Regenerative Sampling

Tutorial - Part 1

Mayank Kakodkar

September 10, 2020

Learning Objectives

For Markov chains

- What is a **steady state simulation** and when is it well posed?
- For any well posed simulation problem:
  - a steady state distribution exists for the chain
  - the chain is **Harris recurrent**
    - (positive recurrent but for measurable spaces)

Harris recurrent chains

- *Aperiodic* Harris recurrent chains are ergodic
- They exhibit regenerative structure under certain conditions
  - Regenerative structure can be exploited to split a long chain into i.i.d *tours* that yield
    - parallelizable estimators
    - confidence intervals
    - bias reduction
- Convergence to steady state is a direct consequence of regenerative structure

1. Bruno and I recently were discussing how aperiodicity conditions can be relaxed
2. We will see one case today (I am working on understanding other cases too)

Extensions

- Most discrete event simulations can be transformed to SMPs (Glynn 2006)
  - These can be converted to Markov processes
- $\therefore$ regenerative sampling has broad applicability

Why is Regenerative structure useful?

1. Lets do the punch line first and start by defining the task

The Steady State simulation problem (SSSP)

- Given
  - a Markov chain $\mathbf{\Phi} = \{X_n : n \geq 1\}$ over state space $\mathcal{S}$ with kernel P
  - a **bounded** real valued function $f : \mathcal{S} \to \mathbb{R}$
- Compute

$$r(f) = \lim_{n\to\infty} \underbrace{\frac{1}{n}\sum_{i=1}^{n} f(X_i)}_{=r_n(f)}$$

if such a limit exists.

1. We know from our own experience/background that this limmit only exists if the markov chain is ergodic
2. For RW on graphs that means a connected non bipartite graph

The natural estimator for SSSP

- For some finite t, and for an arbitrary $X_1 = x$, sample a chain $X_1, \dots, X_t$ and compute

$$r_t(f, x) = \frac{1}{t} \sum_{i=1}^{t} f(X_i)$$

- Clearly $\mathbb{E}[r_t(f, X_1)] = r(f)$ when $t \to \infty$
  - When $r(f)$ is invariant to the starting state

1. The asymptotic nature is due to initialization bias
2. We will cover this invariance later

Disadvantages

- Since initialization bias in $r_t(f, x)$ goes to 0 asymptotically
  - It makes sense to run one long chain rather than multi thread
    - For example mixing times for simple graph RWs are $\geq \frac{\text{dia}(G)^2}{20\log(|V|)}$ (Levin and Peres 2017)
- Confidence intervals and higher moments:
  - Can be upper bounded using eigen values of P e.g. Ribeiro and Towsley (2012)
  - May be estimated by exploiting some structure e.g. Glynn (2006)

1. Eigen value bounds are usually loose and eigen values are not readily available
2. Structure may be hard to find in general chains

Regenerations

- On the other hand **if** chains can be split
  - $Y_j = (X_{T_{j-1}}, X_{T_{j-1}+1}, \ldots, X_{T_j})$
  - Y_j and $\tau_j = T_j - T_{j-1}$ are i.i.d sequences
- We can estimate $r(f)$ as

$$\frac{\frac{1}{J}\sum_{j=1}^{J}\sum_{X \in Y_j} f(X)}{\frac{1}{J}\sum_{j=1}^{J}\tau_j}$$

- Since there is no mixing time/ burn-in multiple copies of the above estimators are all i.i.d estimators for $r(f)$
  - estimators for higher moments can be easily constructed
    - examples in Glynn (2006) and Avrachenkov, Ribeiro, and Sreedharan (2016)

1. Jacknife estimates for bias removal are possible

However splitting is not so easy

How do we get there?

- Well posed steady state simulation
- $\rightarrow$ Harris Recurrence
- $\rightarrow$ Regenerations where Doeblin's condition holds

Well-Posedness of an SSSP

- Recall the goal

$$r(f) = \lim_{n\to\infty} \underbrace{\frac{1}{n}\sum_{i=1}^{n} f(X_i)}_{=r_n(f)}$$

- The problem of computing $r(f)$ given $\mathbf{\Phi}$ is well posed iff

$$\lim_{n\to\infty} \mathbb{E}_x r_n(f) = r(f)$$

  - for any bounded real function f on $\mathcal{S}$ regardless of $x \in \mathcal{S}$
- I.e. the limiting behavior of $\mathbf{\Phi}$ is invariant to how the simulation is started (Glynn 1982)

1. Informally limit needs to exist
2. Formally we say that the simulation problem is wellposed iff the following holds - by convention $\mathbb{E}_x[\cdot] = \mathbb{E}[\cdot|X_1 = x]$

Consequences of Well-Posedness

- If the steady state simulation problem is well posed we can show that
  - there exists π such that $\sum_x \pi(x)P(x,y) = \pi(y)$
  - and $r(f) = \sum_x \pi(x)f(x) = \mathbb{E}_\pi[f(X)]$
  - $\mathbf{\Phi}$ is Harris recurrent (Glynn 1982)

Proof Sketch

- For any $A \subseteq \mathcal{S}$ let $\pi(A) = r(I_A)$
- Due to well posedness

$$r(I_A) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_{X_i \in A}$$

- Sum $r(I_A)$ over all A to see that it is a probability measure over $\mathcal{S}$
- Decomposing f as $f(X) = \sum_{x \in \mathcal{S}} I_{X=x} f(x)$ we see that $r(f) = \sum_x \pi(x) f(x)$

1. Since well posedness applies to all f
2. There also is a measure theoretic proof for how π is the steady state distribution for $\mathbf{\Phi}$ and this convergence implies ergodicity and therefore Harris Recurrence

π-Harris Recurrence

- Given a measure π, $\mathbf{\Phi}$ is π-Harris (Positive) Recurrent if
  - For each $A \subset \mathcal{S}$ if $\pi(A) > 0$ for all $x \in \mathcal{S}$

$$P_x\left(\sum_{i=1}^{\infty} I_{X_i \in A} = \infty\right) = 1$$

- Recall that we had seen that

$$\pi(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_{X_i \in A}$$

1. Seemed to me that HR is a measure theoretic extension of the definition of positive recurrence.
2. Need to read more
3. if the chain doesn't visit A infinitely often the measure would be 0

Pause

- We now know that well posed simulation problems on Markov chains involve Harris recurrent Markov chains
- Why is this important?

Harris Recurrence and Uniform Ergodicity

- If $\mathbf{\Phi}$ is Harris recurrent and aperiodic (Meyn and Tweedie 1993) (Theorem 13.3.3)

$$\| \int \lambda(dx) P^n(x, \cdot) - \pi \| \to 0 \,, \quad n \to \infty \,,$$

- If $\mathcal{S}$ is finite we get uniform ergodicity where for some $\rho < 1$

$$\| P^n(x, \cdot) - \pi \| \le \rho^n \,,$$

A more interesting case (Meyn and Tweedie 1993) (Theorem 16.2.4)

- If, for some m, $\forall x \in \mathcal{S}$ and $A \in \mathcal{B}(\mathcal{S})$, $\mathbf{\Phi}$ satisfies

$$P^m(x, A) \geq (1 - \rho)\nu(A)$$

  where ν is a probability measure and $0 < \rho < 1$ we have

$$\|P^n(x, \cdot) - \pi\| \leq \rho^{n/m}\,,$$

  and vice versa.
- This is called Doeblin's condition.

Splitting Φ into tours

Assume $m = 1$ and rewrite $P(x, A)$ as

$$P(x, A) = (1 - \rho)\nu(A) + \rho Q(x, A)$$

so

$$Q(x, A) = \frac{(P(x, A) - (1 - \rho)\nu(A))}{\rho}$$

- At every step toss a biased coin with $\mathbb{H}$ probability ρ.
- If you sample $\mathbb{T}$ sample the next state from ν else Q
- Note how the chain regenerates at each $\mathbb{T}$
- Mykland, Tierney, and Yu (1995) is a direct application of this result

Proof sketch for convergence

- Start two coupled chains, $\mathbf{\Phi}$ and $\mathbf{\Phi}'$
  - $\mathbf{\Phi}'$ is started in steady state, i.e. is always dist w.p. π
- Both chains use the same random noise generator to sample transitions
  - Once the chains collide,
    - they'll stay together and both will be dist w.p. π
- From the coupling inequality we have

$$\|P^n(x,\cdot) - \pi\| \leq P(\Phi_n \neq \Phi'_n)\,.$$

  - Sampling n heads in a row implies $\Phi_n \neq \Phi'_n$

$$\therefore \|P^n(x,\cdot) - \pi\| \leq P(\text{flipping } n\ \mathbb{H} \text{ in a row}) = \rho^n$$

  which completes the proof.

1. Transitions are as sampled above

Conclusion

- We saw how well posed simulation problems have Harris recurrence
- Doeblin's condition guarantees unifrom ergodicity and enables splitting and thus regenerative sampling

Future Lectures

- Coupling based methods for regeneration in general MCs
- Atom based ergodicity
- Regenerative estimates for finite state MCs

References I

Avrachenkov, Konstantin, Bruno Ribeiro, and Jithin K. Sreedharan. 2016. "Inference in OSNs via Lightweight Partial Crawls." *Sigmetrics* 44 (1): 165–77. https://doi.org/10.1145/2896377.2901477.

Glynn, Peter W. 2006. "Chapter 16 Simulation Algorithms for Regenerative Processes." In *Handbooks in Operations Research and Management Science*, 13:477–500. Elsevier. https://doi.org/10.1016/S0927-0507(06)13016-9.

Glynn, Peter W. 1982. "Regenerative Aspects of the Steady-State Simulation Problem for Markov Chains." STANFORD UNIV CA DEPT OF OPERATIONS RESEARCH.

Levin, David A, and Yuval Peres. 2017. *Markov Chains and Mixing Times*. Vol. 107. American Mathematical Soc.

References II

Meyn, S., and R. Tweedie. 1993. *Markov Chains and Stochastic Stability*. Cambridge University Press. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.34.3433.

Mykland, Per, Luke Tierney, and Bin Yu. 1995. "Regeneration in Markov Chain Samplers." *Journal of the American Statistical Association* 90 (429): 233–41.

Ribeiro, Bruno, and Don Towsley. 2012. "On the Estimation Accuracy of Degree Distributions from Graph Sampling." In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 5240–7. https://doi.org/10.1109/CDC.2012.6425857.