeley | Stanford | 27–24 |
| 80 | 1977 | Stanford | Stanford | 21–3 |
| 81 | 1978 | Berkeley | Stanford | 30–10 |
| 82 | 1979 | Stanford | California | 21–14 |
| 83 | 1980 | Berkeley | California | 28–23 |
| 84 | 1981 | Stanford | Stanford | 42–21 |
| 85 | 1982 | Berkeley | California | 25–20 |
| 86 | 1983 | Stanford | California | 27–18 |
| 87 | 1984 | Berkeley | Stanford | 27–10 |
| 88 | 1985 | Stanford | Stanford | 24–22 |
| 89 | 1986 | Berkeley | California | 17–11 |
| 90 | 1987 | Stanford | Stanford | 31–7 |
| 91 | 1988 | Berkeley | Tie | 19–19 |
| 92 | 1989 | Stanford | Stanford | 24–17 |
| 93 | 1990 | Berkeley | Stanford | 27–25 |
| 94 | 1991 | Stanford | #21 Stanford | 38–21 |
| 95 | 1992 | Berkeley | #14 Stanford | 41–21 |
| 96 | 1993 | Stanford | California | 46–17 |
| 97 | 1994 | Berkeley | California | 24–23 |
| 98 | 1995 | Stanford | Stanford | 29–24 |
| 99 | 1996 | Berkeley | Stanford | 42–21 |
| 100 | 1997 | Stanford | Stanford | 21–20 |
| 101 | 1998 | Berkeley | Stanford | 10–3 |
| 102 | 1999 | Stanford | Stanford | 31–13 |
| 103 | 2000 | Berkeley | Stanford | 36–30 OT |
| 104 | 2001 | Stanford | #13 Stanford | 35–28 |
| 105 | 2002 | Berkeley | California | 30–7 |
| 106 | 2003 | Stanford | California | 28–16 |
| 107 | 2004 | Berkeley | #4 California | 41–6 |
| 108 | 2005 | Stanford | California | 27–3 |
| 109 | 2006 | Berkeley | #21 California | 26–17 |
| 110 | 2007 | Stanford | Stanford | 20–13 |
| 111 | 2008 | Berkeley | California | 37–16 |
| 112 | 2009 | Stanford | California | 34–28 |
| 113 | 2010 | Berkeley | #7 Stanford | 48–14 |
| 114 | 2011 | Stanford | #8 Stanford | 31–28 |
| 115 | 2012 | Berkeley | #22 Stanford | 21–3 |
| 116 | 2013 | Stanford | #10 Stanford | 63–13 |
| 117 | 2014 | Berkeley | Stanford | 38–17 |
| 118 | 2015 | Stanford | #15 Stanford | 35–22 |
| 119 | 2016 | Berkeley | Stanford | 45–31 |
| 120 | 2017 | Stanford | #22 Stanford | 17–14 |
| 121 | 2018 | Berkeley | Stanford | 23–13 |

**Series:** Stanford leads 64–46–11

**Figure 5:** Big Game History

Suppose that we want to fit the following model onto our data: each year, the chances of win of either team only depend on the past game played between those teams. This seems to be a reasonable assumption, as teams don't entirely change within a year, so it might be reasonable to assume that the team which won the previous game is more likely to win the next one than the past year's loser.

In other terms, we can model our problem with the following matrix $P$:

$$P = \begin{bmatrix} p_w & 1 - p_w \\ p_l & 1 - p_l \end{bmatrix} \tag{3}$$

where $p_w \in [0, 1]$ represents the probability for Stanford to win this year given that they won last year, and $p_l \in [0, 1]$ is the probability for Stanford to win if they happened to lose the last Big Game.

You can also see this model as a Markov chain $X_n$ where $X_n \in \{0,1\}$ encodes whether or not Stanford wins the Big Game at year $n$, and in that case, the transition matrix of our Markov chain is precisely $P$! Indeed, if $X_{n-1}=1$, that is if Stanford has won the year before, then with probability $p_w$, you will have $X_n = 1$, and with probability $1-p_w$, you will have $X_n = 0$. The same occurs if $X_{n-1}=0$, with $p_l$ in place of $p_w$.

Now, if we look at our data, and we count all the actual transitions that occurred, we get the following matrix of counts:

$$\hat{P} = \begin{bmatrix} 43 & 20 \\ 20 & 26 \end{bmatrix} \tag{4}$$

we see that the likelihood of our parameters $p_w$ and $p_l$ is as follows:

$$\mathcal{L}(p_w, p_l \mid \mathcal{D}) \propto p_w^{43}(1-p_w)^{20} p_l^{20}(1-p_l)^{26}$$

which is maximal for $\hat{p}_w = \frac{43}{63} \simeq 68\%$ and $\hat{p}_l = \frac{20}{46} \simeq 43\%$.

In conclusion, we estimate that, after a win, Stanford has an estimated probability of taking the next game equal to 68%, but after a loss, this probability falls to 43%. Luckily, it seems that we are on a winning streak!

**Remark** Several last things to get out of our analysis:

- Here, we do not use the score of each game for each prediction, which is certainly a weakness of our model: a large score might indicate even greater chances for the next team to win the next Big Game. We will see other models that do take into account the gap between the two teams.

- On the long run, what proportion of wins can we expect for Stanford in the model (3)? Actually, if we consider that our real transition matrix is $\hat{P}$, we can compute the limit probability to be in the state 1 when the number of games grows very large, i.e. the unconditional probability that Stanford wins a game (no matter what the previous result). It should be actually close to the historical proportion of Stanford wins at the Big Game.

## References


[1] David Frank Percy (2015) Strategy selection and outcome prediction in sport using dynamic learning for stochastic processes, Journal of the Operational Research Society, 66:11, 1840-1849, DOI: 10.1057/jors.2014.137

[2] H. Riddle, Lawrence. (1988). Probability Models for Tennis Scoring Systems. Applied Statistics. 37. 490. 10.2307/2347332.

[3] Sadovski LE and Sadovski AL (1993). Mathematics and Sports. Translated by Makar-Limanov S and edited by Ivanov SAmerican Mathematical Society: Providence.

[4] Spanias D and Knottenbelt WJ (2013). Predicting the outcomes of tennis matches using a low-level point model. IMA Journal of Management Mathematics 24(3): 311320.

[5] Wikipedia Big Game Page https://en.wikipedia.org/wiki/Big_Game_(American_football)