Markov Chains Mixing Times

Ori Gurel-Gurevich Ariel Yadin

December 16, 2007

1 Introduction

Deck of cards. How many shuffles does it take to mix? How do we know it mixes? What does it mean that the deck is mixed?

Examples using Deck: cutting into equal halves. to quarters. "perfect" riffle shuffle: cutting to exact halves and exactly interweaving them.

Exercise. *what happens in a perfect riffle shuffle?*

Conclusion: failure of deterministic procedures. There is no randomness, without randomness (and entropy is it's apostle).

Another example: cutting randomly into two parts. Not enough, in what sense? in the sense that if you knew something about the original arrangement (supposed it was completely ordered), you still know quite a bit about the resulting arrangement.

Usual solution: iterating.

Question: What happens when iterating random cutting?

Answer: It's still just random cutting.

So just iterating some random action is not enough. one need to prove that iteration leads to the desired distribution. Luckily, this is usually quite easy.

Example: top-to-random, does it shuffle?

Exercise. *prove that it is possible to get to any permutation that way.*

Example: riffle/thorp shuffle. Here it is not precisely clear what is a correct mathematical formulation.

Other examples: random to random, random transposition, random adjacent transposition.

Emphasize that the deck is in some specific state, but we are interested mainly in the *distribution* of this state. The state of the deck does not tend to anything, it is the distribution tending to some limit.

Even after proving the limit distribution is as desired, We still need to ask how many iterations does it take to reach it. Certainly, riffle shuffle seems more efficient then top-to-random, but we need to prove it. Usually, the limit is never reached precisely. Instead we define a notion of distance, and ask how many iteration are needed to be ε close to the limit. This is called the mixing time. This is what this course is about (mostly).

Diversion: not enough probability in undergraduate courses, so we might take some (long) detours not strictly related to mixing times.

If we denote the distance you get at time t by $d(t)$ and the time it takes to reach distance ε by $\tau(\varepsilon)$, then how does this function behaves when $\varepsilon \to 0$? This is interesting, but it is *not* the focus of this course. The reason is this: it will be clear from the metric definition that for any given chain $d(t) \to 0$ exponentially fast, because if $d(t) < \varepsilon$ then $d(t + \tau(1/2)) < \varepsilon/2$. Therefore, (a) the asymptotic as $\varepsilon \to 0$ are not very interesting and (b) there's not much difference between $\tau(\varepsilon_1)$ and $\tau(\varepsilon_2)$.

What will we be interested in? we will want to know the asymptotic behaviour for a class of processes, as a function of some parameter.

Example: Shuffling a deck of n cards. How does $\tau(n, \varepsilon)$ behaves as a function of n (for a fixed ε) ?

Besides card shuffling, which is of outmost importance, what other uses is all this? Sampling from large, implicitly defined distributions.

Example: we want to get a deck such that no two adjacent cards has the same suite (?), how to choose uniformly from all such decks? It is quite easy to define a "shuffling" that has this as a limit distribution. Then all you have to do is calculate the mixing time.

This also enables one to make approximation for the counting problem: how many such decks are there? a notable real example is that of approximating the permanent (same as determinant without signs). Hopefully, people from other fields would provide more examples.

2 formalities

Course time and place.

Requirements: linear algebra, probability, a bit of graph theory. Don't hesitate to ask about elementary things.

Grades: Course status not clear. Hopefully, will be official. likely a home-exam, but don't count on it. exercises, possibly obligatory.

resources: Yuval's book. give link. Don't be intimidated by some of the material there (esp. background). http://www.stat.berkeley.edu/ peres/sim.pdf

Course mailing list.

3 Stochastic processes

Stochastic=random.

What is a stochastic process? A sequence of Random Variables / A random sequence of variables.

Examples: independent RVs, constant RV's, identical RVs, cumulative sums of independent RVs.

not necessarily a sequence of numbers.

Examples: Graphs, $G(n, m)$ as a process in m, adding and deleting edges. WWW/small worlds models. Ideas about mixing related to Google's search engine.

State space Ω. We will deal with finite space (in particular, discrete) and finite time - dismissal of measurability issues.

What is a Markov Chain?

Chain=process.

Not-quite-correct definition: x_{t+1} depends only on x_t. Why incorrect/inaccurate? because it is *not* true that in a Markov chain x_7 is independent of x_3.

Correct definition:

$$P\{X_{t+1} = y | X0 = x0, X1 = x1, ..., Xt - 1 = xt - 1, Xt = x\} = P(x, y)$$

This is sometimes called the Markov property.

Of course, P then is a stochastic matrix:

$$\sum_{y \in \Omega} P(x, y) = 1$$

For all $x \in \Omega$.

Frog example. which of the processes is and which isn't a Markov chain?

$$\mu_t = \mu_{t-1} P = \mu_0 P^t$$

Exercise. *If P and Q are stochastic then so is PQ.*

Remark: time-homogenous.