Mean Field Theory in Networks

Vishrant Tripathi

November 30, 2018

Introduction

- **Mean Field Theory** - introduced in the statistical physics community to study phase transitions (Curie [1895] & Weiss [1907])
- Used to study the behavior of **large and complex stochastic models** by considering **simpler deterministic models**
- Mean field approximations to study the performance of networks have become common over the past two decades
- Other applications include epidemic models and game theory

Introduction

- Models to be studied usually consist of a **large number of small interacting components** (Ex. - Ising Model)
- Effect of everyone else on any given individual is approximated by a single averaged effect or a "**mean field**"
- Reduces a many-body problem to a one-body problem

Outline

- We start with a simple epidemic model to describe the mean field approach
- We then look at a more concrete queuing application using the supermarket model
- For both of the above, our discussion closely follows Mitzenmacher's PhD thesis
- We also look at mean field techniques as applied to analyzing medium access protocols

A Simple Epidemic Model

- Population of N people, out of which X are susceptible to a disease and Y have the disease
- Rate of infection $\propto$ the amount of interaction between the susceptible and infected population
- Infected people recover and become immune i.i.d. according to an exponential random variable with rate μ

Can be modeled as a CTMC -

$$q_{(X,Y),(X-1,Y+1)} = \lambda X \frac{Y}{N}$$
$$q_{(X,Y),(X,Y-1)} = \mu Y$$

A Simple Epidemic Model

- Need to analyze time evolution and stationary distribution of this Markov Chain
- A simpler approach is to model the system using differential equations. Let $x = \frac{X}{N}$ and $y = \frac{Y}{N}$, and consider

$$\frac{dx}{dt} = -\lambda xy$$
$$\frac{dy}{dt} = \lambda xy - \mu y$$

A Simple Epidemic Model

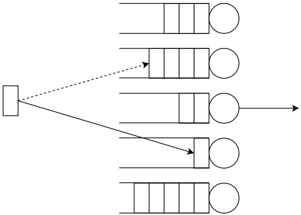
- Consider the expected behavior of the Markov chain over a small time Δt

$$\mathbb{E}[\Delta X] = -N\lambda \frac{X}{N}\frac{Y}{N}\Delta t$$
$$\mathbb{E}[\Delta Y] = (N\lambda \frac{X}{N}\frac{Y}{N} - N\mu\frac{Y}{N})\Delta t$$

- Observe that the differential equations "track" this behavior
- Deterministic path given by the differential equations describes the limiting behavior of the Markov chain as $N \to \infty$
- Informally, a law of large numbers for Markov processes takes effect, and the system must behave close to its expectation

The Supermarket Model

Back to queuing, the supermarket model is used to study distributed load balancing in networks



Poisson arrivals of rate $n\lambda$, where $\lambda < 1$, at a collection of n servers. Choose some constant number d i.i.d. uniform out of n, and join the shortest of these d queues. Service is FIFO, job sizes are exp(1) i.i.d.

The Supermarket Model

- We are interested in the expected time a job spends in the system - a natural measure of system performance
- When $d = 1$, this is identical to n $M/M/1$ queues and is easy to analyze
- However, for $d > 1$, the supermarket model proves difficult to analyze because of dependencies
- Knowing the length of one queue affects the distribution of the length of all the other queues

The Supermarket Model

Recipe to analyze the supermarket model (and applying mean field theory in general) -

- Define an idealized process, corresponding to a system of infinite size
- Analyze this process, which is cleaner and easier because its behavior is completely deterministic
- Relate the idealized system to the finite system, bounding the error between the two for large n

The Supermarket Model

We first define some useful quantities

- $n_i(t) \triangleq$ number of queues with i customers at time t
- $p_i(t) \triangleq \frac{n_i(t)}{n}$ (fraction of queues with i customers at time t)
- $s_i(t) \triangleq \sum_{k=i}^{\infty} p_i(t)$ (fraction of queues with at least i customers)

The key observation is that the system evolution can be described easily using the infinite dimensional vector $(s_0, s_1, s_2, \dots)$ as the state.

The Infinite System

- For any finite n, the system evolves as a Markov chain over the state space $(s_0, s_1, s_2, \ldots)$.
- We now introduce a deterministic infinite dimensional system whose time evolution is given by -

$$\frac{ds_i}{dt} = \lambda(s_{i-1}^d - s_i^d) - (s_i - s_{i+1}), \forall i \geq 1;$$
$$s_0 = 1.$$

- The claim is that this system of differential equations closely tracks the supermarket model for large n

The Infinite System

Informally, consider a small time interval Δt

- Number of arrivals in this interval $= \lambda n \, \Delta t$
- $\mathbb{P}$(arriving customer joins queue of size $i-1$) $= s_{i-1}^d - s_i^d$
- $\mathbb{P}$(customer leaves queue of size i) $= n_i \Delta t = n(s_i - s_{i+1})\Delta t$

Finally, we have

$$n\Delta s_i = \lambda n(s_{i-1}^d - s_i^d)\Delta t - n(s_i - s_{i+1})\Delta t$$

This is the same as the differential equations we set up earlier.

Fixed Point Analysis

- To understand the behavior of the infinite system we set up, we want to analyze its fixed point
- A fixed point $\vec{p}$ is a point such that if $\vec{s}(t') = \vec{p}$, then $\vec{s}(t') = \vec{p}$ for all $t' > t$
- To find such a fixed point we set $\frac{ds_i}{dt} = 0, \forall i$.
- Finally, if $d \geq 2$, we get $s_i = \lambda^{\frac{d^i - 1}{d-1}}, \forall i$
- In fact, it can be shown that this is the unique fixed point assuming average number of jobs per queue is finite

Convergence to Fixed Point

- Denote the fixed point found earlier by $(\pi_0, \pi_1, \pi_2, \ldots)$
- Consider the potential function $\phi(t) = \sum_{i=0}^{\infty} w_i |s_i(t) - \pi_i|$, where $w_i \geq 1$ are suitably chosen
- Mitzenmacher showed that if $\phi(0) < \infty$, then $\phi(t)$ decreases exponentially to zero, i.e. the system converges exponentially quickly to its fixed point (proof on blackboard)

The Power of 2 Choices

- If $d = 1$, the supermarket system behaves like n $M/M/1$ queues
- The fixed point for this system is given by $s_i = \lambda^i, \forall i$
- From our earlier analysis, if $d \geq 2$, the fixed point is given by $s_i = \lambda^{\frac{d^i - 1}{d - 1}}, \forall i$

Thus, if 2 or more choices are given to each job entering the system, the fraction of queues with length at least i decreases double exponentially with i as compared to just exponentially, when $d = 1$.

The Power of 2 Choices

- The expected time spent in the system by a packet entering at time t is given by $\sum_{i=0}^{\infty} s_i(t)^d$ (Proof on blackboard)
- Expected time spent by a packet in the infinite system is arbitrarily close to $T_d(\lambda) = \sum_{i=0}^{\infty} \pi_i^d$
- From standard queuing theory $T_1(\lambda) = \frac{1}{1-\lambda}$
- From fixed point analysis, $T_d(\lambda) = \sum_{i=1}^{\infty} \lambda^{\frac{d^i - d}{d-1}} \leq c_d \log(\frac{1}{1-\lambda})$ for some constant c_d dependent on d (Proof on blackboard)

Thus, giving 2 or more choices to each job leads to an **exponential reduction** in the time spent in the system, as compared to choosing a queue at random.

From Infinite to Finite

- The supermarket model is an example of a **density dependent family of jump Markov processes**
- Informally, a family of one parameter Markov processes, where the parameter signifies population size, area, volume etc.
- States normalized to densities and the transition rates depend only on these densities
- Infinite system is the limiting model as the parameter grows arbitrarily large

From Infinite to Finite:Kurtz's Theorem

- Kurtz's theorem provides a law of large numbers and Chernoff-like bounds for relating the infinite system to finite systems of such a family
- Starting from the same initial point over a very short time interval both systems behave similarly
- Assuming some form of **Lipschitzness** holds, the transition rates remain close for points that are close enough
- Repeating this argument inductively, we can bound how far the processes separate over any interval $[0, T]$

More Details

There are 2 main tools involved in proving Kurtz's theorem

- Lipschitzness
- Gronwall's inequality

Lemma

Gronwall's inequality - *Let $f(t)$ be a bounded function such that*

$$f(t) \leq \epsilon + \delta \int_0^t f(s)ds,$$

then we have

$$f(t) \leq \epsilon e^{\delta t}.$$

The approach is very similar to the way stochastic approximation and descent methods in optimization are viewed as a noisy discretization of corresponding ODEs

Finite System

Finally, for the finite supermarket model with $d \geq 2$, Mitzenmacher proves the following result -

Theorem

For any fixed T, the expected time spent in an initially empty supermarket system by a job over the interval [0,T] is given by

$$\sum_{i=1}^{\infty} \lambda^{\frac{d^i-1}{d-1}} + o(1),$$

where the o(1) term goes to zero as $n \to \infty$

Extensions

The mean field approach has 3 strengths - simplicity, generality, and accuracy. This makes it easy to apply to more complicated load balancing systems that have

- customer types,
- bounded buffers, or
- general service time distributions

Supermarket Model: Extensions

Consider **two types of customers**

- One type gets to choose only 1 queue; each customer is of this type w.p. 1-p
- The more privileged customer gets to choose 2 queues; each customer is of this type w.p. p

We modify the differential equations as follows -

$$\frac{ds_i}{dt} = \lambda p(s_{i-1}^2 - s_i^2) + \lambda(1-p)(s_{i-1} - s_i) - (s_i - s_{i+1})$$

Supermarket Model: Extensions

Consider **bounded buffers** of size b at each server. If an arriving job sees all d of its sampled queues to be full, it exits the system.

- The system can be represented by the state $(s_0, s_1, ..., s_b)$
- Metric of interest - probability job is turned away
- Given fixed point $\vec{\pi}$, this probability is given by π_b^d

We modify the differential equations as follows -

$$\frac{ds_i}{dt} = \lambda(s_{i-1}^d - s_i^d) - (s_i - s_{i+1}), \forall i < b;$$
$$\frac{ds_b}{dt} = \lambda(s_{b-1}^d - s_b^d) - s_b.$$

Supermarket Model: Extensions

Consider **constant job sizes** instead of being exponentially distributed. To maintain the Markov chain property, we use Erlang's method of stages with r stages

- The state of a queue is given not by number of jobs but by the total number of stages remaining $r(\#\text{ of waiting jobs}) + (\text{remaining stages of current job})$
- We maintain $(s_0, s_1, ...)$ as before but based on the number of stages in each queue
- The modified differential equations are given by -

$$\frac{ds_i}{dt} = \lambda(s_{j-r}^d - s_j^d) - r(s_j - s_{j+1})$$

Medium Access Control

- In a seminal paper published in 1998, Giuseppe Bianchi introduced a simple decoupled model of the IEEE 802.11 DCF
- Bianchi's model had great intuitive appeal and predictive success leading to a whole line of works
- We will discuss how the decoupling assumption made by Bianchi can be justified using mean field theory

Bianchi's Model

Recall the setup for IEEE 802.11 setup from HW-2

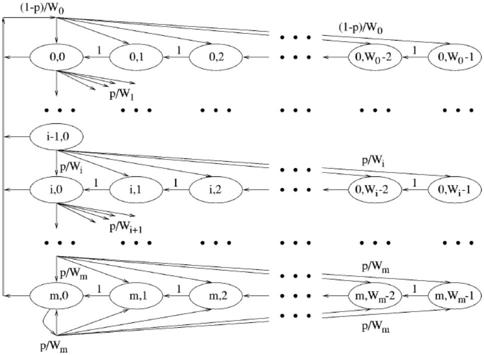


Figure: Image from Bianchi[1998]

Bianchi's Model

- Note that, in general, one would need to maintain a Markov chain that has as its current state the backoff stage and backoff counter for every node
- However, Bianchi makes the **decoupling assumption** - probability of dropping a packet is i.i.d. p for every node
- Now, it suffices to maintain just one chain for a generic node
- Intuitively, all the other nodes influence any given node similarly on average, having a "mean-field" like affect

Mean Field Justifications

- While Bianchi's model predicts throughput with remarkable accuracy, he did not give a rigorous justification for the decoupling assumption
- This was resolved by two works which formally justified the decoupling assumption by using mean field techniques
  - Bordenave, McDonald and Proutière [2005]
  - Sharma, Ganesh and Key [2006]
- However, they deviate from modeling 802.11 DCF exactly, since they assume backoff counters to be geometrically distributed (but with the same mean)

Mean Field Justification

- We focus on the paper by Sharma, Ganesh and Key.
- System state - number of stations in each backoff stage

$$\boldsymbol{X}_n(t) = (X_{n0}(t), X_{n1}(t), ..., X_{nM}(t))$$

- Consider the drift

$$f^{(n)}(x^{(n)}) = \mathbb{E}[\boldsymbol{X}_n(t+1) - \boldsymbol{X}_n(t) | \boldsymbol{X}_n(t) = x^{(n)}]$$

- This is easy to compute using the geometric backoff counter assumption (transmit with probability $2/(W_j - 1)$ in backoff stage j)

Mean Field Justification

- They use the "fluid-limit" scaling $Y_n(t) = \frac{\boldsymbol{X}_n(\lfloor nt \rfloor)}{n}$ to show that
$$\lim_{n\to\infty} \sup_{0\le s\le t} ||Y_n(s) - Y(s)|| = 0 \text{ a.s.}$$
where $Y(\cdot)$ is the solution to
$$\frac{dY(t)}{dt} = F(Y(t)),$$
and $F(\cdot)$ is given by
$$F(x) = \lim_{n\to\infty} f^{(n)}(nx)$$

Mean Field Justification

- The proof goes along similar lines - showing that $F(\cdot)$ is Lipschitz and using Gronwall's identity
- The authors also show that $F(x) = 0$ has a unique solution and hence the DE has a unique fixed point (corresponding to the decoupling assumption)
- However, convergence to this fixed point is only shown for $M = 1$
- Stability for general M conjectured, metastable counterexamples also reported in literature using different modeling assumptions

Conclusion

General recipe for applying mean fields

- Identify a deterministic limiting system that you believe approximates the behavior of a complicated stochastic system
- Identify "nice" properties of the deterministic system, use Kurtz's theorem or something similar to show that it is indeed a good approximation
- Analyze the deterministic system - find its fixed point/s, show convergence/stability around fixed point
- Use this to reason about original stochastic system

References

Our discussion of power of two choices, epidemic models and mean field theory as applied to queuing closely follows Prof. Mitzenmacher's PhD thesis -

- Mitzenmacher, M. D. (1996). "The Power of Two Choices in Randomized Load Balancing."

For an excellent introduction to mean field theory from a statistical physics viewpoint, see

- Kadanoff, L. P. (2009). "More is the same; phase transitions and mean field theories." Journal of Statistical Physics

References

For an actual implementation based on power-of-2-choices see -

- Ousterhout, K., Wendell, P., Zaharia, M., & Stoica, I. (2013). "Sparrow: distributed, low latency scheduling."

For the wireless protocols part of the lecture, see

- Bianchi, G. (2000). "Performance analysis of the IEEE 802.11 distributed coordination function."
- Duffy, K. R. (2010). "Mean field Markov models of wireless local area networks."
- Bordenave, C., McDonald, D., & Proutiere, A. (2005). "Random multi-access algorithms-a mean field analysis."
- Sharma, G., Ganesh, A., & Key, P. (2009). "Performance analysis of contention based medium access control protocols."

References

For examples of works using mean fields in networks, see

- Chaintreau, A., Le Boudec, J. Y., & Ristanovic, N. (2009). "The age of gossip: spatial mean field regime."
- Ying, L., Srikant, R., & Kang, X. (2015). "The power of slightly more than one sample in randomized load balancing."
- Gamarnik, D., Tsitsiklis, J. N., & Zubeldia, M. (2016). "Delay, memory, and messaging tradeoffs in distributed service systems."
- Ying, L. (2017). "Stein's Method for Mean Field Approximations in Light and Heavy Traffic Regimes."
- Li, B., Ramamoorthy, A., & Srikant, R. (2016). "Mean-field-analysis of coding versus replication in cloud storage systems."

The End