SIMULATION

PETER ALPING

2018-05-24

TABLE OF CONTENTS

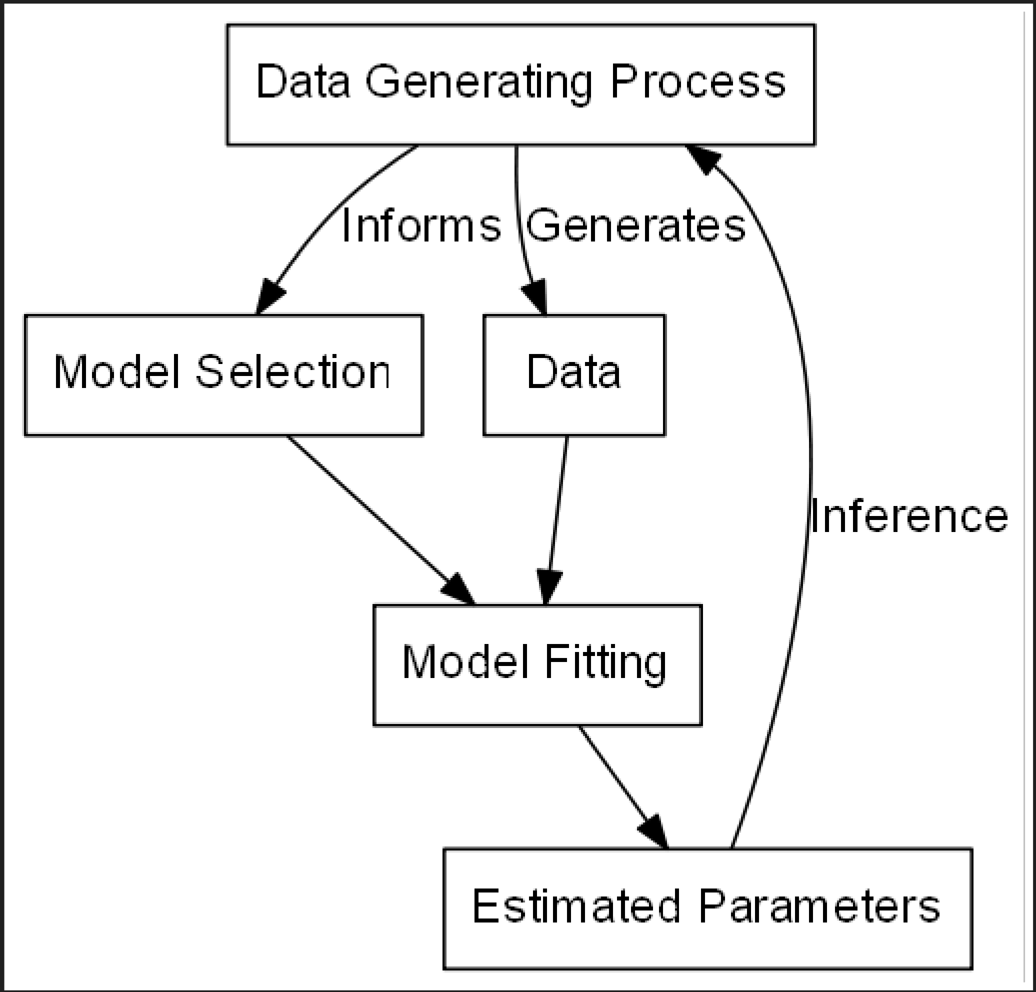
INTRODUCTION

Simulation is the imitation of the operation of a real-world process or system – Wikipedia

OUR TYPICAL SITUATION

We want to know more about a biological process for which we have measured some data

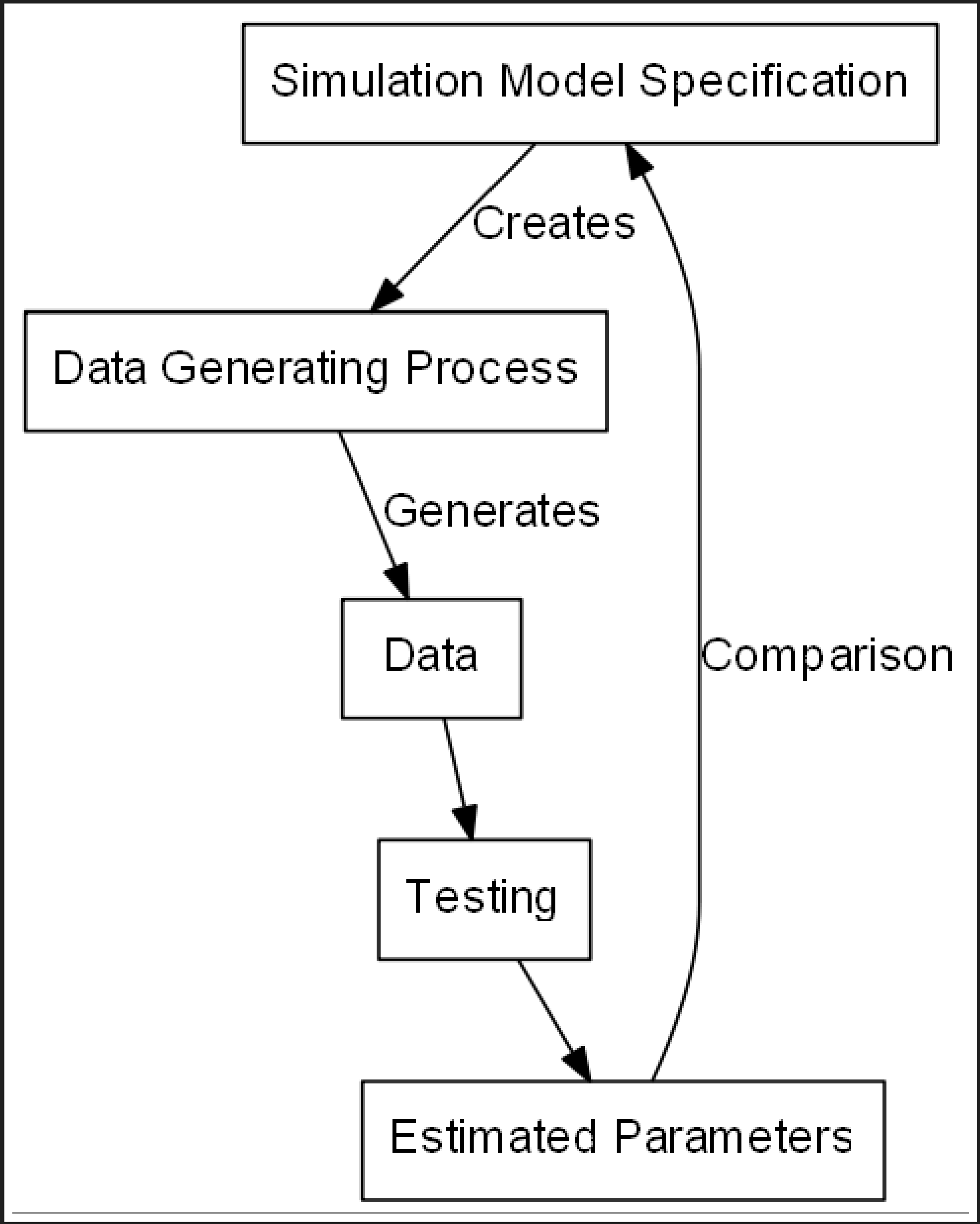
- An biological process generates data
- We measure and collect this data
- Invent model how we **think** data was generated
- Fit the model to our data
- Giving us the parameters for the model
- Inferences about the data-generating process


Data Generating Process
Informs
Generates
Model Selection
Data
Inference
Model Fitting
Estimated Parameters

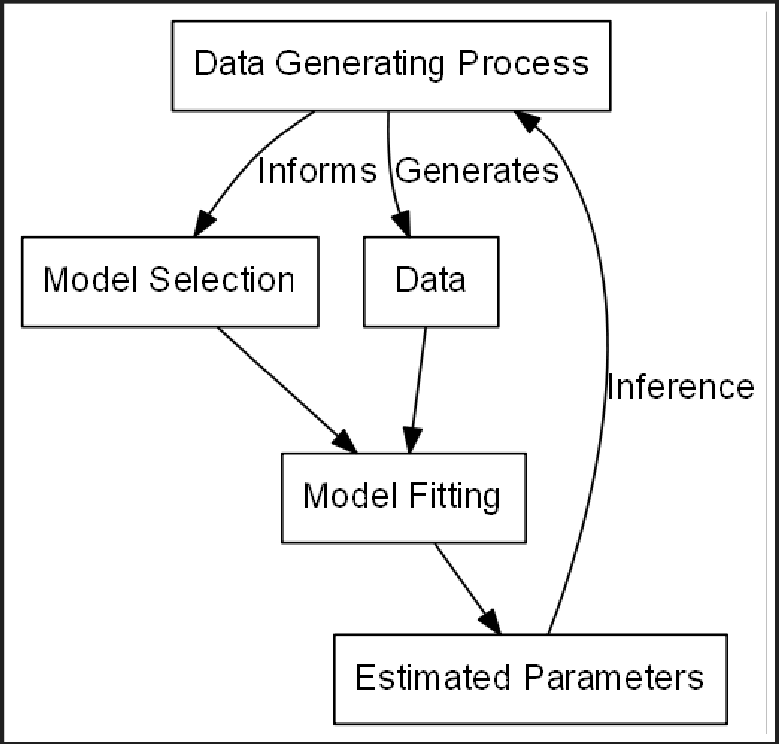
THE SIMULATION SITUATION

We want to know more about how something behaves, given data of specific type

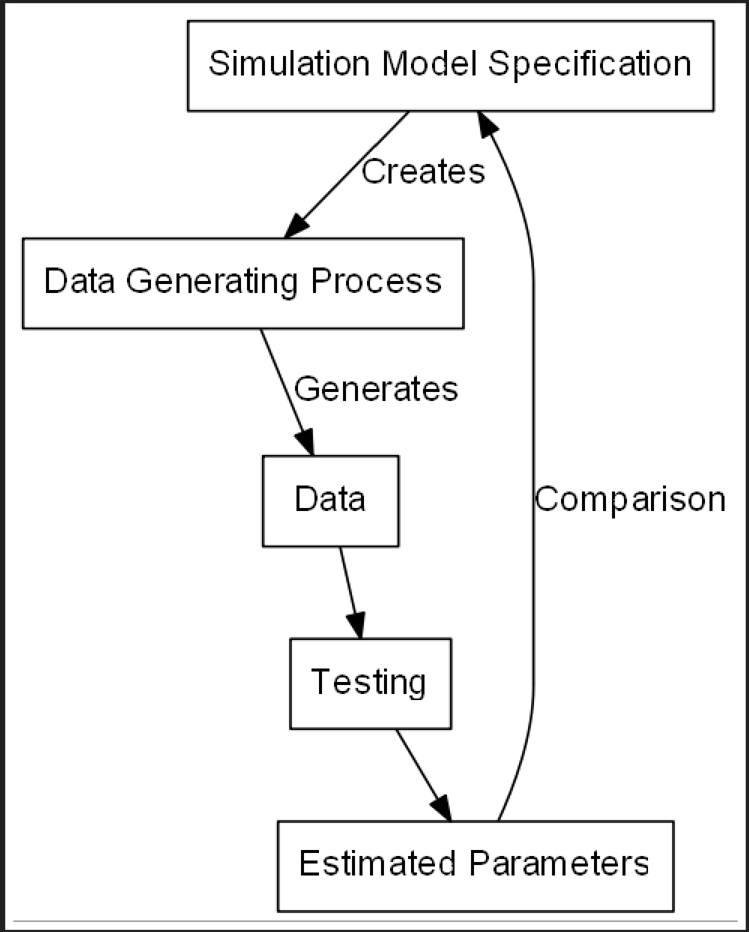
- Specify a model for the simulated data
- Including model parameters
- **Generate data using this model**
- Test something given our data
- Compare the result to our "true" model


Simulation Model Specification
Creates
Data Generating Process
Generates
Data
Testing
Estimated Parameters
Comparison

Typical situation



Simulation situation



BASICS

LINEAR REGRESSION

WHEN DO WE USE IT?

- Continuous dependent variable (y)
- Cont./discrete independent variables ($x_1, \ldots, x_p$)
- Error normally distributed

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \epsilon$
- $\epsilon \sim normal(0, \sigma)$

IN GLM NOTATION

- Dependent variable from normal distribution (y)
- Cont./discrete independent variables (x_1, …, x_p)

- $E(y) = \mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$
- $y \sim normal(\mu, \sigma)$

LOGISTIC REGRESSION

WHEN DO WE USE IT?

- Binary dependent variable (y)
- Cont./discrete independent variables (x_1, …, x_p)
- No common error distribution independent of predictor values

---

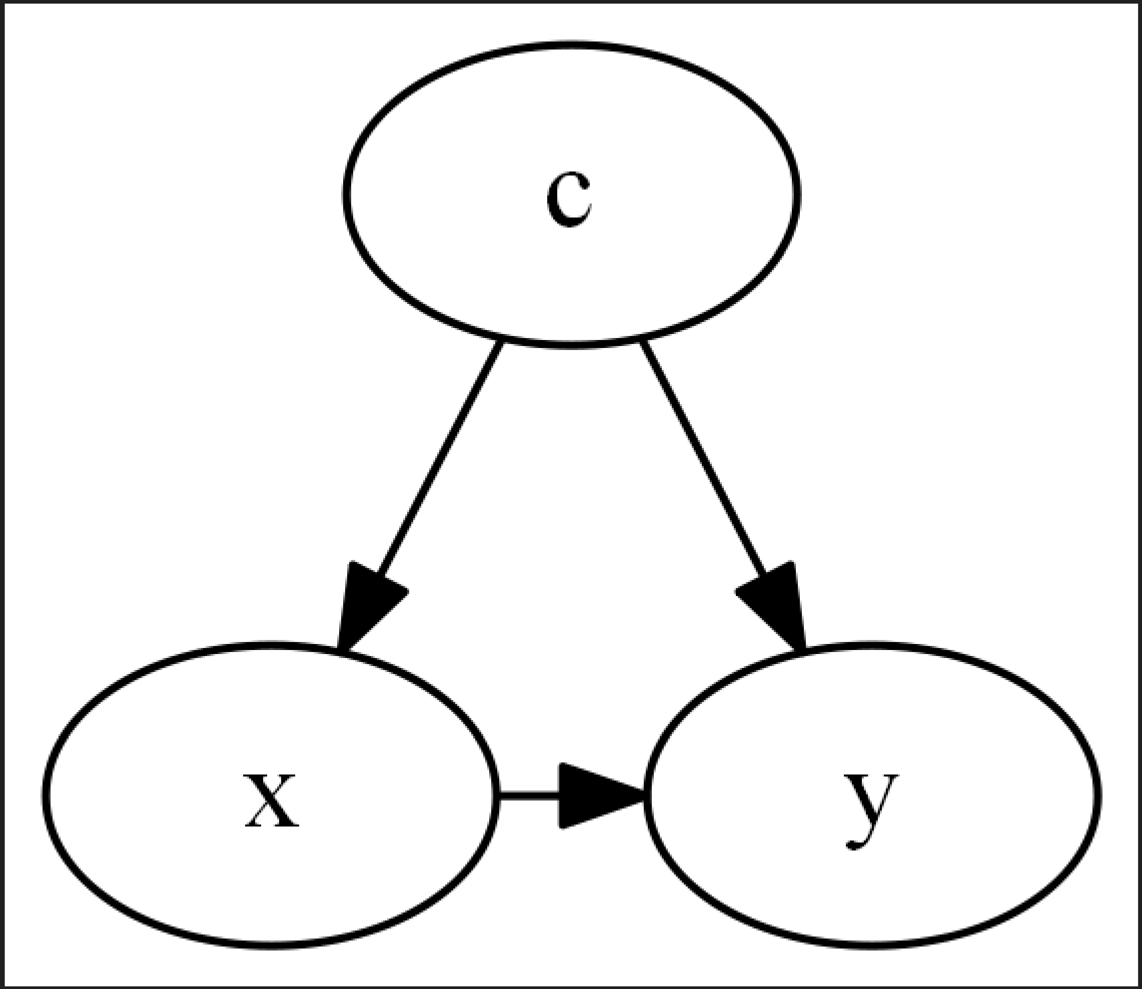
- $logit(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \epsilon$
- $logit(y) = log(\frac{y}{1-y})$
- ϵ has no independent distribution

IN GLM NOTATION

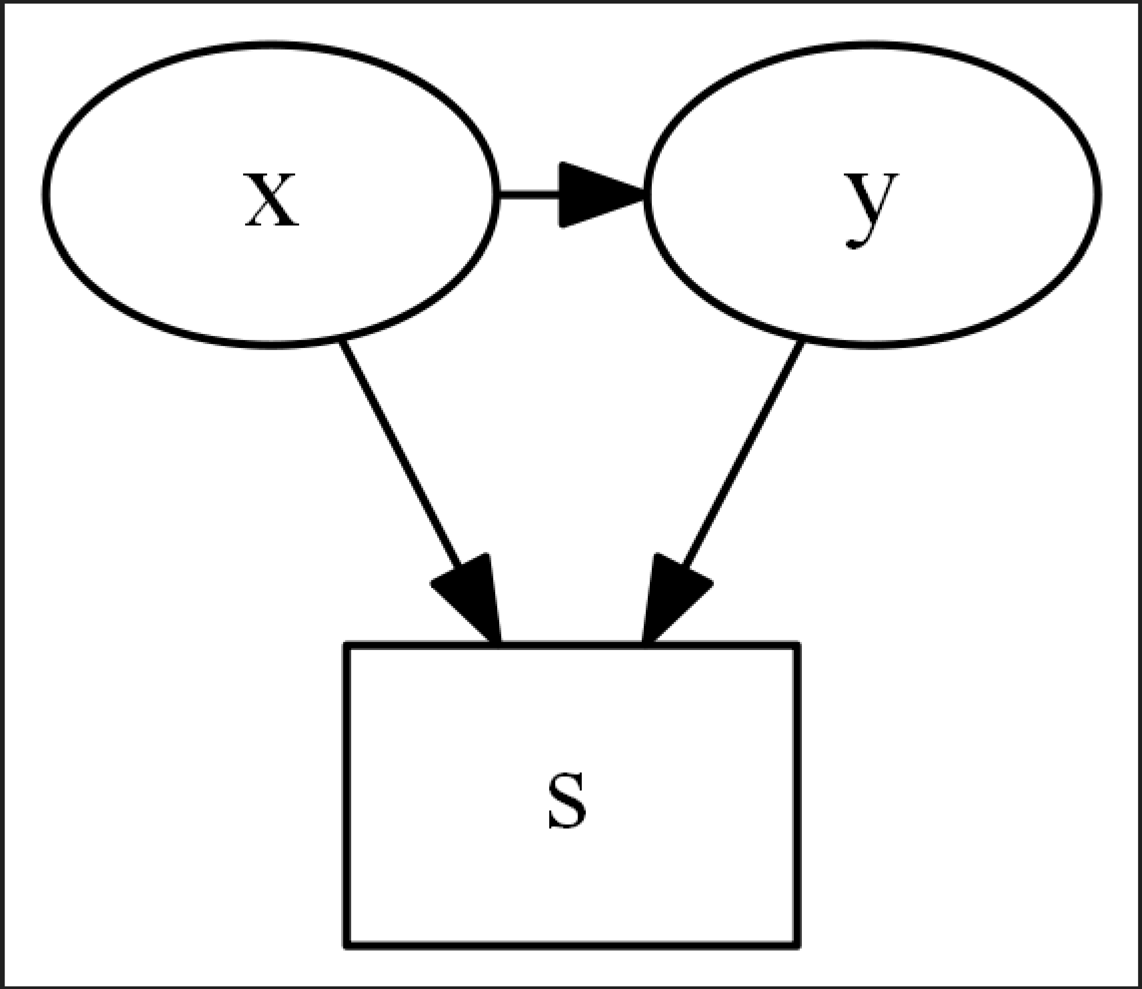
- Dep. var. from binomial distr. with 1 trial (y)
- Cont./discrete independent variables (x_1, …, x_p)

- $E(y) = \mu = logit^{-1}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p$
- $logit^{-1}(x) = \frac{e^x}{1+e^x}$
- $y \sim binomial(1, \mu)$

CONFOUNDING


c
x
y

SELECTION BIAS


x
y
s

GENERALIZED LINEAR MODELS

- $E(y) = \mu = g^{-1}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)$
- $y \sim distribution(\mu)$

- $E(y) = \mu = g^{-1}(\boldsymbol{X}\boldsymbol{\beta})$
- $y \sim distribution(\mu)$