Stochastic Compartmental Models

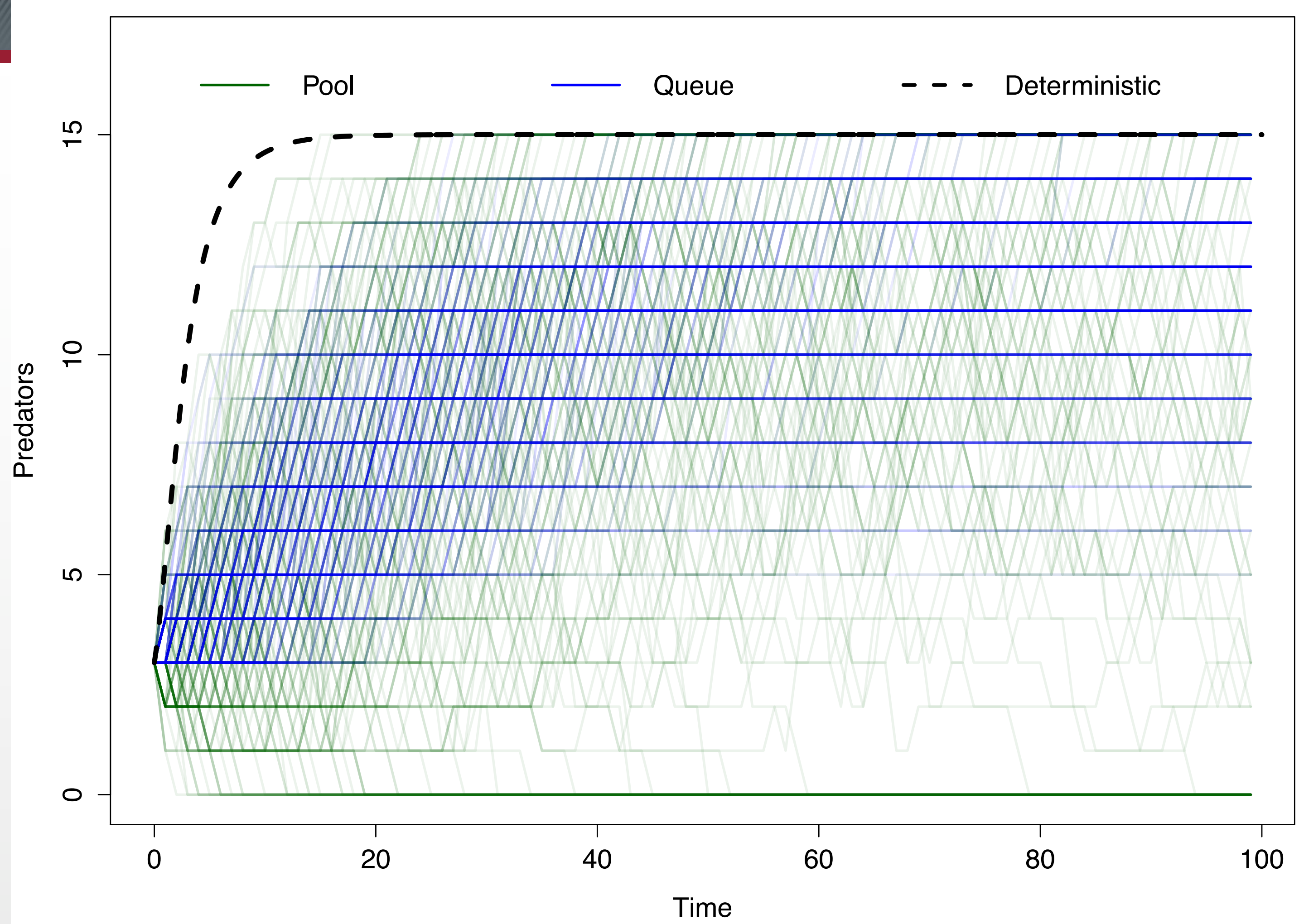
Questions from Last Class?

Consider the Attofox

- Deterministic models can result in fractional units
- These often don't make sense biologically
  - Half a sick person, or the "Attofox" (i.e. 1×10^{-18} foxes)
- In a large population, this might not matter
  - 0.0001% of New York city is still ~20 people
- This is a much bigger deal in a small population, where you start talking about the difference between zero and one

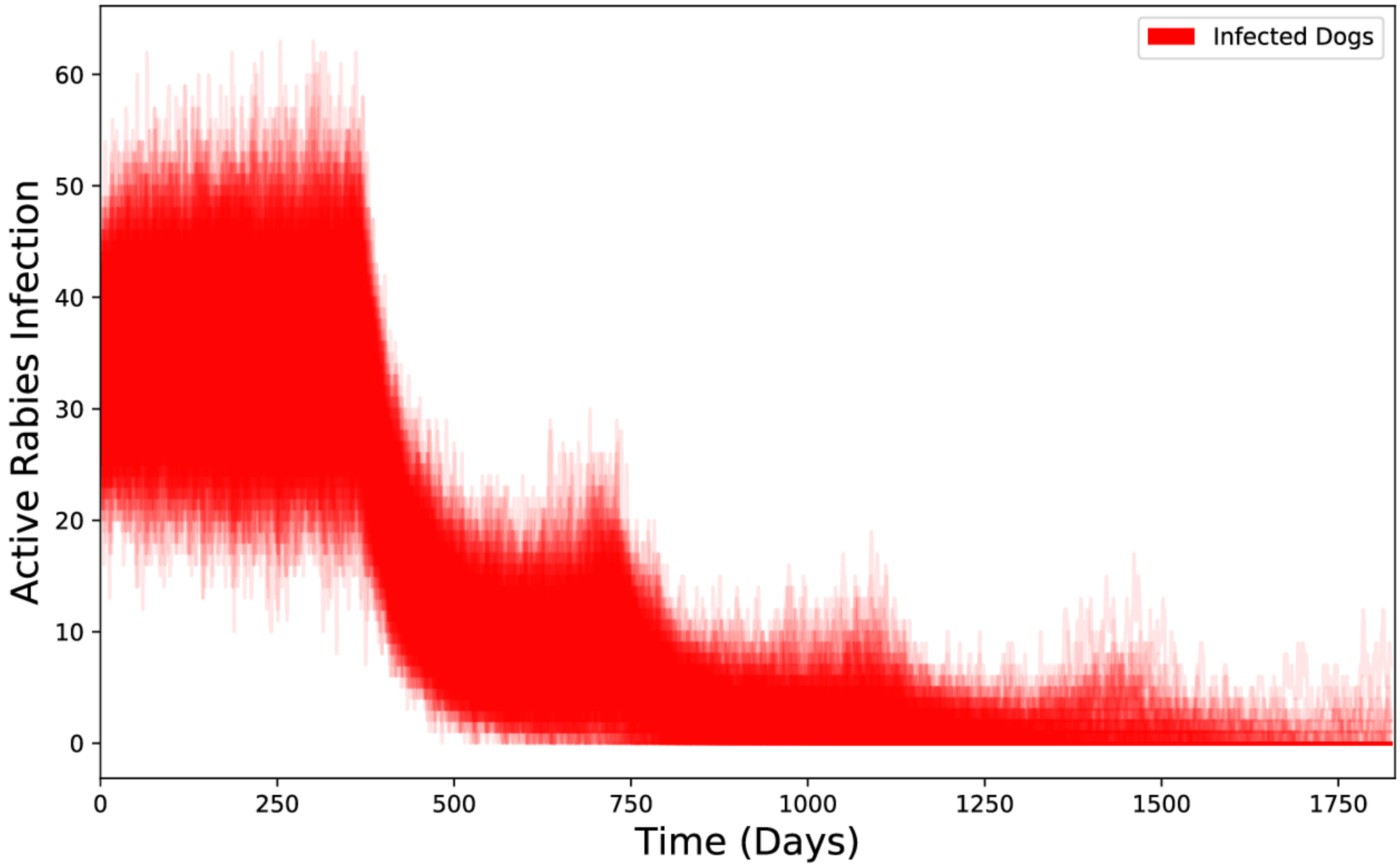
Stochastic Extinction

- Extinction due to randomness
- I = 0 in an SIR model
- In a deterministic model, this is essentially impossible – there will always be some infinitesimally small population of infected individuals who can reignite an outbreak
- This can have very large consequences for planning – for example, the periodic fluctuations we see when we add births to a model may not occur

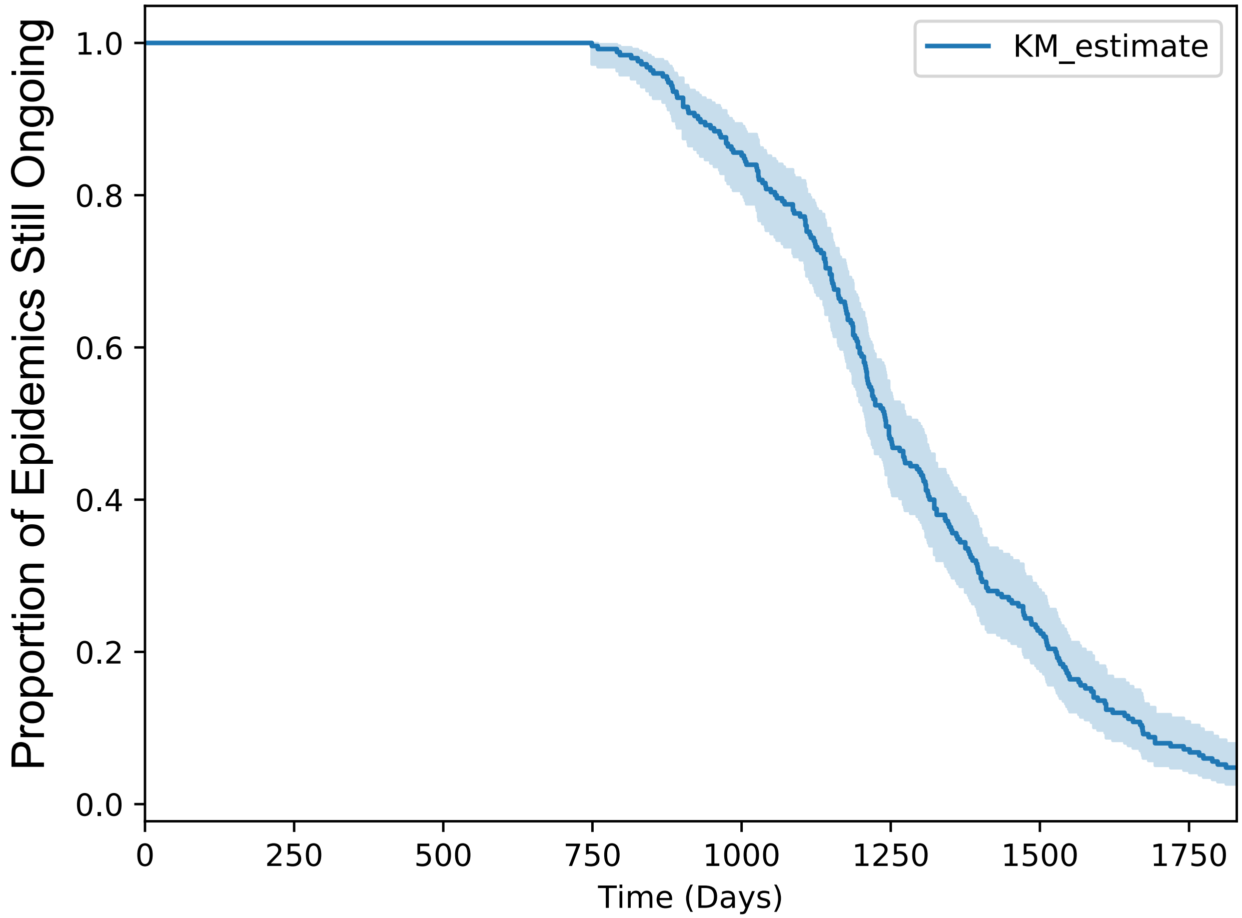

Pool
Queue
Deterministic
Predators
0
5
10
15
0
20
40
60
80
100
Time

Other Aspects of Randomness

- $R_0 > 1$ doesn't *guarantee* an outbreak, if merely makes it more likely
- Similarly, $R_0 < 1$ can still cause enough cases that we might view it as a public health problem
- Can induce a great deal of uncertainty into the system just due to random chance



Infected Dogs
Active Rabies Infection
Time (Days)
0
10
20
30
40
50
60
0
250
500
750
1000
1250
1500
1750



Proportion of Epidemics Still Ongoing
1.0
0.8
0.6
0.4
0.2
0.0
KM_estimate
0
250
500
750
1000
1250
1500
1750
Time (Days)

More Subtlety

- What does "steady-state" mean in this context
- As mentioned previously, 1/10th of a room leaving and everyone having a 10% chance of leaving have very different possibilities
- These are the kind of things you have to consider when working on stochastic and noisy systems

How Small is "Small"

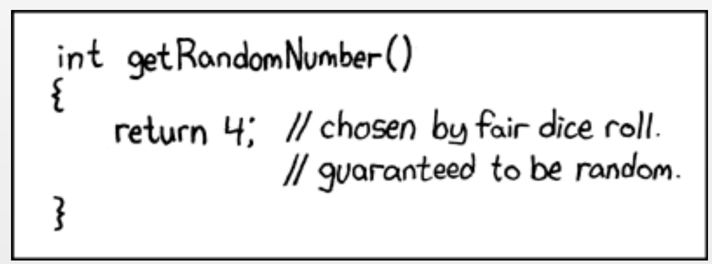
- This is something of an open question, with a lot of "well obviously…" type answers
- New York is large
- An 18-bed ICU is small
- But an entire hospital is also small
- As is Pullman, WA
- But what about Spokane?
- My assertion: For most animal and wildlife systems, you're almost certainly working with a small population

Ways of Implementing Stochasticity

- Stochastic Differential Equations
  - Similar to Ordinary Differential Equations (discussed last week) but with the addition of a random noise process
  - Have many of the same appealing traits (analytical solutions etc.)
- Stochastic simulation methods
  - What we're going to cover

Random Numbers

- Stochastic models use tremendous amounts of random numbers
- How do random number generators work?
- What is a “seed” and why do I care?
- V&V using random numbers

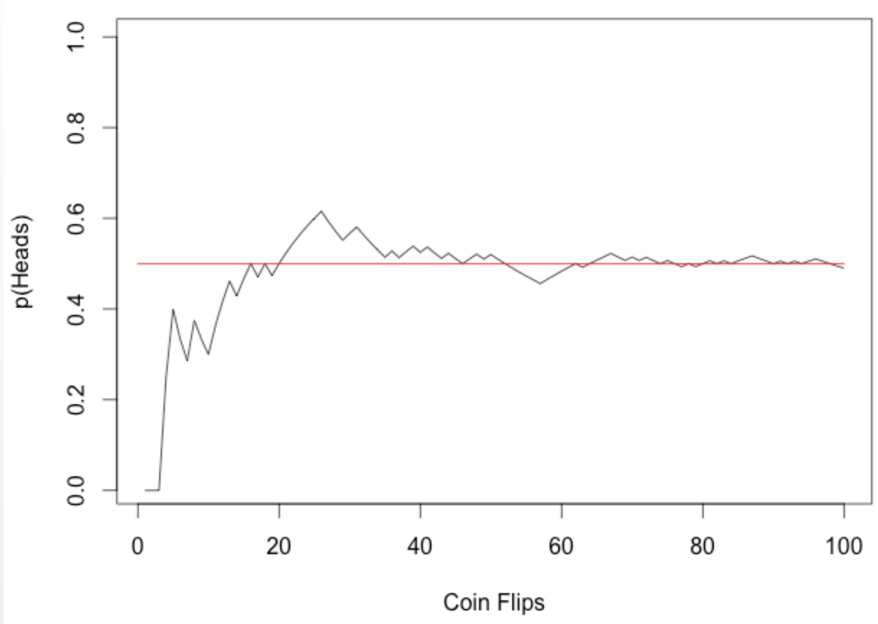


Analysis

- Every time you change a parameter, you create a new counterfactual scenario
- Many/most simulations are very amenable to basic statistical analysis – t-tests, ANOVA, etc.
- Caveats
  - p-values do not mean what you think they mean
  - Plot all your data at least once – multimodal, non-normal, etc. distributions are quite common

Why Simulations Worry About Sample Size

- Law of Large Numbers
- *Not* statistical power
- Goal is to converge on an answer and minimize the impact of extreme random numbers



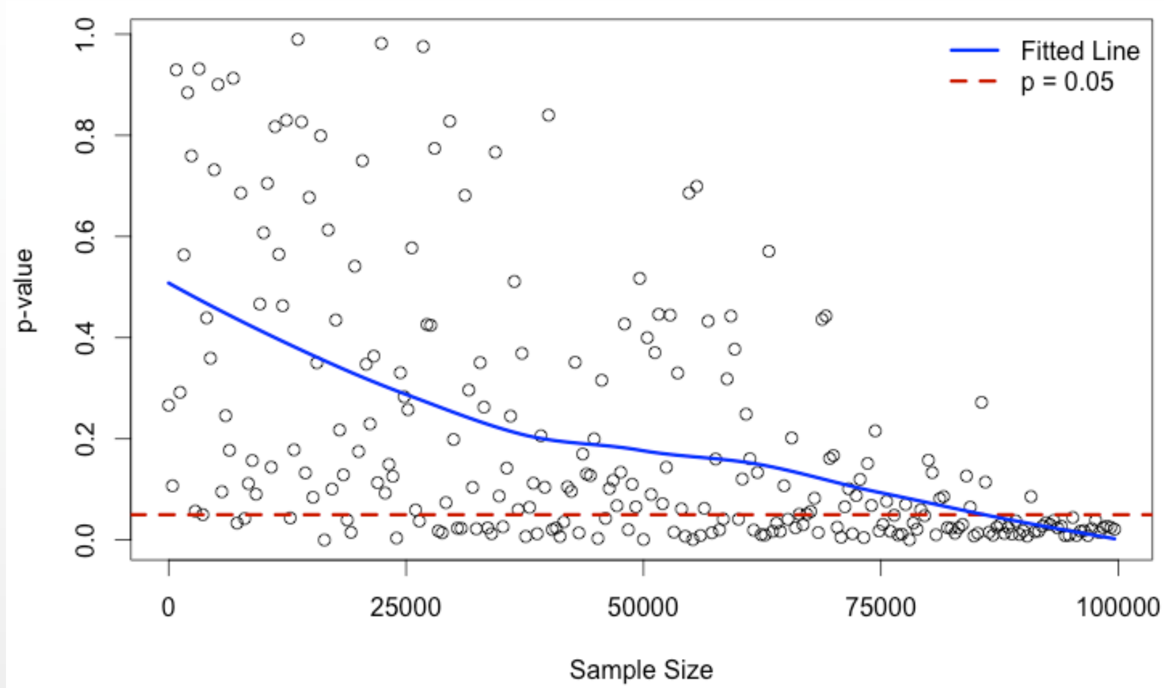
On P-values

- They're a dangerous thing at the best of times
- Observational Study:
  - *f*(Sample Size, Effect Size, Test,α)
  - All but sample size essentially fixed
  - Sample size is hard to increase – limited source population, recruitment is hard and expensive
- Simulation Study:
  - All those factors
  - But what determines simulation sample size?
    - *f*(Computing Power, Patience)
  - Power is now something trivially modified by the researcher
    - Clusters, cloud computing, three-day weekends

- True difference:

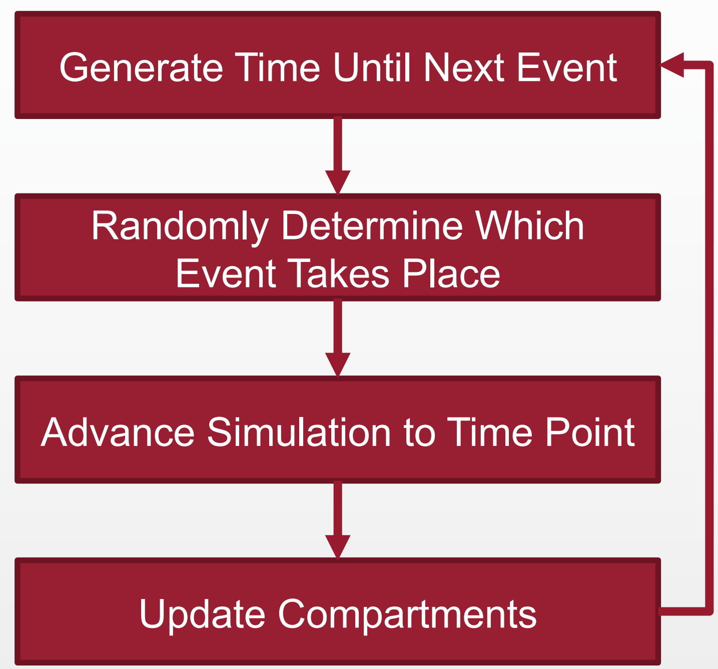
RR = 3.37 vs. 3.3701

- All it took was 100,000 runs of the model
- Average 3.37 seconds / run
- 4 processor cores = 25,000 runs / core
- ~ 7 hours of wall time
- All of that overnight
- It could have been even shorter if I had been more actively fishing



Gillespie’s Direct Method

- Originally used by Daniel Gillespie for computational chemistry
- Exact stochastic simulation of ODE-based compartmental models
  - Faster approximate algorithms are available
- Converts deterministic *rates* into stochastic probabilities, and treats individuals as indivisible, integer-valued units
- Scales poorly as population increases (time between events goes down)



Rates to Probabilities

- This is a pretty direct translation
- If something happens every 2 days, the probability of it happening on a single day is…1/2

Totaling Up All Rates

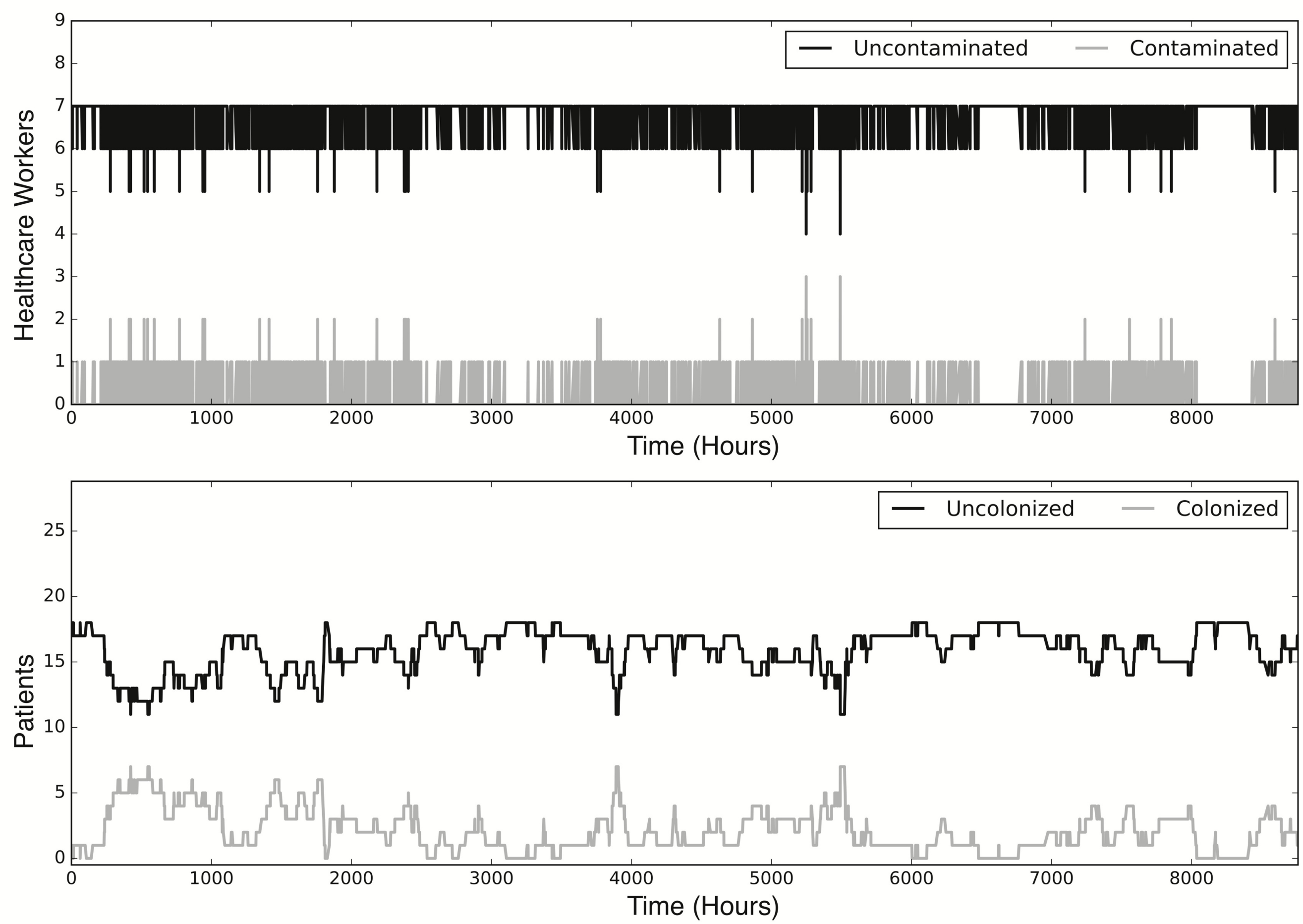
- Let’s consider two events
- One that takes place every 2 days
- One that takes place every 4 days
- *An* event takes place every 1.33 days
  - 0.5 + 0.25 = 0.75. 1/0.75 = 1.33

Randomly Determine Which Event Happens

- Draw an event randomly weighted by the frequency those events occur
- For example:
  - 0 to 0.667 = Event 1
  - 0.668 to 1 = Event 2
- This means events that happen more frequently are *more likely* but not guaranteed

Update Compartments

- Big advantage of the Gillespie Algorithm (and many stochastic methods) is the treatment of compartment populations as integer valued units
- *An* event occurred – someone moved from one compartment to another
- It's useful to think of these as a list of reactions/transitions
  - S to I, I to R, etc.

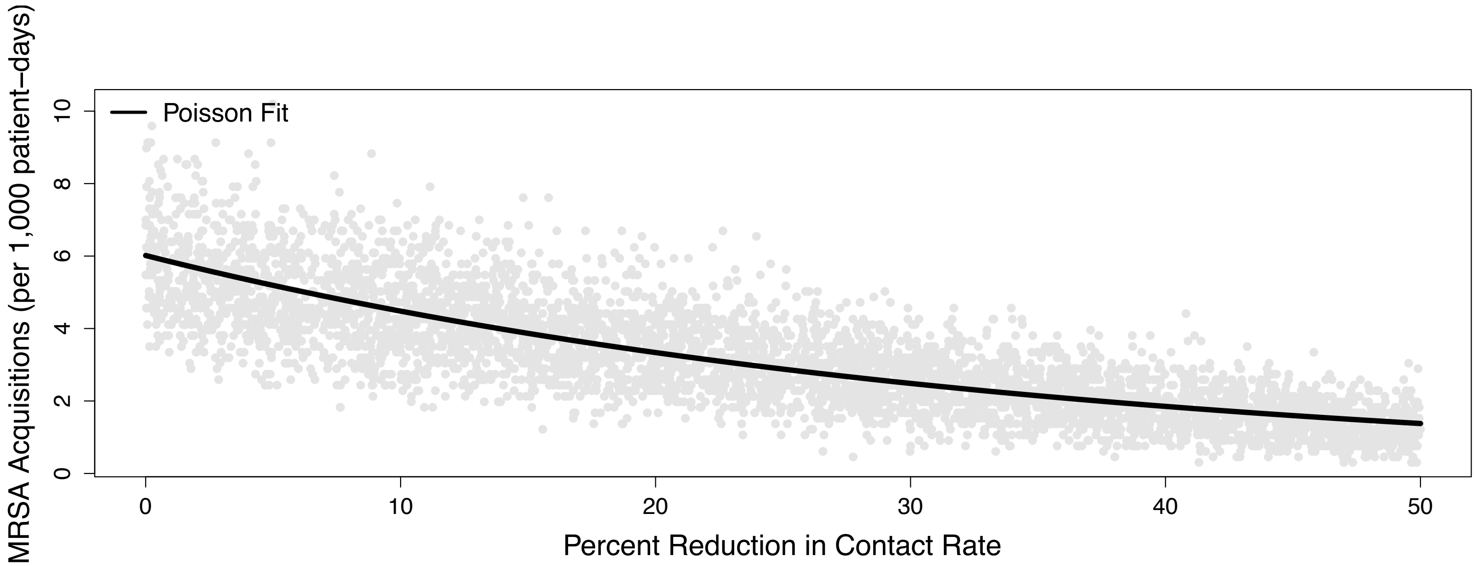

Uncontaminated
Contaminated
Healthcare Workers
Time (Hours)
Uncolonized
Colonized
Patients
Time (Hours)

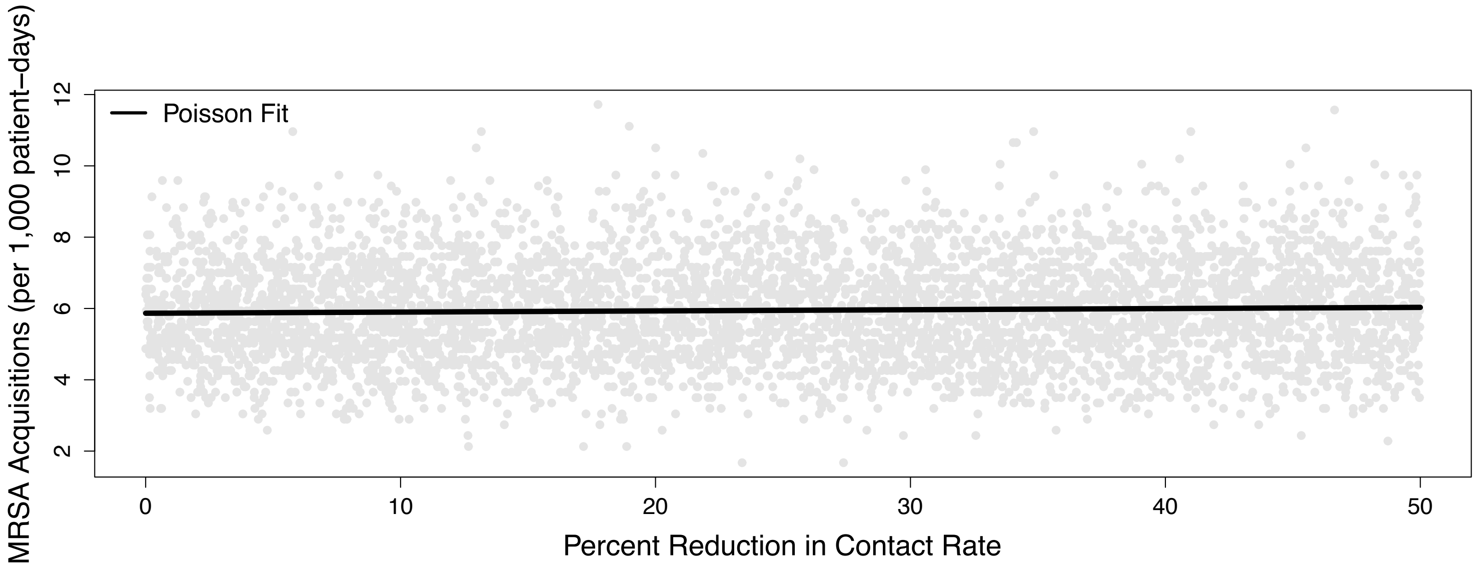
Other Algorithms

- Gillespie's Direct Method (what we just went over) is an exact method (it gives you a statistically valid realization of a particular stochastic process)
- It's also very computationally intensive
- It scales poorly with population size (more people means more events means slower simulations)
- There are *approximate* methods that are much faster
  - Tau leaping
  - These work well, but may behave strangely at thresholds

Other Types of Uncertainty

- Stochastic models *only* account for uncertainty due to random chance
- They do not address uncertainty in parameter estimates
- That can be done by drawing your parameters from a distribution and repeatedly solving the equations
  - Either stochastically or deterministically


Poisson Fit
MRSA Acquisitions (per 1,000 patient–days)
0
2
4
6
8
10
0
10
20
30
40
50
Percent Reduction in Contact Rate


Poisson Fit
MRSA Acquisitions (per 1,000 patient–days)
2
4
6
8
10
12
0
10
20
30
40
50
Percent Reduction in Contact Rate