# Approximate Bayesian Computation

Dr. Jarad Niemi

STAT 615 - Iowa State University

December 5, 2017

# Outline

- Stochastic kinetic models
- Approximate Bayesian computation

# Stochastic kinetic models

Imagine a *well-mixed* system in *thermal equilibrium* with

- $N$ species: $S_1, \ldots, S_N$ with
- number of molecules $X_1, \ldots, X_N$ with elements $X_j \in \mathbb{Z}^+$
- which change according to $M$ reactions: $R_1, \ldots, R_M$ with
- propensities $a_1(x), \ldots, a_M(x)$.
- The propensities are given by $a_j(x) = \theta_j h_j(x)$
- where $h_j(x)$ is a known function of the system state.
- If reaction $j$ occurs, the state is updated by the stoichiometry $\nu_j$ with
- elements $\nu_{ij} \in \{-2, -1, 0, 1, 2\}$, i.e. reaction orders 0,1, and 2.

# Michaelis-Menton System

The Michaelis-Menton system has $N = 4$ species:

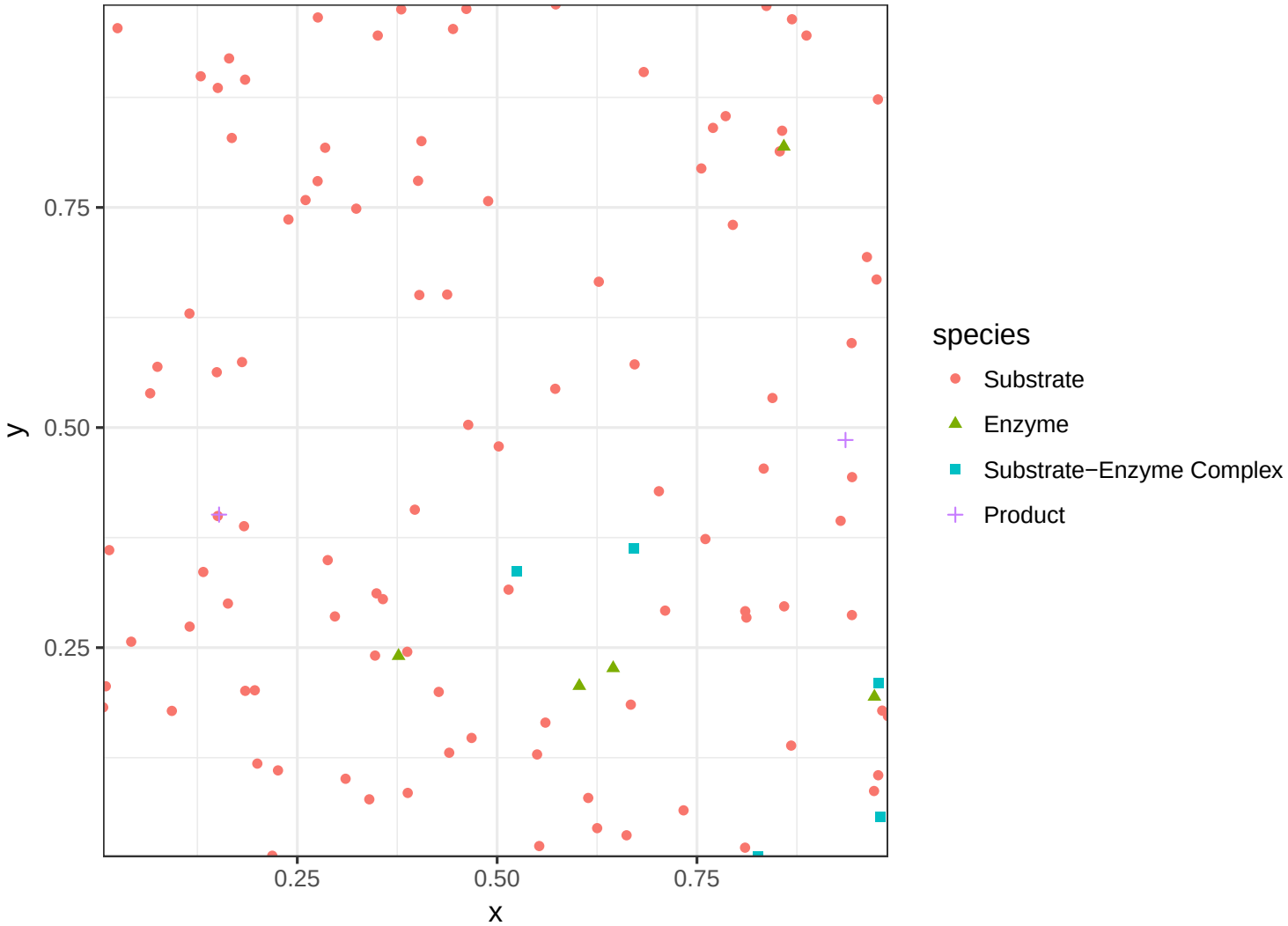
- Substrate (S),
- Enzyme (E),
- Substrate-Enzyme Complex (SE), and
- Product (P).

The $M = 3$ reactions as well as their propensities and stoichiometries are

| Reaction | Propensity | Stoichiometry S | E | SE | P |
|---|---|---|---|---|---|
| S + E $\longrightarrow$ SE | $\theta_1 X_S X_E$ | -1 | -1 | 1 | |
| SE $\longrightarrow$ S + E | $\theta_2 X_{SE}$ | 1 | 1 | -1 | |
| SE $\longrightarrow$ P+E | $\theta_3 X_{SE}$ | | 1 | -1 | 1 |

where $\theta = (\theta_1, \theta_2, \theta_3)$ is the parameter of interest.

# Michaelis-Menton snapshot



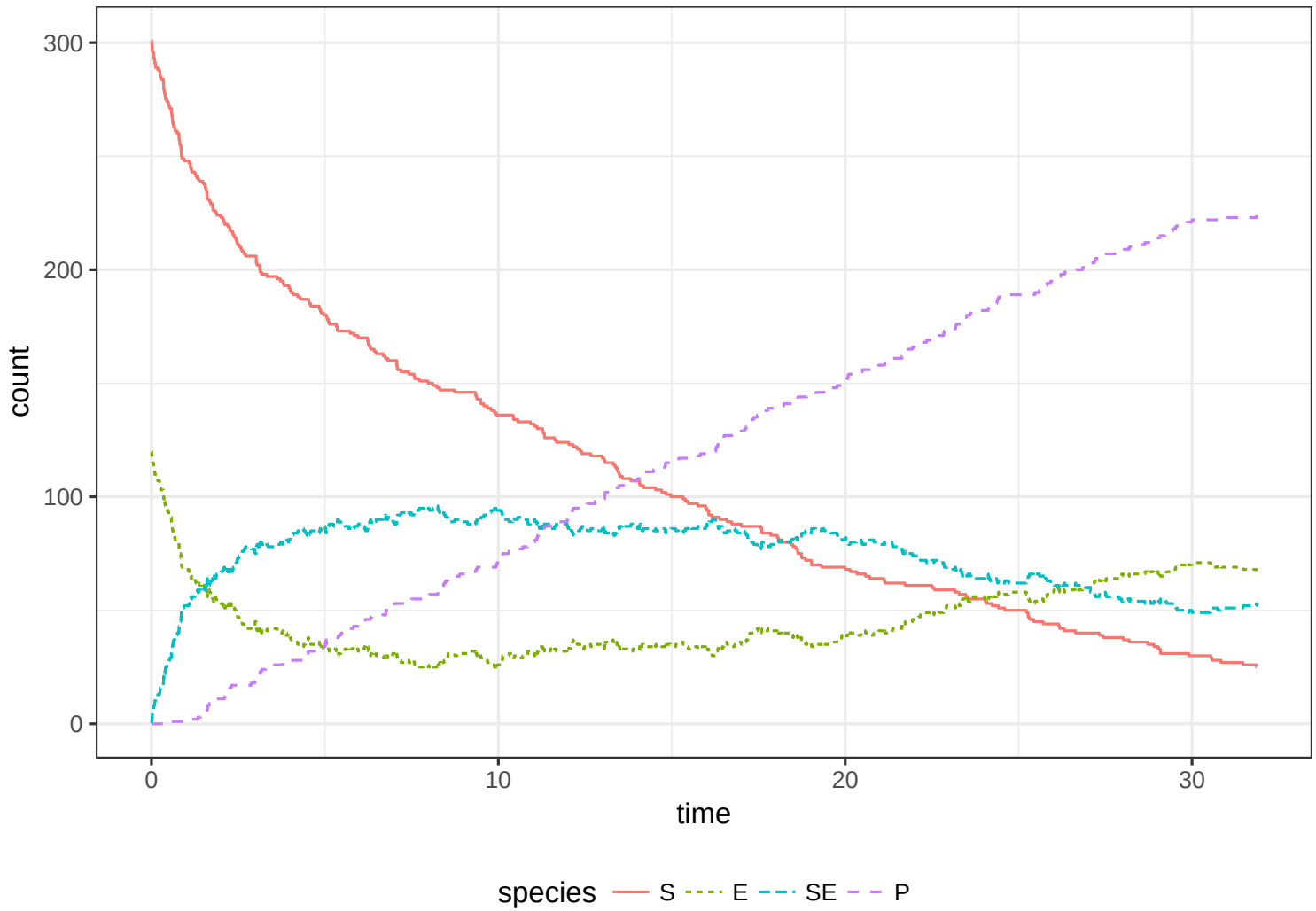
# Gillespie algorithm

- If reaction $j \in \{1, \ldots, M\}$ has the following probability

$$\lim_{dt \to 0} P(\text{reaction } j \text{ within the interval } (t, t+dt)|X_t) = a_j(X_t)dt,$$

  then this defines a continuous-time Markov jump process.
- Then a realization from this model can be obtained using the Gillespie algorithm:
  1. For $j \in \{1, \ldots, M\}$, calculate $a_j(X_t)$.
  2. Calculate $a_0(X_t) = \sum_{j=1}^{M} a_j(X_t)$.
  3. Simulate a reaction time $\tau \sim Exp(a_0(X_t))$
  4. Simulate a reaction id $k \in \{1, \ldots, M\}$ with probability $a_k(X_t)/a_0(X_t)$
  5. Update $X$ according to $v_k$ and time by $\tau$.

# Michaelis-Menton Gillespie Simulation



# Complete observations

Suppose you observe all system transitions:

- $n$ reactions occur in the interval $[0, T]$
- $t_1, \ldots, t_n$ are the reaction times
- $r_1, \ldots, r_n$ are the reaction indicators, $r_i \in \{1, \ldots, M\}$

Then inference can be performed based on the likelihood

$$L(\theta) \propto \prod_{j=1}^{M} \theta_j^{n_j} \exp\left(-\theta_j I_j\right)$$

where

$$n_j = \sum_{i=1}^{n} \mathrm{I}(r_i = j) \quad \# \text{ of } j \text{ reactions}$$

$$I_j = \int_0^T h_j(X_t)dt = \sum_{i=1}^{n} h_j(X_{t_{i-1}})(t_i - t_{i-1}) + h_j(X_{t_n})[T - t_n]$$

# Inference

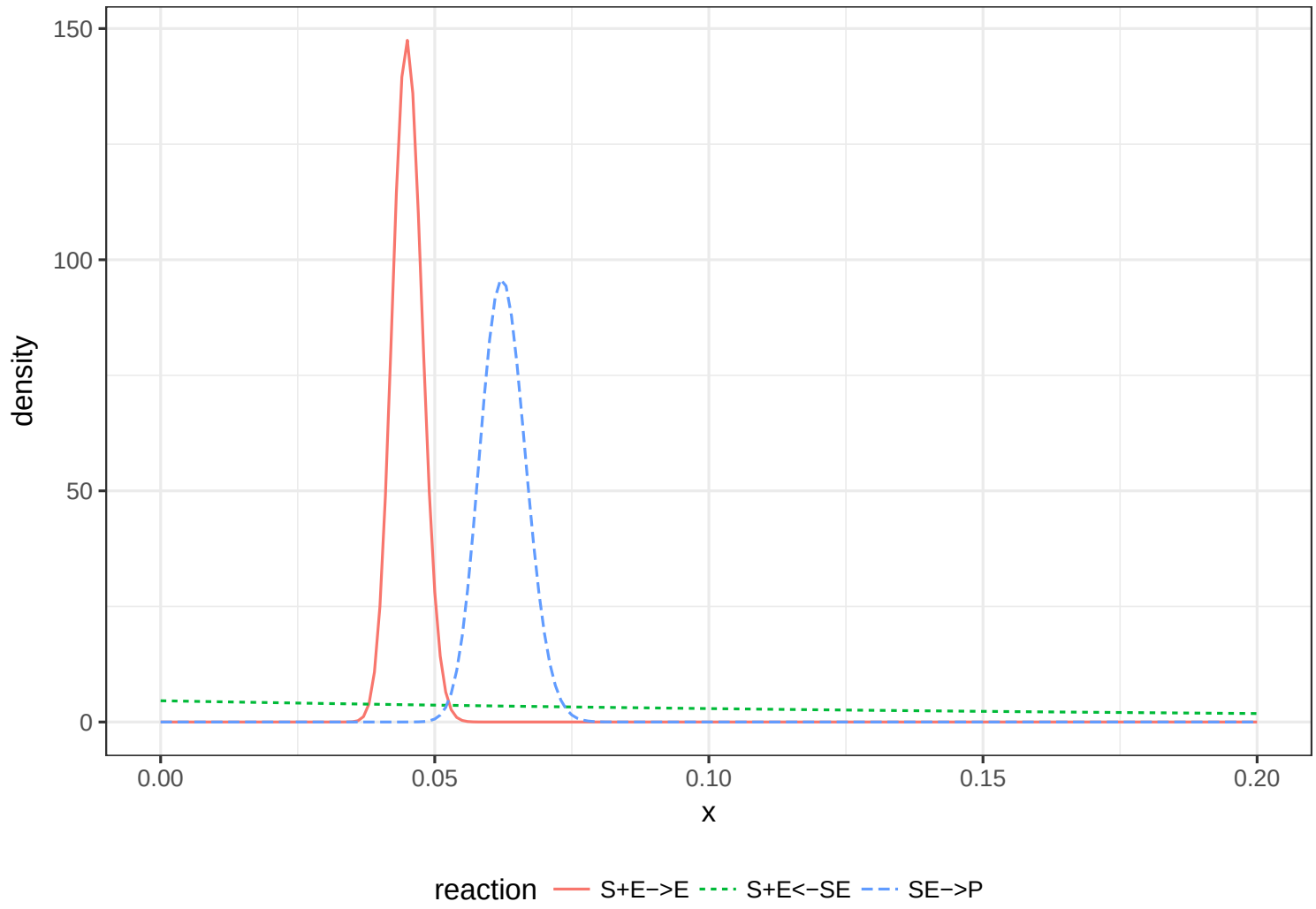
- Maximum likelihood estimation

$$\hat{\theta}_j = \frac{n_j}{I_j}$$

- Conjugate Bayesian inference

$$\begin{aligned} p(\theta) &= \prod_{j=1}^{M} Ga(\theta_j; \alpha_j, \beta_j) \\ p(\theta|X) &= \prod_{j=1}^{M} Ga(\theta_j; \alpha_j + n_j, \beta_j + I_j) \\ E[\theta_j|X] &= \frac{\alpha_j + n_j}{\beta_j + I_j} \end{aligned}$$

# Michaelis-Menton Complete Data Inference

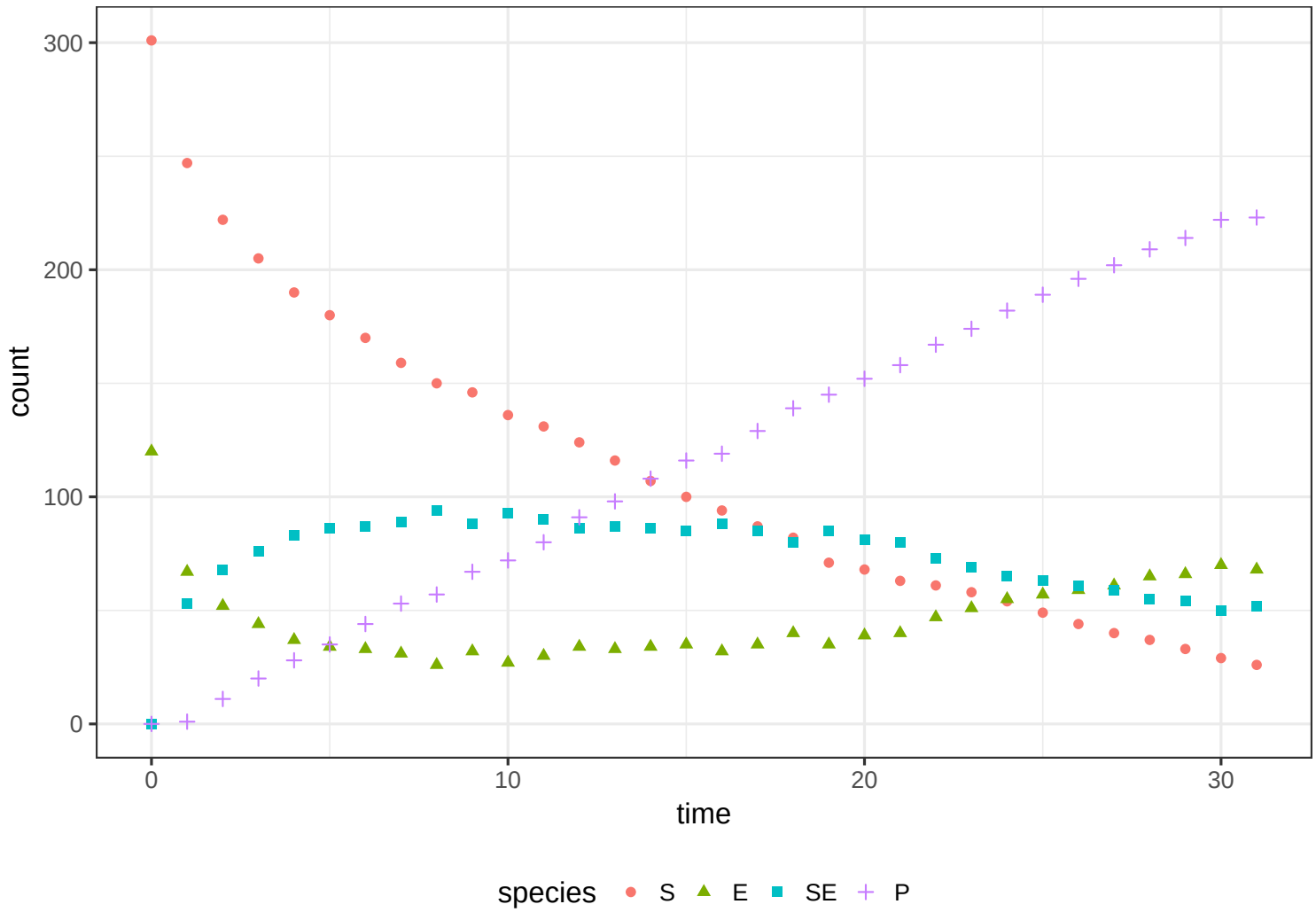


# Discrete observations

Suppose you only observe the system at discrete-times:

- For simplicity, observe the system at times $t = 1, 2, \ldots, T$.
- At these times, we observe $y_t = X_t$ the system state.
- But do not observe the system between these times.

# Michaelis-Mention discrete observations



# Inference

Inference is still performed based on the likelihood

$$L(\theta) = p(y|\theta) = p(t, y)$$

but this is the solution to the chemical master equation

$$\frac{\partial}{\partial t}p(t, y) = \sum_{j=1}^{M} \left(a_j(y - v_m)p(t, y - v_m) - a_j(y)p(t, y)\right)$$

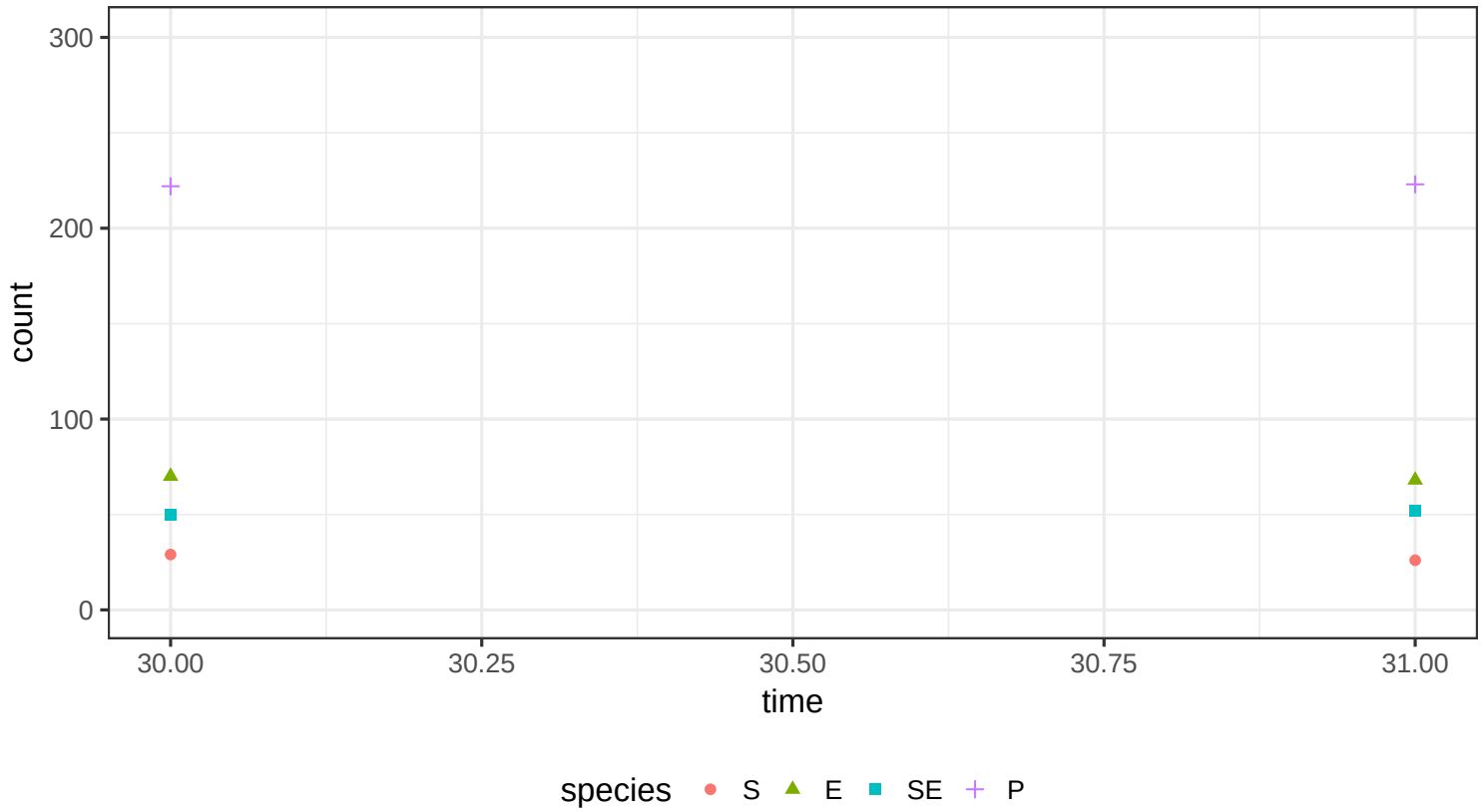
For constitutive production $h(X_t) = 1$ and $a(X_t) = \theta$, we still have

$$L(\theta) \propto \theta^n \exp\left(-\theta I\right)$$

with

$$n = y_T - y_0 \qquad I = \int_0^T 1dt = T$$

# Reversible isomerization




- How many reactions occurred in the interval $[30, 31]$?
- What is $\int_{30}^{31} X_{SE} dt$?

# Summary

- With complete observations and independent gamma priors, the posterior is

$$p(\theta|X) = \prod_{j=1}^{M} Ga(\theta_j; \alpha_j + n_j, \beta_j + I_j)$$

where

$$\begin{aligned} n_j &= \textstyle\sum_{i=1}^{n} \mathrm{I}(r_i = j) \\ I_j &= \textstyle\int_0^T h_j(X_t)dt = \sum_{i=1}^{n} h_j(X_{t_{i-1}})(t_i - t_{i-1}) + h_j(X_{t_n})[T - t_n] \end{aligned}$$

- For discrete observations, the likelihood is analytically intractable and therefore no closed form exists for the posterior (or MLEs).

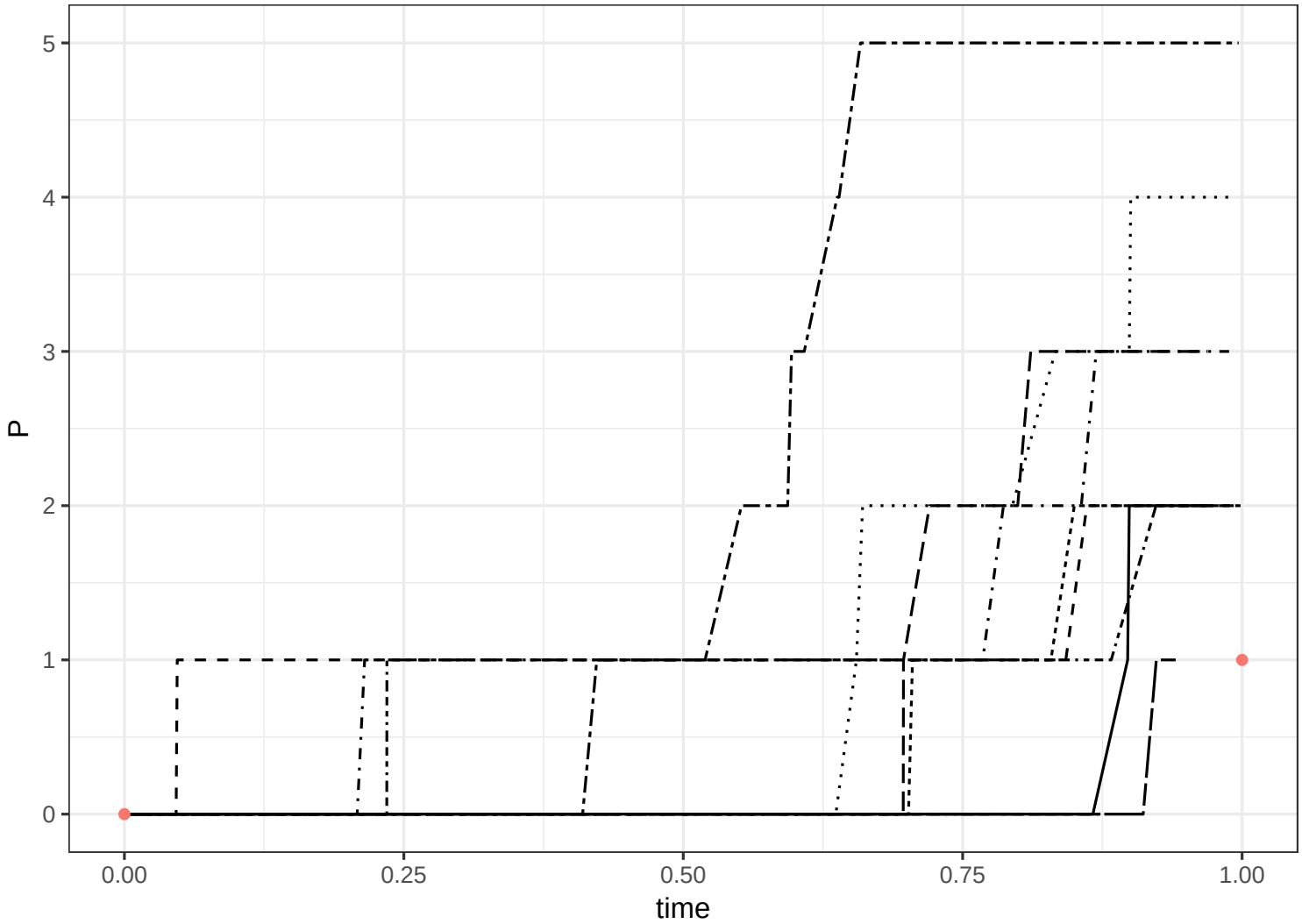
# The idea

- But we can simulate from the model using the Gillespie algorithm!!
- Intuitively, if we
  1. pick a set of parameters,
  2. simulate a realization using these parameters,
  3. and it matches our data,
  4. then these parameters should be reasonable.

Our goal is to formalize this through

1. Rejection sampling
2. Gibbs sampling

# Simulations from the prior



# Rejection sampling

Our objective is samples from the posterior

$$\begin{aligned} p(\theta|y) &= \int p(\theta, X|y)dX \propto \int p(y|X)p(X|\theta)p(\theta)dX \\ &= \int \prod_{t=1}^{n} \mathrm{I}(y_t = X_t)p(X|\theta)p(\theta)dX \end{aligned}$$

A rejection sampling procedure is

1. Sample $\theta \sim p(\theta)$.
2. Sample $X \sim p(X|\theta)$ a.k.a. Gillespie
3. If $y_t = X_t$ for $t = 1, 2, \ldots, T$, then
4. $\theta$ is a sample from $p(\theta|y)$ and
5. $\theta, X$ is a sample from $p(\theta, X|y)$.

# Gibbs sampling

Our objective is samples from the posterior

$$p(\theta|y) = \int p(\theta, X|y)dX \propto \int p(y|X)p(X|\theta)p(\theta)dX$$

A Gibbs sampling procedure is

1. Start with $\theta^{(0)}, X^{(0)}$
2. For $k = 1, \ldots, K$,
   a. Sample $\theta^{(k)} \sim p(\theta|X^{(k-1)})$
   b. Sample $X^{(k)} \sim p(X|\theta^{(k)}, y)$ a.k.a. rejection sampling

$\theta^{(k)}, X^{(k)}$ converge to samples from $p(\theta, X|y)$

# An approximate posterior

- Intuitively, if we
  1. pick a set of parameters,
  2. simulate a realization using these parameters,
  3. and it is similar to our data,
  4. then these parameters should be reasonable.

- We can formalize this using
  - Approximate Bayesian computation

# Approximate Bayesian computation (ABC)

Our approximate objective is samples from the posterior

$$p(\theta|y) = \int p(\theta, X|\rho \leq \epsilon)dX \propto \int \mathrm{I}(\rho \leq \epsilon)p(X|\theta)p(\theta)dX$$

where $\rho = \rho(y, X)$ is a measure of the difference between your data $y$ and simulations $X$.

- Choice of $\epsilon$ reflects tension between computability and accuracy.
  - As $\epsilon \to \infty$,
    - $p(\theta|\rho \leq \epsilon) \xrightarrow{d} p(\theta)$
    - acceptance probability converges to 1
  - As $\epsilon \to 0$,
    - $p(\theta|\rho \leq \epsilon) \xrightarrow{d} p(\theta|y)$
    - acceptance probability decreases

# ABC rejection sampling

Let $\rho = \sum_{t=1}^{n} |y_t - X_t|$ and $\epsilon = n$,

An ABC rejection sampling procedure is

1. Sample $\theta \sim p(\theta)$
2. Sample $X \sim p(X|\theta)$ a.k.a. Gillespie
3. If $\rho(y, X) \leq \epsilon$, then
4. $\theta$ is a sample from $p(\theta|\rho \leq \epsilon)$ and
5. $\theta, X$ is a sample from $p(\theta, X|\rho \leq \epsilon)$.

# ABC Gibbs sampling

Let $\rho = \sum_{t=1}^{n} |y_t - X_t|$ and $\epsilon = n$,

A Gibbs sampling procedure is

1. Start with $\theta^{(0)}, X^{(0)}$
2. For $k = 1, \ldots, K$,
   a. Sample $\theta^{(k)} \sim p(\theta | X^{(k-1)})$
   b. Sample $X^{(k)} \sim p(X | \theta^{(k)}, \rho \leq \epsilon)$ a.k.a. rejection sampling

$\theta^{(k)}, X^{(k)}$ converge to samples from $p(\theta, X | \rho \leq \epsilon)$

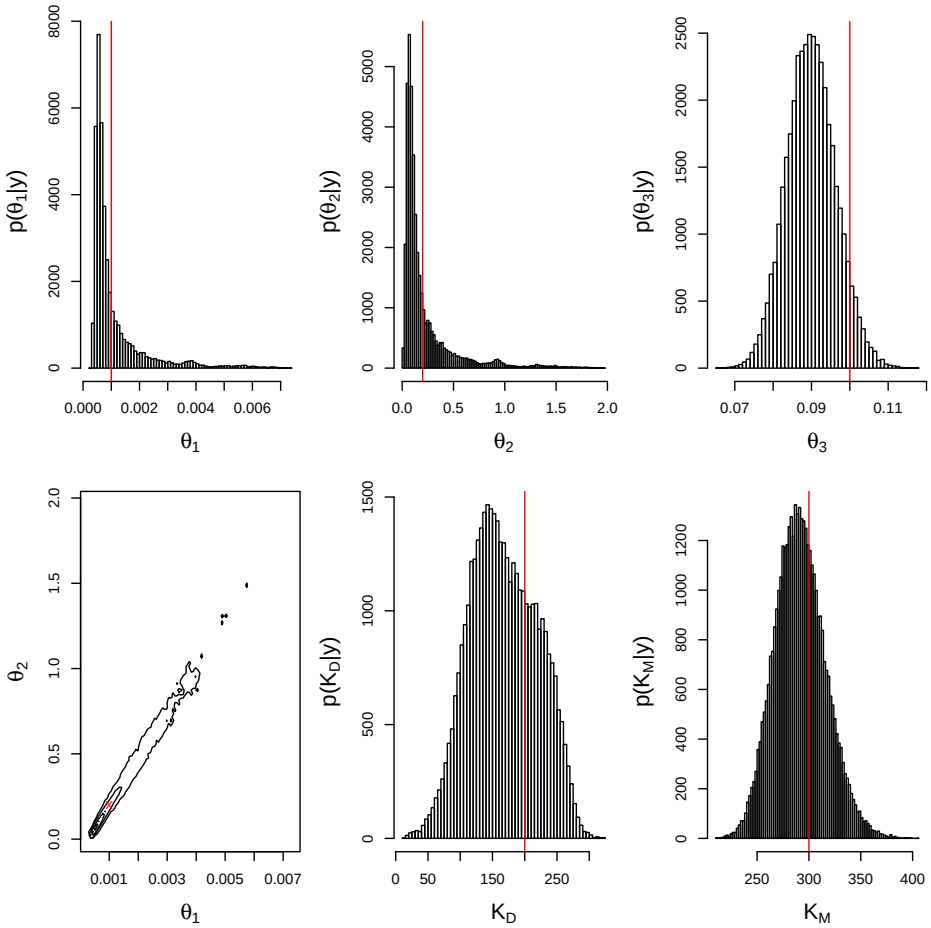
# Michaelis-Menton system

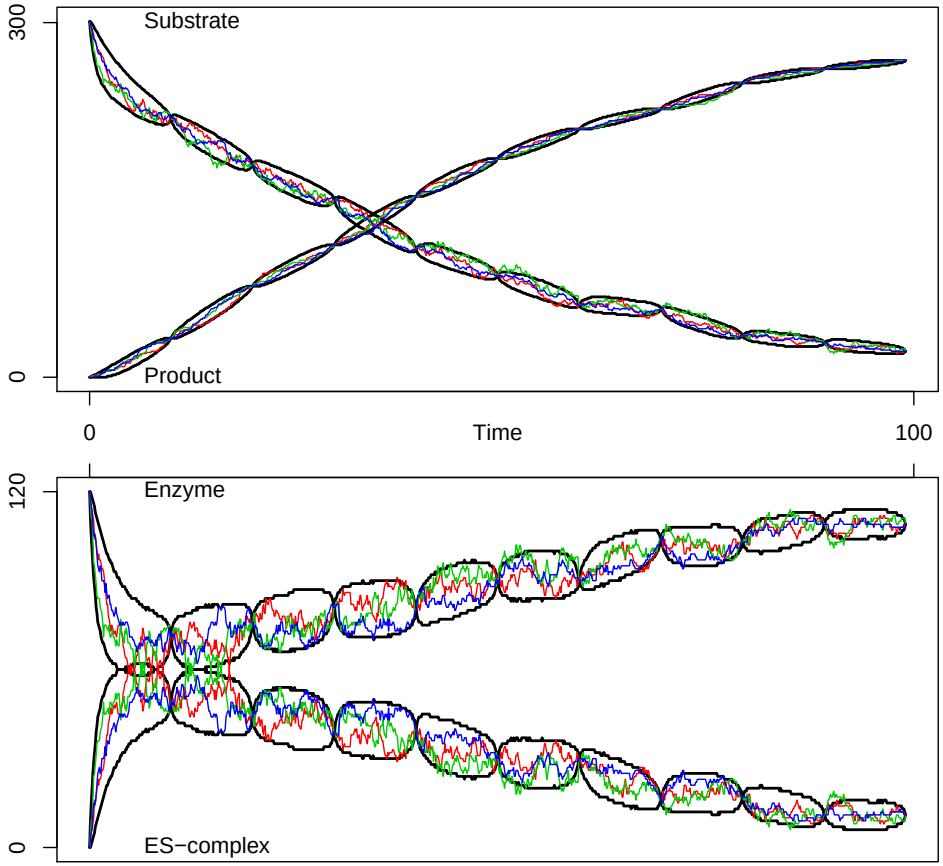
$$E + S \underset{\theta_2}{\overset{\theta_1}{\rightleftharpoons}} ES \xrightarrow{\theta_3} E + P$$

Table : Measurements taken from a simulated Michaelis-Mention system with parameters $\theta_1 = 0.001$, $\theta_2 = 0.2$, and $\theta_3 = 0.1$.

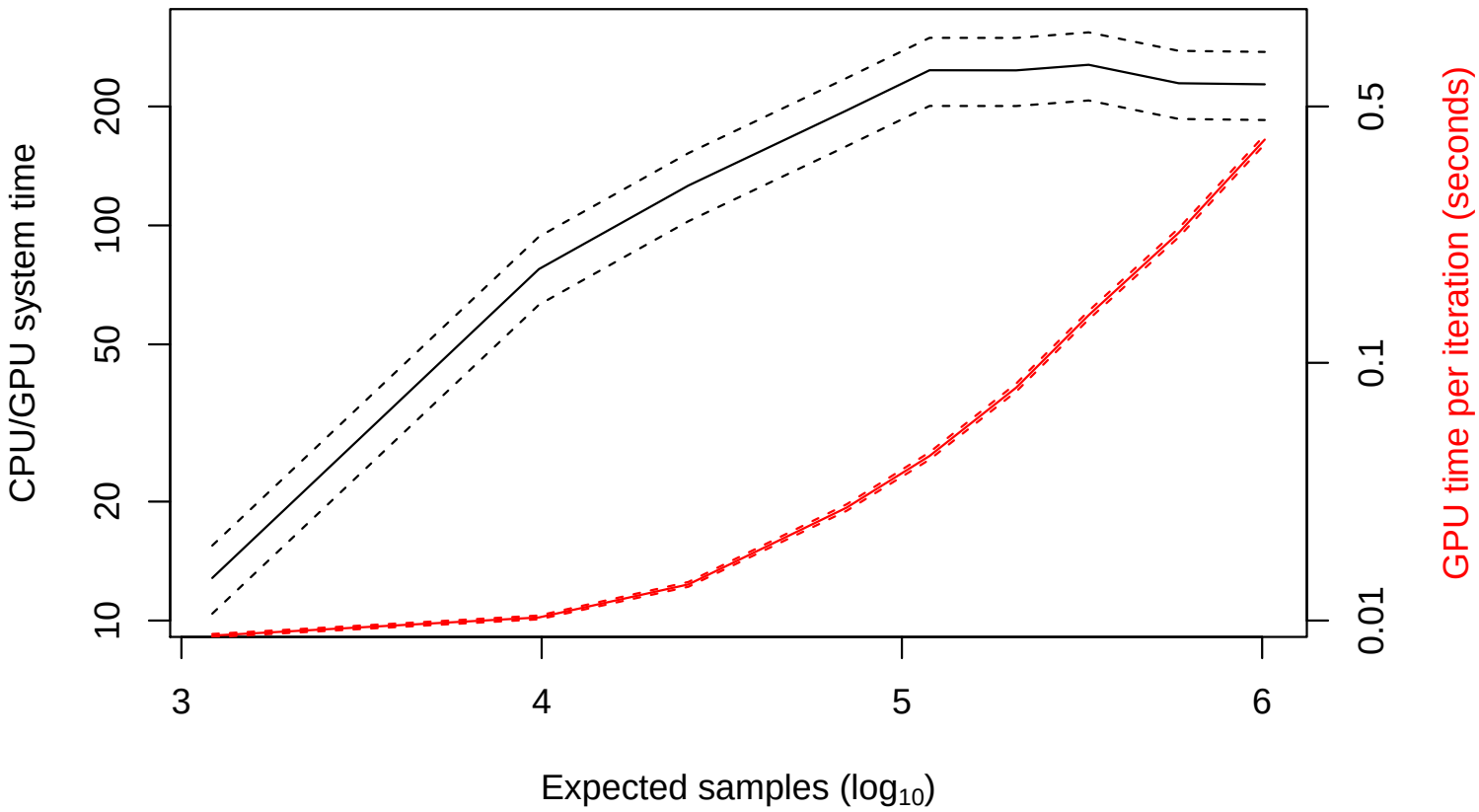
| Time | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $E$ | 120 | 71 | 76 | 81 | 80 | 90 | 90 | 104 | 103 | 109 | 109 |
| $S$ | 301 | 219 | 180 | 150 | 108 | 86 | 61 | 52 | 35 | 29 | 22 |

With $\epsilon = 0$ (i.e. draws from $p(\theta|y)$),



Since rejection sampling is inherently parallel, run this algorithm on a graphical processing unit:



# Summary

- Bayesian inference in discretely observed SCKMs
  - Goal: $p(\theta|y) \propto p(y|\theta)p(\theta)$
  - Likelihood, $L(\theta) = p(y|\theta)$, is analytically intractable
  - Sampling methods are required, e.g. rejection and/or Gibbs
  - Acceptance rate can be unacceptably low
- Approximate Bayesian computation (ABC) in SCKMs
  - Goal: $p(\theta|\rho \leq \epsilon) \propto p(\rho \leq \epsilon|\theta)p(\theta)$
  - $\rho = \rho(y, X)$ measures the difference between data and a simulation
  - $\epsilon$ balances computability with accuracy
  - Readily accommodates bounded errors, e.g. $y_t = X_t \pm \epsilon$
- ABC generally
  - More general than SKMs, e.g. phylogenetic trees
  - Building $\rho$ is an art, often use sufficient statistics of the data
  - Not useful for unbounded errors, e.g. $y_t = X_t + \epsilon_t, \epsilon_t \sim N(0, \sigma^2)$
  - Current debate about usefulness for model selection