Approximate Bayesian Computation

Alireza Shafaei - April 2016

The Problem

- Given a dataset $\mathcal{D}$, we are interested in $\mathcal{P}(\theta|\mathcal{D})$.

The Problem

- Given a dataset $\mathcal{D}$, we are interested in $\mathcal{P}(\theta|\mathcal{D})$.

$$\mathcal{P}(\theta|\mathcal{D}) = \frac{\mathcal{P}(\mathcal{D}|\theta)\pi(\theta)}{\mathcal{P}(\mathcal{D})} \propto \mathcal{P}(\mathcal{D}|\theta)\pi(\theta)$$

The Problem

- Given a dataset $\mathcal{D}$, we are interested in $\mathcal{P}(\theta|\mathcal{D})$.

$$\mathcal{P}(\theta|\mathcal{D}) = \frac{\mathcal{P}(\mathcal{D}|\theta)\pi(\theta)}{\mathcal{P}(\mathcal{D})} \propto \mathcal{P}(\mathcal{D}|\theta)\pi(\theta)$$

$$\mathcal{P}(\mathcal{D}) = \int_{\Theta} \mathcal{P}(\mathcal{D}|\theta)\pi(\theta)\,\mathrm{d}\theta$$

The Problem - A Review

- Previously we looked at the general problem of handling **high-dimensional integrals** and **unnormalized probability** functions.

$$\mathcal{P}(x) = \frac{1}{Z} p^*(x)$$

The Problem - A Review

- Rejection Sampling
  - Given $p^*(x), q(x), M \text{ s.t.} \frac{p^*(x)}{q(x)} \leq M \ \forall x.$
  - $x \sim q(x)$
  - Accept x with probability

$$\frac{p^*(x)}{M \cdot q(x)}$$

The Problem - A Review

- Importance Sampling

$$\int_x f(x)p(x)\,\mathrm{d}x$$

The Problem - A Review

- Importance Sampling

$$\int_x f(x)p(x)\,\mathrm{d}x = \frac{\int_x f(x)p^*(x)\,\mathrm{d}x}{\int_x p^*(x)\,\mathrm{d}x}$$

The Problem - A Review

- Importance Sampling

$$\int_x f(x)p(x)\,\mathrm{d}x = \frac{\int_x f(x)p^*(x)\,\mathrm{d}x}{\int_x p^*(x)\,\mathrm{d}x} = \frac{\int_x f(x)\frac{p^*(x)}{q(x)}q(x)\,\mathrm{d}x}{\int_x \frac{p^*(x)}{q(x)}q(x)\,\mathrm{d}x}$$

The Problem - A Review

- Importance Sampling

$$\int_x f(x)p(x)\,\mathrm{d}x = \frac{\int_x f(x)p^*(x)\,\mathrm{d}x}{\int_x p^*(x)\,\mathrm{d}x} = \frac{\int_x f(x)\frac{p^*(x)}{q(x)}q(x)\,\mathrm{d}x}{\int_x \frac{p^*(x)}{q(x)}q(x)\,\mathrm{d}x}$$

$$\int_x f(x)p(x)\,\mathrm{d}x \approx \frac{\sum_{l=1}^{L} f(x_l)\frac{p^*(x_l)}{q(x_l)}}{\sum_{l=1}^{L} \frac{p^*(x_l)}{q(x_l)}}$$

The Problem - A Review

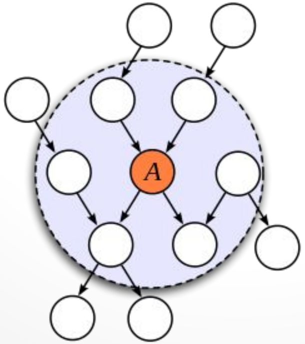
- Markov chain Monte Carlo

The Problem - A Review

- Markov chain Monte Carlo
- Use a transition $q(\theta^{t+1}|\theta^{t})$ to move in the space.

The Problem - A Review

- Markov chain Monte Carlo
- Use a transition $q(\theta^{t+1}|\theta^{t})$ to move in the space.
    - Gibbs sampling.



The Problem - A Review

- Markov chain Monte Carlo
- Use a transition $q(\theta^{t+1}|\theta^{t})$ to move in the space.
  - Gibbs sampling.
  - Metropolis-Hastings algorithm.

The Problem - A Review

- Markov chain Monte Carlo
- Use a transition $q(\theta^{t+1}|\theta^t)$ to move in the space.
  - Gibbs sampling.
  - Metropolis-Hastings algorithm.
  - Reversible Jump MCMC (non-parametric)

The Problem

- Given a dataset $\mathcal{D}$, we are interested in $\mathcal{P}(\theta|\mathcal{D})$.

$$\mathcal{P}(\theta|\mathcal{D}) = \frac{\mathcal{P}(\mathcal{D}|\theta)\pi(\theta)}{\mathcal{P}(\mathcal{D})} \propto \mathcal{P}(\mathcal{D}|\theta)\pi(\theta)$$

The Problem

- Given a dataset $\mathcal{D}$, we are interested in $\mathcal{P}(\theta|\mathcal{D})$.

$$\mathcal{P}(\theta|\mathcal{D}) = \frac{\mathcal{P}(\mathcal{D}|\theta)\pi(\theta)}{\mathcal{P}(\mathcal{D})} \propto \mathcal{P}(\mathcal{D}|\theta)\pi(\theta)$$

- What if we can't calculate $\mathcal{P}(\mathcal{D}|\theta)$?

The Problem

- Apparently applies to a lot of problems in biology.

The Problem

- Apparently applies to a lot of problems in biology.
- Given a parameter θ you can simulate the execution.

The Problem

- Apparently applies to a lot of problems in biology.
- Given a parameter θ you can simulate the execution.
- $\mathcal{P}(\mathcal{D}|\theta)$ Could be intractable or simply no mathematical derivation of it exists.

Approximate Bayesian Computation

1. Draw $\theta \sim \pi(\theta)$

Approximate Bayesian Computation

1. Draw $\theta \sim \pi(\theta)$
2. Simulate $\tilde{\mathcal{D}} \sim \mathcal{P}(\cdot|\theta)$

Approximate Bayesian Computation

1. Draw $\theta \sim \pi(\theta)$
2. Simulate $\tilde{\mathcal{D}} \sim \mathcal{P}(\cdot|\theta)$
3. Accept if $\rho(\mathcal{D}, \tilde{\mathcal{D}}) < \epsilon$

Approximate Bayesian Computation

1. Draw $\theta \sim \pi(\theta)$
2. Simulate $\tilde{\mathcal{D}} \sim \mathcal{P}(\cdot|\theta)$
3. Accept if $\rho(\mathcal{D}, \tilde{\mathcal{D}}) < \epsilon$

- $\epsilon \to \infty \Rightarrow \theta \sim \pi(\theta)$

Approximate Bayesian Computation

1. Draw $\theta \sim \pi(\theta)$
2. Simulate $\tilde{\mathcal{D}} \sim \mathcal{P}(\cdot|\theta)$
3. Accept if $\rho(\mathcal{D}, \tilde{\mathcal{D}}) < \epsilon$

- $\epsilon \to \infty \Rightarrow \theta \sim \pi(\theta)$
- $\epsilon \to 0 \Rightarrow \theta \sim \mathcal{P}(\theta|\mathcal{D})$

Approximate Bayesian Computation

1. Draw $\theta \sim \pi(\theta)$
2. Simulate $\tilde{\mathcal{D}} \sim \mathcal{P}(\cdot|\theta)$
3. Accept if $\rho(\mathcal{S}(\mathcal{D}), \mathcal{S}(\tilde{\mathcal{D}})) < \epsilon$

- $\epsilon \to \infty \Rightarrow \theta \sim \pi(\theta)$
- $\epsilon \to 0 \Rightarrow \theta \sim \mathcal{P}(\theta|\mathcal{D})$

Discussion

- Randomly sampling θ from the prior each time is 'too wasteful'.
  - We want to explore the space to accept more often.

Discussion

- Randomly sampling θ from the prior each time is 'too wasteful'.
  - We want to explore the space to accept more often.
  - Sampling from the prior does not incorporate current observations.

Discussion

- Randomly sampling θ from the prior each time is 'too wasteful'.
  - We want to explore the space to accept more often.
  - Sampling from the prior does not incorporate current observations.
- How do we choose $\rho(\cdot, \cdot), \mathcal{S}(\cdot), \epsilon$?

Approximate MCMC

1. Propose $\theta' \sim Q(\theta'|\theta)$

Approximate MCMC

1. Propose $\theta' \sim Q(\theta'|\theta)$
2. Simulate $\tilde{D} \sim \mathcal{P}(.|\theta')$

Approximate MCMC

1. Propose $\theta' \sim Q(\theta'|\theta)$
2. Simulate $\tilde{\mathcal{D}} \sim \mathcal{P}(.|\theta')$
3. If $\rho(\mathcal{S}(\mathcal{D}), \mathcal{S}(\tilde{\mathcal{D}})) < \epsilon$
    a. Accept θ' with probability

$$\min(1, \frac{\pi(\theta')Q(\theta'|\theta)}{\pi(\theta)Q(\theta|\theta')})$$

Approximate Gibbs

- Let's assume $\theta = (\theta_1, \theta_2)$
  - $\mathcal{P}(\theta_1 | \mathcal{D}, \theta_2)$ is known.
  - $\mathcal{P}(\theta_2 | \mathcal{D}, \theta_1)$ is unknown.

Approximate Gibbs

- Let's assume $\theta = (\theta_1, \theta_2)$
  - $\mathcal{P}(\theta_1|\mathcal{D}, \theta_2)$ is known.
  - $\mathcal{P}(\theta_2|\mathcal{D}, \theta_1)$ is unknown.

1. $\theta_1^{t+1} \sim \mathcal{P}(\theta_1|\mathcal{D}, \theta_2)$

Approximate Gibbs

- Let's assume $\theta = (\theta_1, \theta_2)$
  - $\mathcal{P}(\theta_1|\mathcal{D}, \theta_2)$ is known.
  - $\mathcal{P}(\theta_2|\mathcal{D}, \theta_1)$ is unknown.

1. $\theta_1^{t+1} \sim \mathcal{P}(\theta_1|\mathcal{D}, \theta_2)$
2. $\theta_2^* \sim \pi(\theta_2)$

Approximate Gibbs

- Let's assume $\theta = (\theta_1, \theta_2)$
  - $\mathcal{P}(\theta_1|\mathcal{D}, \theta_2)$ is known.
  - $\mathcal{P}(\theta_2|\mathcal{D}, \theta_1)$ is unknown.

1. $\theta_1^{t+1} \sim \mathcal{P}(\theta_1|\mathcal{D}, \theta_2)$
2. $\theta_2^* \sim \pi(\theta_2)$
   - $\tilde{\mathcal{D}} \sim \mathcal{P}(.|\theta_1^{t+1}, \theta_2^*)$

Approximate Gibbs

- Let's assume $\theta = (\theta_1, \theta_2)$
  - $\mathcal{P}(\theta_1|\mathcal{D}, \theta_2)$ is known.
  - $\mathcal{P}(\theta_2|\mathcal{D}, \theta_1)$ is unknown.

1. $\theta_1^{t+1} \sim \mathcal{P}(\theta_1|\mathcal{D}, \theta_2)$
2. $\theta_2^* \sim \pi(\theta_2)$
   - $\tilde{\mathcal{D}} \sim \mathcal{P}(.|\theta_1^{t+1}, \theta_2^*)$
   - $\rho(\mathcal{S}(\mathcal{D}), \mathcal{S}(\tilde{\mathcal{D}})) < \epsilon \Rightarrow \theta_2^{t+1} = \theta_2^*$
   - else go to 2.

Discussion

Pros

- Likelihood is not needed.

Discussion

Pros

- Likelihood is not needed.
- Easy to implement and parallelize.

Discussion

Pros

- Likelihood is not needed.
- Easy to implement and parallelize.

Cons

- Lot’s of tuning.

Discussion

Pros

- Likelihood is not needed.
- Easy to implement and parallelize.

Cons

- Lot's of tuning.
- For complex problems, sampling from the prior is frustrating because it does not incorporate the evidence.

Discussion

Pros

- Likelihood is not needed.
- Easy to implement and parallelize.

Cons

- Lot’s of tuning.
- For complex problems, sampling from the prior is frustrating because it does not incorporate the evidence.
- How good is our approximation?

Thank you!

References

1. Wilkinson, Richard, and Simon Tavaré. "Approximate Bayesian Computation: a simulation based approach to inference."
2. https://en.wikipedia.org/wiki/Gibbs_sampling
3. Barber, David. Bayesian reasoning and machine learning. Cambridge University Press, 2012.
4.