Approximate Bayesian Computation (ABC)

Isaac J. Michaud

October 27, 2017

Introduction - Bayesian Statistics

- Treat unknown parameters as random variables
  - Distribution describes your uncertainty about the parameter’s true value
  - Data is incorporated to enhance your understanding
  - Loosely: get data to reduce variance of parameter distribution
- Approximate Bayesian computation is a technique for applying Bayes’ rule to compute these updates without many assumptions
- We will do some simple examples
- Apply ABC to locate special nuclear material

Bayesian Inference (in 1 minute)

- $f(x|\theta)$ - Statistical model (how data is generated)
- $\pi(\theta)$ - Prior parameter density (your belief before new data)
- $\pi(\theta|x)$ - Posterior parameter density (your belief after new data)

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\lambda)\pi(\lambda)d\lambda}$$

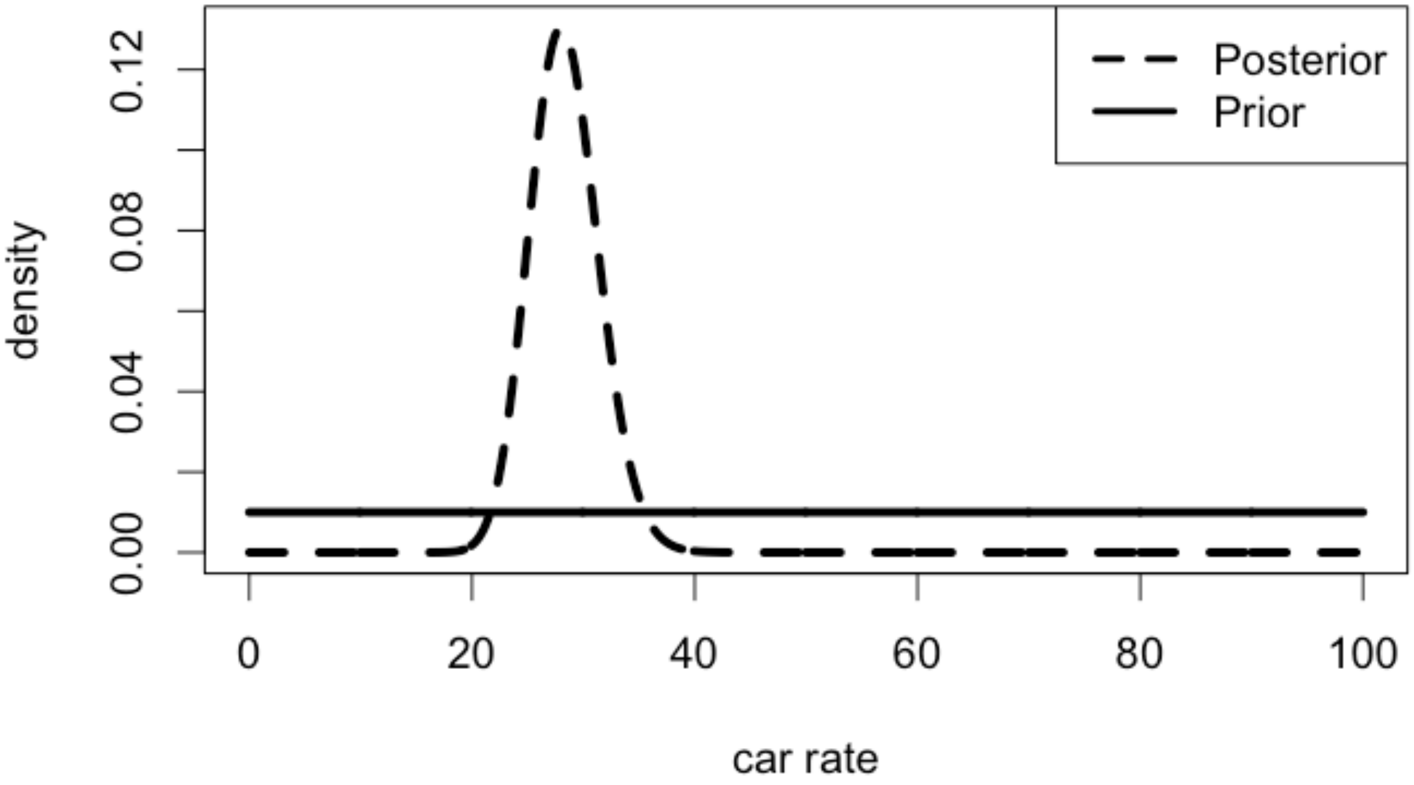
- Inference is done through the posterior distribution of the parameter
- Posterior mean or median, credible intervals, etc.
- Computing the posterior is the name of the Bayesian game

Example: Designing a Highway

- Should we build a 2, 3, or 4 lane highway to ensure fluid circulation?
- $X \sim pois(\theta)$ - number of cars that pass an off ramp in 1 minute
- $\theta \sim unif(0,100)$ - prior belief about the traffic rate
- Data: {25, 27, 32}
- Likelihood: $\mathcal{L}(\theta|\boldsymbol{X} = \{25, 27, 32\}) = f(\boldsymbol{X} = \{25, 27, 32\}|\theta)$
- Posterior: $\pi(\theta|\boldsymbol{x}) \propto \frac{e^{-3\theta}\theta^{84}}{25!27!32!}\left(\frac{1}{100}\right)\mathrm{I}(\theta \in [0,100])$

$\theta|x \sim Gamma(85,1/3)$ (well almost…)

Example: Designing a Highway



The posterior mean is 28.33 (higher than $\bar{x} = 28$).

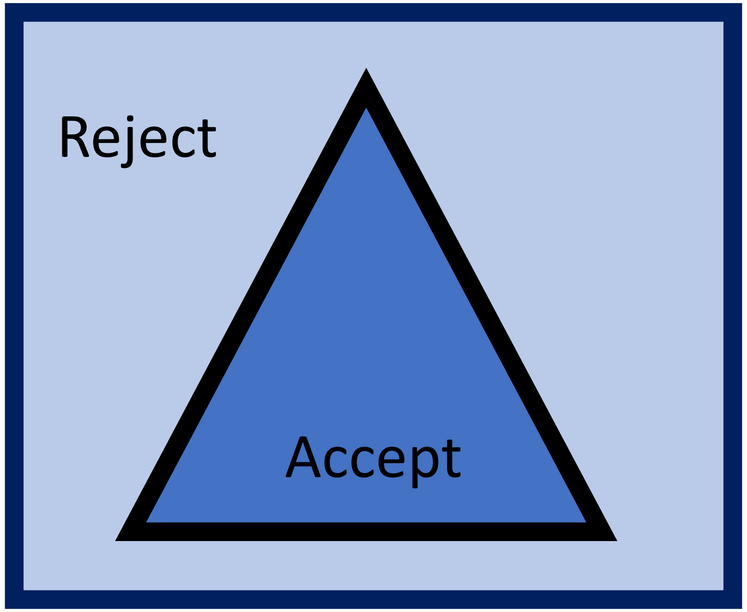
Computing a Bayesian Posterior

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\lambda)\pi(\lambda)d\lambda}$$

- Analytically…
- Integration
  - ➢Numerically (get out your trapezoid rule)
  - ➢Monte Carlo/Importance Sampling
- Markov Chain Monte Carlo (MCMC)
  - ➢Sample from Markov chain whose stationary distribution is $\pi(\theta|x)$
  - ➢Includes Gibbs sampling
- Sequential Monte Carlo (SMC) and Particle Filters

Accept/Reject Sampling

That stuff is too complicated… let's simplify…



Algorithm 1:

1. Draw θ_0 from $\pi(\theta)$
2. Draw x_0 from $f(x|\theta_0)$
3. If $x_0 = x$ then accept θ_0, else reject θ_0

- Repeat until N posterior samples are obtained
- The accepted θ are being sampled from the posterior
- Computationally infeasible (especially if $f(x|\theta)$ is continuous)

Approximate Bayesian Computation (ABC)

- Relax the equality by requiring "closeness"

Algorithm 2:

1. Draw θ_0 from $\pi(\theta)$
2. Draw x_0 from $f(x|\theta_0)$
3. If $d(x_0, x) < \epsilon$ then accept θ_0, else reject θ_0

- d is a suitably chosen metric (typically Euclidean)
- Repeat until N approximate posterior samples are obtained
- Sampling from $\pi(\theta|d(x_0, x) < \epsilon)$ which is not the true posterior

Why would you do this?

- Subtle difference between model $f(x|\theta)$ and likelihood $\mathcal{L}(\theta|x)$
- Model describes “what if” scenarios (forward direction)
- Likelihood explains what has happened (backward direction)

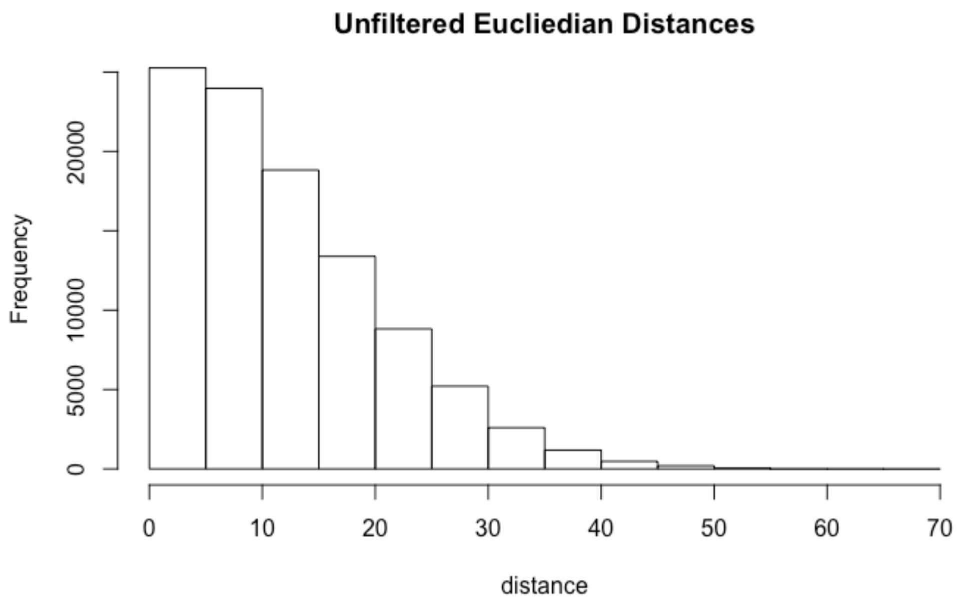
Example:

1. You don’t study for the Basic Exam. How are you going to do?
2. You did poorly on the Basic Exam. Why?

- It is often easier to create models than compute likelihoods
  - Population Genetics, Astronomy, Systems Biology
  - Stochastic dynamical systems, simulations, etc.

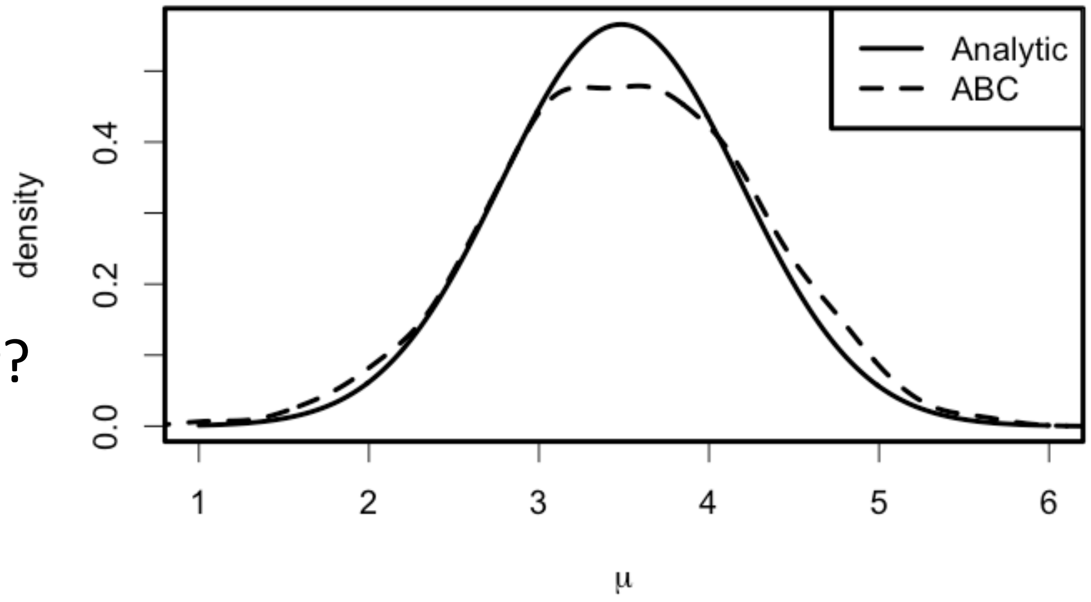
Example: Normal Model

- Model $X_i \sim N(\mu, 1), iid, i = 1, \ldots, n$
- Prior $\mu \sim N(0, 100)$
- Observed Data: $x_1 = 3, x_2 = 4$
- Posterior $\mu | x_1, x_2 \sim N(3.48, 0.50)$



ABC Approach:

- $\mathrm{M} = 100{,}000$ prior draws of μ
- $\epsilon = 0.6$, accepted the smallest 1000 distances
- Results are good, but can we make them better?



Improving ABC Performance

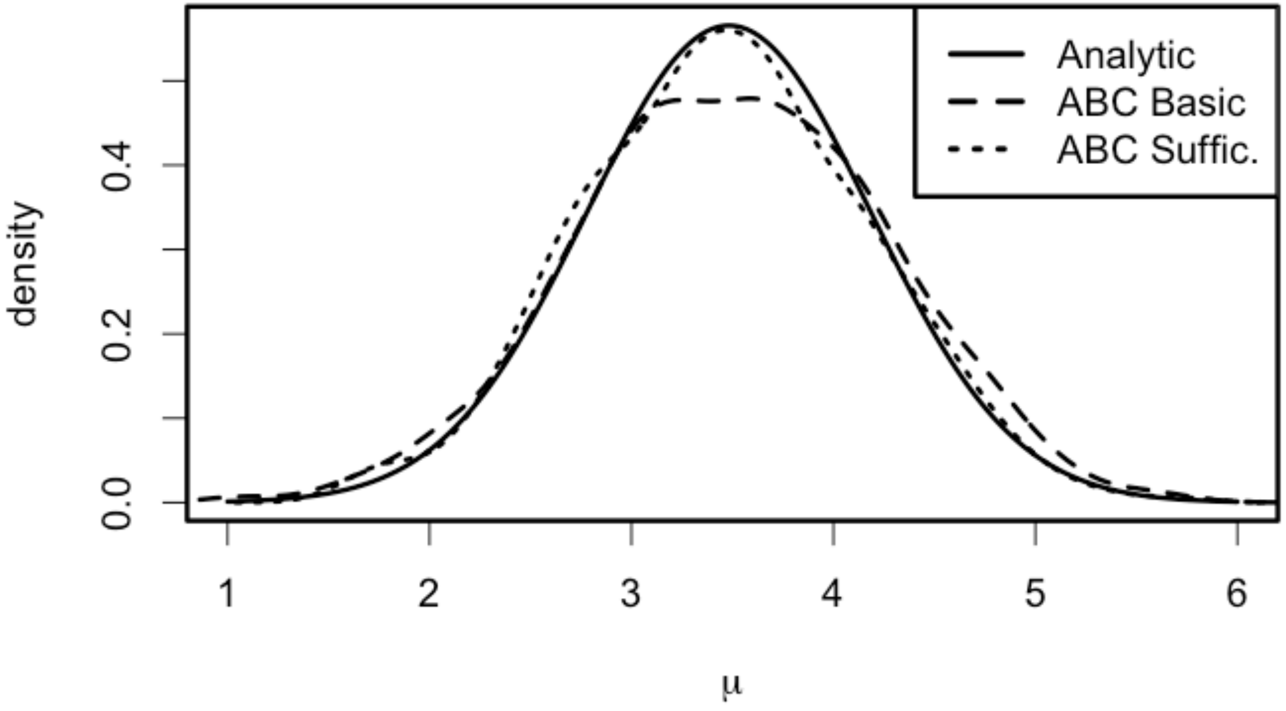
- Make M larger and ϵ smaller (can I have my PhD now?)
- Work smarter not harder! Reduce the dimension of the problem
- **Definition:** A statistic $T(\boldsymbol{X})$ is <u>sufficient</u> for parameter θ if the conditional distribution of $\boldsymbol{X}$ given $T(\boldsymbol{X})$ is free of θ, or in other words

$$f(\boldsymbol{X}|\theta) = \mathrm{g}(T(\boldsymbol{X})|\theta)h(\boldsymbol{X})$$

- Thanks Casella and Berger!
- Key idea: a sufficient statistic summarizes all the information about the parameter that is contained a sample (e.g. sample of 100 summarized by a sample mean)

ABC with Sufficiency

- Back to our normal example
- $X_1, X_2 \sim N(\mu, 1) \rightarrow \bar{X} \sim N\left(\mu, \frac{1}{\sqrt{2}}\right)$
- $M = 100{,}000$ prior draws of μ
- Accepted the smallest 1000 distances
- ABC posterior definitely improved
- Sufficiency gives you more bang for your computational dollar



Wait One Second!

- The reason for doing ABC was we couldn’t compute the likelihood.
- How can we factor a likelihood we can’t compute?

Magic!

- Approximate sufficient statistics
- Use methods like PCA and maximum entropy to find informative functions of the data and hope for the best

Theory

Let $\boldsymbol{x}^*$ be the observed data and $p_{T|\theta}$ be the density of the sufficient statistic then:

1. If $t \to p_{T|\theta}\ (t|\theta)$ is continuous and uniformly bounded in the neighborhood of $t^* = T(\boldsymbol{x}^*)$ for all θ then as $\epsilon \to 0^+$ we sample θ from the true posterior $p(\theta|\boldsymbol{x}^*)$ using the ABC algorithm

2. The average number of samples required to get a single accepted parameter is $\mathcal{O}(\epsilon^{-q})$ where $q = \dim(t^*)$

Radiation Sensor Networks (CNEC)

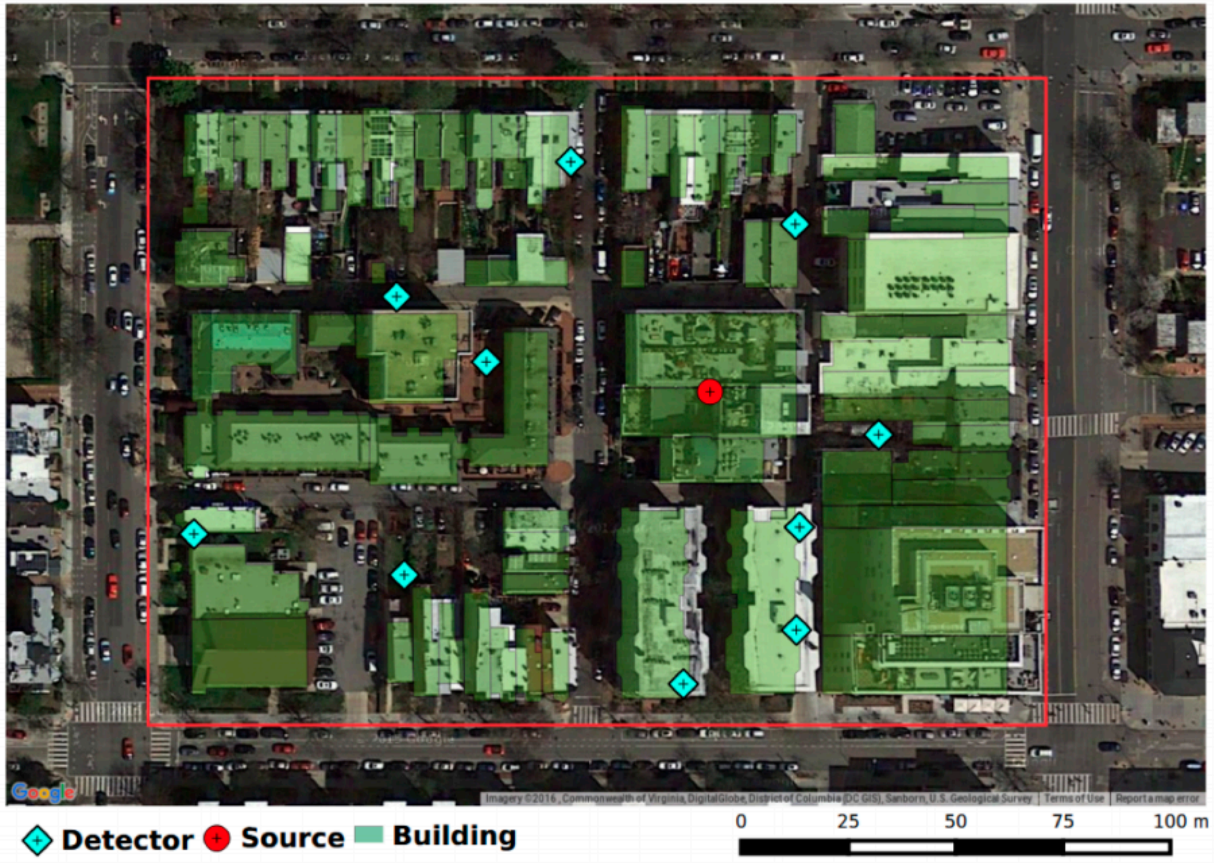
- Radiation source in an urban environment
- Sensors observe gamma ray counts, $Pois(\Gamma_i)$

$$\Gamma_i = \underbrace{\frac{\epsilon_i A_i I_0}{4\pi |r_i - r_0|^2} \exp\left(-\sum_j \sigma_{n_j} s_{n_j}\right)}_{\text{Radiation Source}} + B_i$$

- Goal is to infer source location and intensity (x,y,i)

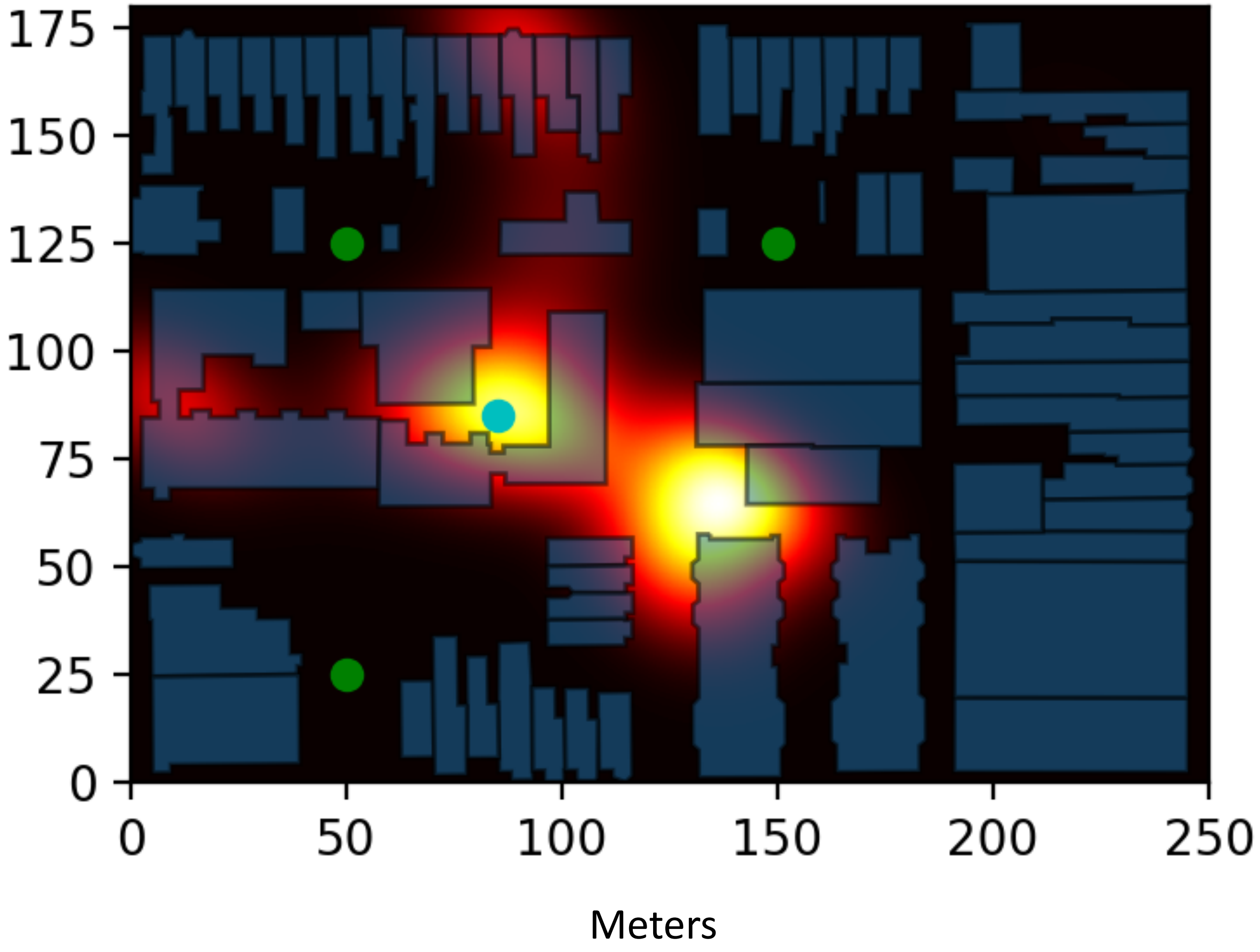
Issues:

- Small amount of data
- Likelihood is not smooth
- Needs to be real-time



Source Posterior

- Source (cyan dot)
  - ➢Uniform priors
- Three sensors (green dots)
  - ➢Five obs. per sensor
- Sensor mean responses are sufficient (three dimensional)
- $M = 50{,}000$ samples
- Accept 1000 best parameters
- Run time $\approx$ 30 minutes (single core)



Caveat Emptor

- You need to be cautious
  - ➢ABC can be applied in really complicated settings
  - ➢You will probably not have analytic or numerical verification of posterior
- ABC is computationally expensive (true of most Bayesian methods)
- Consider scaling sufficient statistics so that a single statistic does not dominate the distance metric
- Don’t be happy with the first thing that comes out
  - ➢Uses different epsilons
  - ➢Use ABC multiple times with different random seeds, compare results
  - ➢Check for sensitivity to the tuning parameters

Conclusions

- ABC provides a quick and dirty way of estimating the posterior
- Works in cases where you cannot or are unwilling to compute a likelihood function
- These problems occur frequently in science when there is a forward model that describes the evolution of a system over time
- Although computationally expensive it is easily parallelizable

References and More Examples

- If you like Bayesian statistics and socks:

http://www.sumsar.net/blog/2014/10/tiny-data-and-the-socks-of-karl-broman/

Voss, Jochen. "Beyond Monte Carlo." *An Introduction to Statistical Computing: A Simulation-based Approach*: 181-211.

Marin, Jean-Michel, et al. "Approximate Bayesian computational methods." *Statistics and Computing* (2012): 1-14.

R package: abc