UNIVERSITY OF BIRMINGHAM

MINI COURSE: STATISTICAL INFERENCE FOR STOCHASTIC EPIDEMIC MODELS

Dr Panayiota Touloupou
School of Mathematics
University of Birmingham

LMS Undergraduate summer school
14 - 25 July 2025
University of Bath

OVERALL LEARNING OUTCOMES

By the end of this mini-course, students will be able to:

- Understand the structure and purpose of compartmental epidemic models (e.g., SIS, SIR).
- Formulate and compute transition probabilities for disease spread and recovery.
- Simulate stochastic discrete-time epidemic models.
- Apply likelihood-based and Bayesian methods to estimate model parameters from complete epidemic data.
- Implement simulations and inference procedures in R.
- Interpret model results in the context of infectious disease dynamics.

INTRODUCTION TO EPIDEMIC MODELS

WHY DO WE MODEL EPIDEMICS?

Infectious diseases remain one of the major causes of human mortality and suffering.

Imagine a new flu virus appears:

- How many people will get sick?
- How fast will it spread?
- When will it peak?
- Will it die out on its own, or will it infect everyone?

Mathematical models help us answer these questions.

WHAT IS AN EPIDEMIC MODEL?

A mathematical model that describes how a disease spreads in a population.

Help us:

- **Understand:** How does the disease spread? What factors are most important?
- **Predict:** What might happen next? When will the peak occur?
- **Control:** What if we vaccinate people? What if people wear masks? Models help us test these "what if" scenarios before they happen in real life.
- **Inform policy:** Help public health officials make decisions (e.g., when to close schools, when to recommend masks).

Real-world examples: COVID-19, flu, measles.

COMPARTMENTAL MODELS

Population (N) is divided into compartments based on disease status:

- **S = Susceptible:** Healthy, can catch the disease.
- **I = Infectious:** Sick, can transmit the disease.
- **R = Recovered (or Removed):** Immune or removed from population (or died).
- **E = Exposed:** Infected but not yet infectious (latent period).

Each person is in one compartment at a time.

Individuals move between compartments over time.

EXAMPLES OF COMPARTMENTAL MODELS

Model	Compartments	Description
SIS	S → I → S	No immunity; reinfection possible.
SIR	S → I → R	Permanent immunity after recovery.
SEIR	S → E → I → R	Includes latent (exposed) stage.

DETERMINISTIC VS. STOCHASTIC MODELS

Feature	Deterministic	Stochastic
Output	Same every time. Single, fixed trajectory.	Random. Range of possible trajectories, showing variability.
Use	Large populations, smooth changes.	Small populations, uncertainty.
Tool	Ordinary Differential Equations (ODEs).	Probability distributions (e.g., Binomial).

DISCRETE-TIME STOCHASTIC MODELS

System **observed at fixed intervals** (e.g., daily, i.e. time evolves in steps: $t = 0, 1, 2, \dots, T-1$).

Events occur with a certain probability, called **transition probabilities**.

For example:

- **Event 1 - Infection (S → I):** A susceptible person gets infected by an infectious person.
- **Event 2 - Recovery (I → R or I → S):** An infectious person recovers. Depending on the model, they either become immune (R) or become susceptible again (S).

TRANSLATING RATES INTO PROBABILITIES

In many real-world processes, events occur continuously over time at a certain rate. For example, in our epidemic models:

- β is the **infection rate** (e.g., contacts leading to infection per susceptible per infectious person per unit time).
- γ is the **recovery rate** (e.g., recoveries per infectious person per unit time).

Rates to probabilities:

- If event rate is λ, probability of not occurring in Δt is $e^{-\lambda \Delta t}$.
- Probability of occurring is $1 - e^{-\lambda \Delta t}$.
- For discrete time, $\Delta t = 1$.

THE SIR MODEL (SUSCEPTIBLE-INFECTIOUS-RECOVERED)

Story:

- Get sick, recover, permanent immunity.

Assumptions:

- Closed population.
- Homogeneous mixing.
- Permanent immunity.

State of the system:

- At any time t, we know the number of people in each compartment: (S_t, I_t, R_t).
- $N = S_t + I_t + R_t$.

THE SIR MODEL: PROBABILITIES

Infection probability:

- Rate: β (infection rate per susceptible per infectious per unit time).
- Individual probability: $1 - e^{-\frac{\beta I_t}{N}}$.
- Number of new infections: $X_t \sim Binomial\left(S_t, 1 - e^{-\frac{\beta I_t}{N}}\right)$.

Recovery probability:

- Rate: γ (recovery rate per infectious per unit time).
- Individual probability: $1 - e^{-\gamma}$.
- Number of new recoveries: $Y_t \sim Binomial(I_t, 1 - e^{-\gamma})$.

THE SIR MODEL: TRANSITIONS AND UPDATES

Summary:

- $X_t \sim Binomial\left(S_t, 1 - e^{-\frac{\beta I_t}{N}}\right)$ (new infections),
- $Y_t \sim Binomial(I_t, 1 - e^{-\gamma})$ (new recoveries).

Transition from (S_t, I_t, R_t) to $(S_{t+1}, I_{t+1}, R_{t+1})$:

- $S_{t+1} = S_t - X_t,$
- $I_{t+1} = I_t + X_t - Y_t,$
- $R_{t+1} = R_t + Y_t.$

SIMULATING SIR STOCHASTIC EPIDEMICS

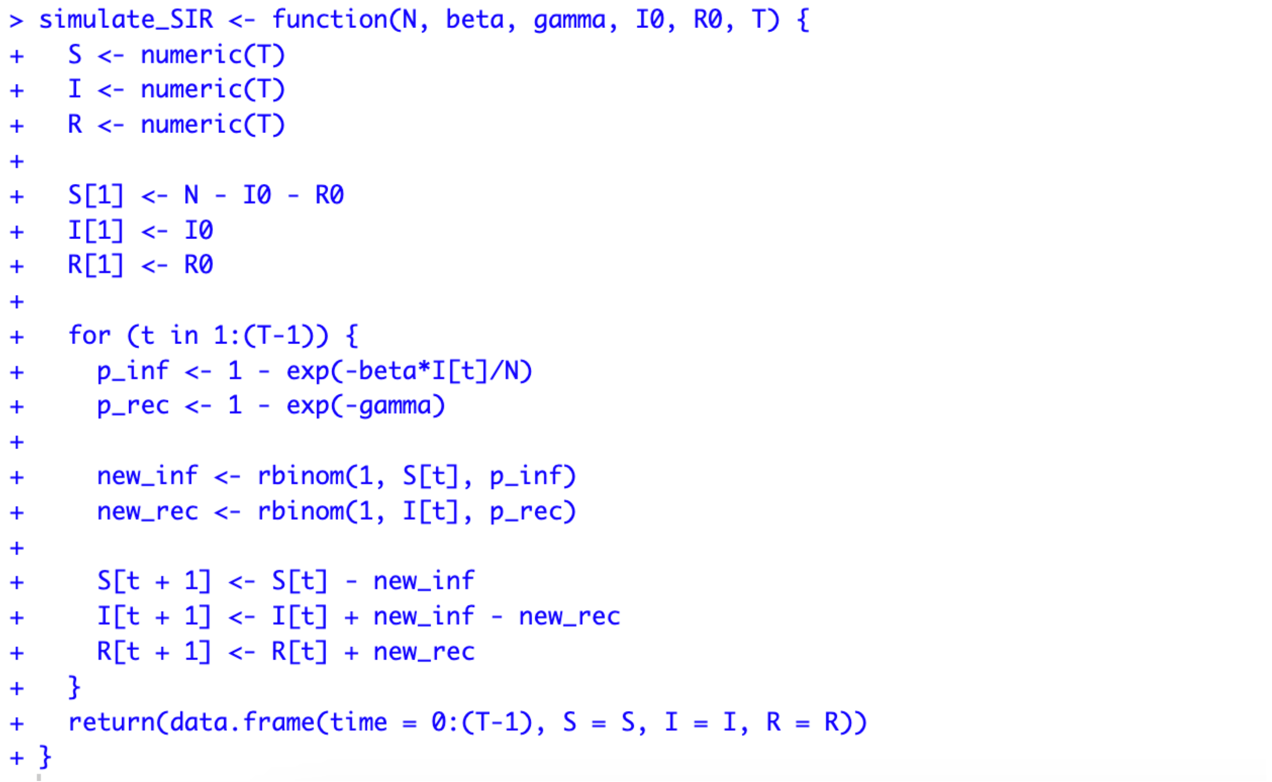
Simulation algorithm (general steps):

1. Initialize: Set initial compartment counts and parameters.

2. Loop through time steps ($t = 1, 2, \dots, T - 1$):
   - Calculate probabilities:
     - Use current counts to get probability of infection and probability of recovery.
   - Draw random events:
     - Draw X_t and Y_t from Binomial distributions.
   - Update compartments:
     - Calculate S_{t+1}, I_{t+1} and R_{t+1} based on drawn events.
   - Store results:
     - Save counts for each step.

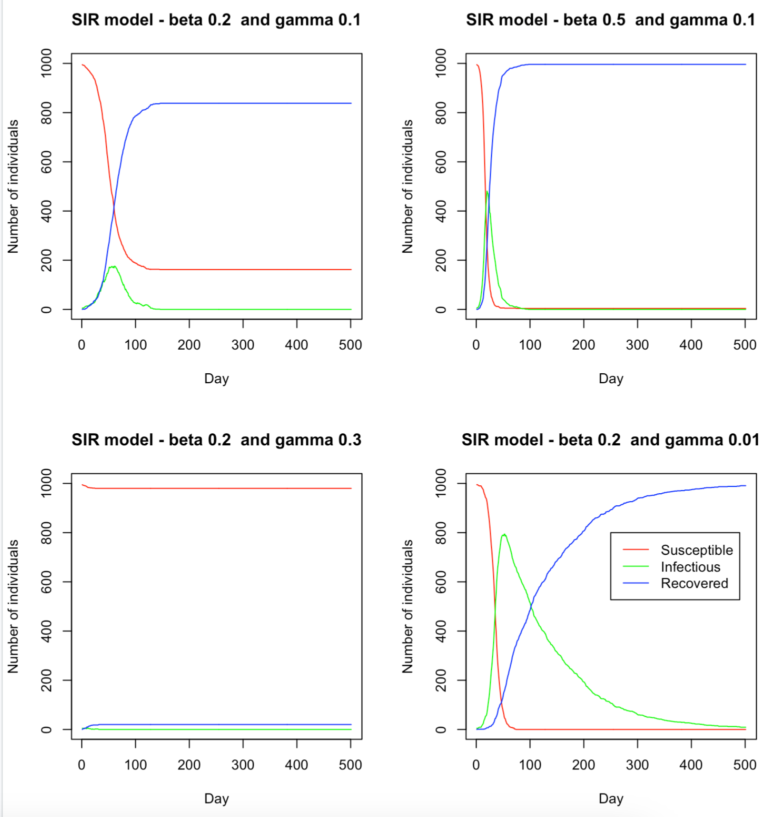
3. Repeat: Continue the loop until the end of your simulation period.

R CODE: SIMULATING SIR STOCHASTIC EPIDEMICS



```
> simulate_SIR <- function(N, beta, gamma, I0, R0, T) {
+   S <- numeric(T)
+   I <- numeric(T)
+   R <- numeric(T)
+
+   S[1] <- N - I0 - R0
+   I[1] <- I0
+   R[1] <- R0
+
+   for (t in 1:(T-1)) {
+     p_inf <- 1 - exp(-beta*I[t]/N)
+     p_rec <- 1 - exp(-gamma)
+
+     new_inf <- rbinom(1, S[t], p_inf)
+     new_rec <- rbinom(1, I[t], p_rec)
+
+     S[t + 1] <- S[t] - new_inf
+     I[t + 1] <- I[t] + new_inf - new_rec
+     R[t + 1] <- R[t] + new_rec
+   }
+   return(data.frame(time = 0:(T-1), S = S, I = I, R = R))
+ }
```

SIR SIMULATIONS: INTERPRETING THE OUTPUT



- Peak: when I_t is highest.
- Final size: total number of people infected.
- Epidemic dies out.
- Effect of parameters:
  - Higher β: faster spread.
  - Higher γ: faster recovery.

THE SIS MODEL (SUSCEPTIBLE-INFECTIOUS-SUSCEPTIBLE)

Story:

- Get sick, recover, can get sick again (no permanent immunity).

Assumptions:

- Closed population.
- Homogeneous mixing.
- No permanent immunity.

State of the system:

- At any time t, we know the number of people in each compartment: (S_t, I_t).
- $N = S_t + I_t$.

THE SIS MODEL: PROBABILITIES

Infection probability: Same as SIR.

- Number of new infections: $X_t \sim Binomial\left(S_t, 1 - e^{-\frac{\beta I_t}{N}}\right)$.

Recovery probability: Same as SIR but instead of moving to R, recovered individuals go back to S.

- Number of new recoveries: $Y_t \sim Binomial(I_t, 1 - e^{-\gamma})$.

Transition from (S_t, I_t) to (S_{t+1}, I_{t+1}):

- $S_{t+1} = ?$
- $I_{t+1} = ?$

THE SIS MODEL: PROBABILITIES

Infection probability: Same as SIR.

- Number of new infections: $X_t \sim Binomial\left(S_t, 1 - e^{-\frac{\beta I_t}{N}}\right)$.

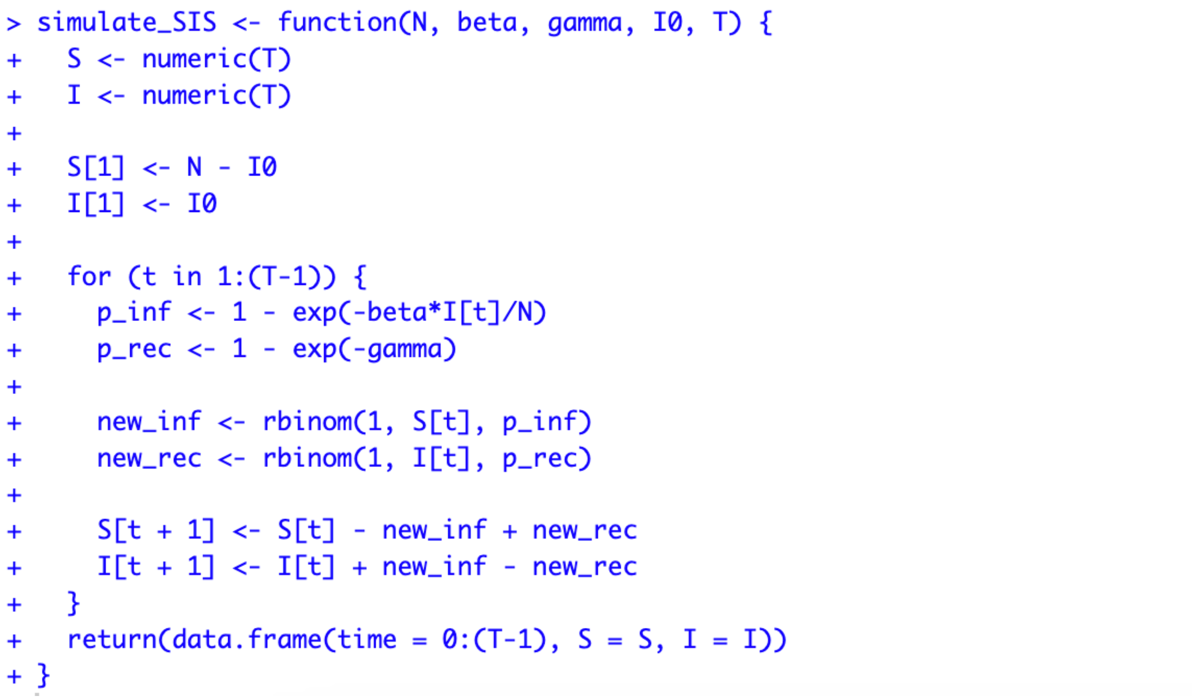
Recovery probability: Same as SIR but instead of moving to R, recovered individuals go back to S.

- Number of new recoveries: $Y_t \sim Binomial(I_t, 1 - e^{-\gamma})$.

Transition from (S_t, I_t) to (S_{t+1}, I_{t+1}):

- $S_{t+1} = S_t - X_t + Y_t$,
- $I_{t+1} = I_t + X_t - Y_t$.

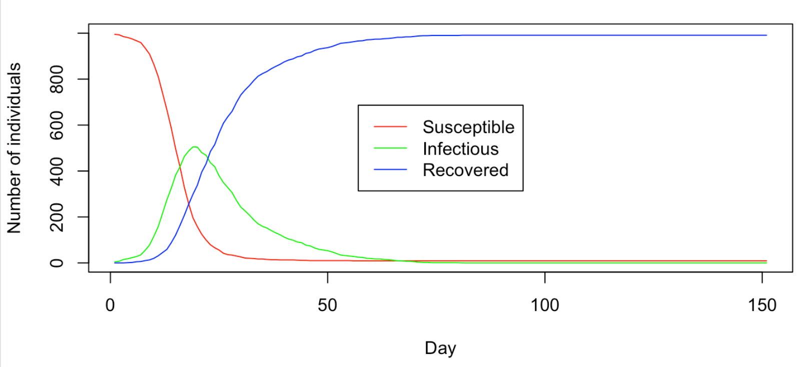
R CODE: SIMULATING SIS STOCHASTIC EPIDEMICS



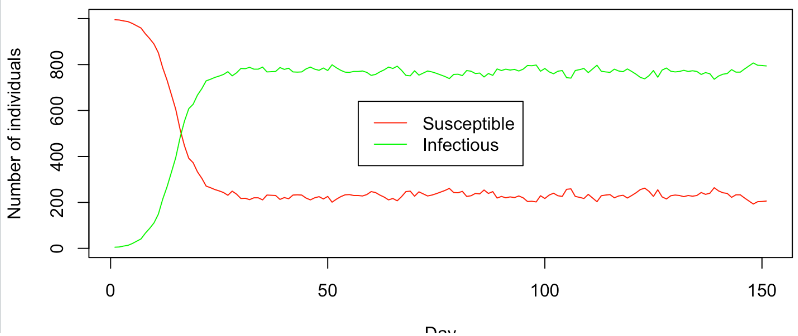
```
> simulate_SIS <- function(N, beta, gamma, I0, T) {
+   S <- numeric(T)
+   I <- numeric(T)
+
+   S[1] <- N - I0
+   I[1] <- I0
+
+   for (t in 1:(T-1)) {
+     p_inf <- 1 - exp(-beta*I[t]/N)
+     p_rec <- 1 - exp(-gamma)
+
+     new_inf <- rbinom(1, S[t], p_inf)
+     new_rec <- rbinom(1, I[t], p_rec)
+
+     S[t + 1] <- S[t] - new_inf + new_rec
+     I[t + 1] <- I[t] + new_inf - new_rec
+   }
+   return(data.frame(time = 0:(T-1), S = S, I = I))
+ }
```

SIR VS. SIS MODELS: A COMPARISON FROM SIMULATIONS

SIR model - beta 0.5 and gamma 0.1



SIS model - beta 0.5 and gamma 0.1



Model structure **affects long-term outcomes**:

- SIR: infection rises, peaks, then dies out.
- SIS: infection may stabilize at a non-zero level, indicating the disease becomes endemic (persists indefinitely) in the population.

STATISTICAL INFERENCE FOR STOCHASTIC EPIDEMIC MODELS

WHAT IS STATISTICAL INFERENCE

We have real-world data from an epidemic. We also have mathematical models (like SIR, SIS) with unknown parameters (β, γ).

The goal of inference is to use the observed data to learn about unknown parameters.

- In our case: estimate β and γ from observed epidemic data.

Two main approaches:

1. Frequentist (e.g., Maximum Likelihood Estimation - MLE).
2. Bayesian (e.g., Posterior distributions).

MAXIMUM LIKELIHOOD ESTIMATION

MAXIMUM LIKELIHOOD PARAMETER ESTIMATION

Maximum likelihood estimation: a method that determines values for the parameters of a model, given some observed data.

The parameter values are found such that they maximise the likelihood of the model or equivalent the log likelihood of the model.

Differentiate the log likelihood with respect to the parameters and set the derivatives to zero.

We use the log likelihood because:

- It turns products into sums.
- It's easier to compute and optimize.

THE LIKELIHOOD FUNCTION

It is a function of parameters, given the data denoted by:

$$L(\theta|Data) = P(Data|\theta).$$

- In our case, the parameters θ are: β (infection rate) and γ (recovery rate).

It quantifies how well a given set of parameters explains the observed data. Higher likelihood means better fit.

Find the parameter values that maximize this function, or equivalently that maximize the log-likelihood, which is the natural logarithm of the likelihood:

$$\log\big(L(\theta|Data)\big).$$

RECAP – THE STOCHASTIC SIR MODEL

Observed epidemic data:

- S_t, I_t and R_t, for $t = 0, 1, 2, \dots, T - 1$.

From this, we can compute:

- $X_t = S_t - S_{t+1}$ (new infections).
- $Y_t = R_{t+1} - R_t$ (new recoveries).

From the model, we assume:

- $X_t \sim Binomial\left(S_t, 1 - e^{-\frac{\beta I_t}{N}}\right)$ (new infections).
- $Y_t \sim Binomial(I_t, 1 - e^{-\gamma})$ (new recoveries).

CONSTRUCTING THE LOG-LIKELIHOOD FOR SIR MODEL

Likelihood contribution for a single time step (t to $t + 1$):

For a given time step, if we start with $\mathrm{S_t}$ susceptibles and $\mathrm{I_t}$ infectious individuals, and we observe X_t new infections and Y_t new recoveries, the likelihood is given by:

$$L_t(\beta, \gamma \mid \mathrm{S_t}, \mathrm{R_t}, \mathrm{I_t}) = P(X_t \mid \mathrm{S_t}, \mathrm{I_t}, \beta)\, P(Y_t \mid \mathrm{I_t}, \gamma)$$
$$= \binom{\mathrm{S_t}}{X_t} (p_{inf})^{X_t} \, (1 - p_{inf})^{\mathrm{S_t} - X_t} \binom{\mathrm{I_t}}{Y_t} (p_{rec})^{Y_t} \, (1 - p_{rec})^{\mathrm{I_t} - Y_t},$$

where

- $p_{inf} = 1 - e^{-\frac{\beta\, I_t}{N}},$
- $p_{rec} = 1 - e^{-\gamma}.$

CONSTRUCTING THE LOG-LIKELIHOOD FOR SIR MODEL

Total likelihood for the entire epidemic:

If we have observed data over T time steps (from $\mathrm{t} = 0$ to $T - 1$), the total likelihood is the product of the probabilities for each individual step:

$$L\,(\beta, \gamma \mid \mathrm{S_t},\, \mathrm{R_t}, \mathrm{I_t}) = \prod_{t=0}^{T-2} L_t(\beta, \gamma \mid \mathrm{S_t},\, \mathrm{R_t}, \mathrm{I_t}) \text{ [Markov property],}$$

and equivalently the total log-likelihood is given by:

$$\log(L\,(\beta, \gamma \mid \mathrm{S_t},\, \mathrm{R_t}, \mathrm{I_t})) = \sum_{t=0}^{T-2} \log(L_t(\beta, \gamma \mid \mathrm{S_t},\, \mathrm{R_t}, \mathrm{I_t})).$$

PARAMETER ESTIMATION

MLE provides point estimates – single "best guess" values for each parameter.

How would you estimate β and γ from data?

$$\frac{d\ \log(L\ (\beta,\gamma|\mathrm{S_t,\ R_t,I_t}))}{d\beta} = 0 \text{ and } \frac{d\ \log(L\ (\beta,\gamma|\mathrm{S_t,\ R_t,I_t}))}{d\gamma} = 0.$$

Properties of log [= ln here]

1. $\ln(x * y)$
$= \ln(x) + \ln(y)$

2. $\ln\ (x^y)$
$= y \ln(x)$

3. $\ln\ (e^x)$
$= x \ln(e) = x$

PARAMETER ESTIMATION

MLE provides point estimates – single "best guess" values for each parameter.

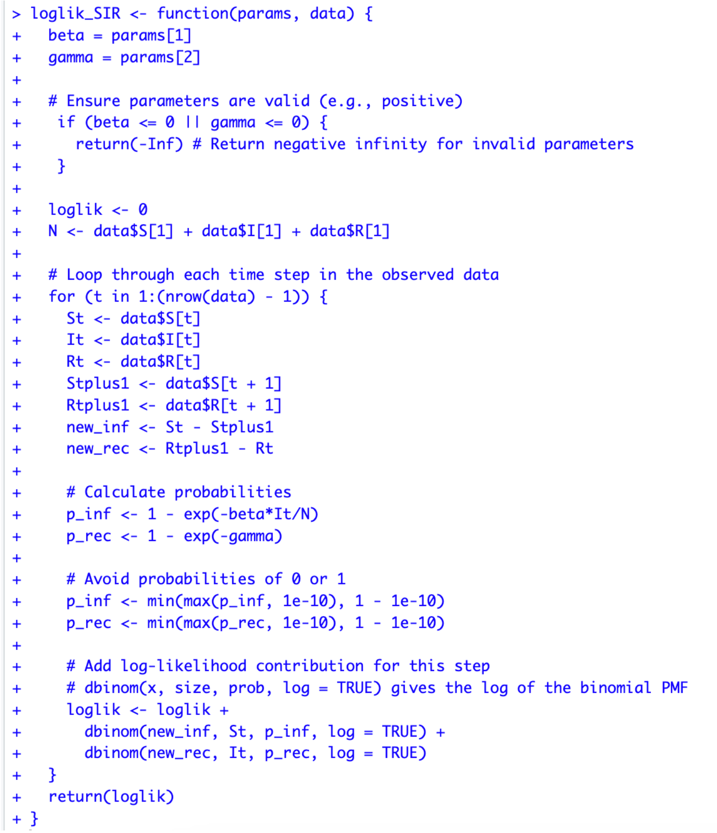
How would you estimate β and γ from data?

$$\frac{d\,\log(L\,(\beta,\gamma|\mathrm{S_t},\ \mathrm{R_t},\mathrm{I_t}))}{d\beta} = 0 \text{ and } \frac{d\,\log(L\,(\beta,\gamma|\mathrm{S_t},\ \mathrm{R_t},\mathrm{I_t}))}{d\gamma} = 0.$$

For most stochastic epidemic models, we cannot find the MLEs by simply taking derivatives of the log-likelihood and setting them to zero (analytical solution).

Instead we use computer algorithms to search for the maximum value of the log-likelihood function. For example, we use optim() in R.

R CODE: LOG-LIKELIHOOD FUNCTION



```
> loglik_SIR <- function(params, data) {
+   beta = params[1]
+   gamma = params[2]
+
+   # Ensure parameters are valid (e.g., positive)
+    if (beta <= 0 || gamma <= 0) {
+      return(-Inf) # Return negative infinity for invalid parameters
+    }
+
+   loglik <- 0
+   N <- data$S[1] + data$I[1] + data$R[1]
+
+   # Loop through each time step in the observed data
+   for (t in 1:(nrow(data) - 1)) {
+     St <- data$S[t]
+     It <- data$I[t]
+     Rt <- data$R[t]
+     Stplus1 <- data$S[t + 1]
+     Rtplus1 <- data$R[t + 1]
+     new_inf <- St - Stplus1
+     new_rec <- Rtplus1 - Rt
+
+     # Calculate probabilities
+     p_inf <- 1 - exp(-beta*It/N)
+     p_rec <- 1 - exp(-gamma)
+
+     # Avoid probabilities of 0 or 1
+     p_inf <- min(max(p_inf, 1e-10), 1 - 1e-10)
+     p_rec <- min(max(p_rec, 1e-10), 1 - 1e-10)
+
+     # Add log-likelihood contribution for this step
+     # dbinom(x, size, prob, log = TRUE) gives the log of the binomial PMF
+     loglik <- loglik +
+       dbinom(new_inf, St, p_inf, log = TRUE) +
+       dbinom(new_rec, It, p_rec, log = TRUE)
+   }
+   return(loglik)
+ }
```

Recap:

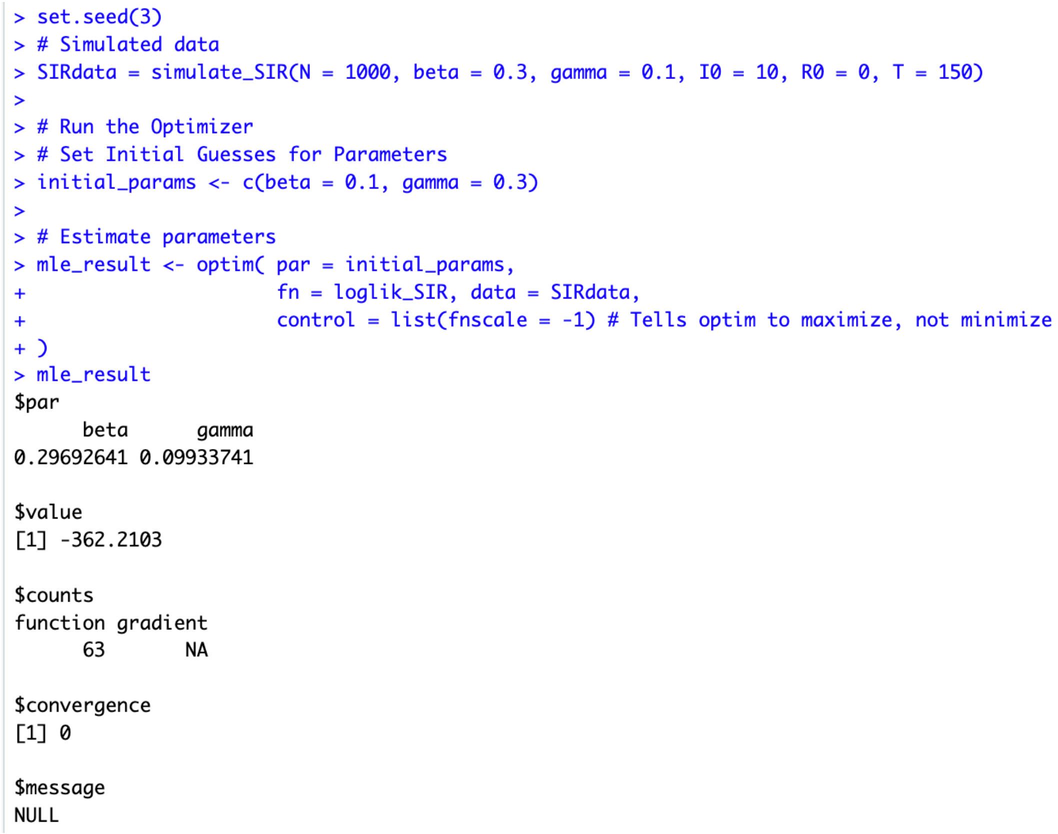
- Observed epidemic data:
  - S_t, I_t and R_t, for $t = 0, 1, 2, \dots, T - 1$.
- From this, we can compute:
  - $X_t = S_t - S_{t+1}$ (new infections).
  - $Y_t = R_{t+1} - R_t$ (new recoveries).
- Likelihood for each step:
  - $L_t(\beta, \gamma \mid \mathrm{S_t}, \mathrm{R_t}, \mathrm{I_t}) = P(X_t \mid \mathrm{S_t}, \mathrm{I_t}, \beta)\; P(Y_t \mid \mathrm{I_t}, \gamma)$

$$= \binom{\mathrm{S_t}}{X_t} (p_{inf})^{X_t} \, (1 - p_{inf})^{\mathrm{S_t} - X_t} \binom{\mathrm{I_t}}{Y_t} (p_{rec})^{Y_t} \, (1 - p_{rec})^{\mathrm{I_t} - Y_t},$$

where

- $p_{inf} = 1 - e^{-\frac{\beta I_t}{N}}$,
- $p_{rec} = 1 - e^{-\gamma}$.

IMPLEMENTING MLE IN R



```
> set.seed(3)
> # Simulated data
> SIRdata = simulate_SIR(N = 1000, beta = 0.3, gamma = 0.1, I0 = 10, R0 = 0, T = 150)
>
> # Run the Optimizer
> # Set Initial Guesses for Parameters
> initial_params <- c(beta = 0.1, gamma = 0.3)
>
> # Estimate parameters
> mle_result <- optim( par = initial_params,
+                      fn = loglik_SIR, data = SIRdata,
+                      control = list(fnscale = -1) # Tells optim to maximize, not minimize
+ )
> mle_result
$par
      beta      gamma
0.29692641 0.09933741

$value
[1] -362.2103

$counts
function gradient
      63       NA

$convergence
[1] 0

$message
NULL
```

QUANTIFYING UNCERTAINTY WITH MLE

While MLE gives point estimates, we also need to know how certain we are about these estimates.

- Standard errors:
  - These measure the precision of our estimates. Smaller standard errors mean more precise estimates.
- Confidence intervals (CIs):
  - Based on the standard errors and the assumption that the parameter estimates are approximately normally distributed (especially for large datasets).
  - A 95% confidence interval for beta means that if we were to repeat the data collection and estimation process many times, 95% of these intervals would contain the true value of beta.

BAYESIAN INFERENCE

WHAT IS BAYESIAN INFERENCE?

Frequentist view (MLE): Find the value that makes the data most likely.

- Gives a point estimate.

Bayesian view: Treat the unknown parameters as random variables.

- Gives a distribution over possible values.

The Bayesian philosophy: Updating beliefs with data.

- We start with an initial belief about the parameters (our prior).
- We update this belief using the observed data (via the likelihood),
- To get a refined belief (our posterior).

BAYES' THEOREM: THE FOUNDATION

The mathematical rule for this update is **Bayes' Theorem**:

$$p(\theta|Data) = \frac{p(Data|\theta)\,p(\theta)}{p(Data)},$$

where:

- θ: unknown parameters,
- $p(\theta)$: prior – your belief before seeing the data (based on previous studies or expert opinion),
- $p(Data|\theta)$: likelihood,
- $p(\theta|Data)$: posterior – updated belief,
- $p(Data)$: evidence or normalizing constant (doesn’t depend on θ).

BAYES' THEOREM: THE FOUNDATION

The mathematical rule for this update is **Bayes' Theorem**:

$$p(\theta|Data) = \frac{p(Data|\theta)\, p(\theta)}{p(Data)}.$$

For parameter estimation, we don't need to calculate $p(Data)$ directly because it doesn't depend on θ. This means:

$$p(\theta|Data) \propto p(Data|\theta)\, p(\theta),$$

(the posterior is proportional to the likelihood times the prior).

WHY USE BAYESIAN INFERENCE?

- Gives a distribution over parameters, not just point estimates.
- Naturally incorporates uncertainty.
- Can include prior knowledge (e.g., from past outbreaks).
  - Incorporate existing knowledge: If we know from previous research that a recovery rate for a particular disease is typically around 0.1, we can encode that.
- Especially useful when data is limited or noisy.

EXAMPLE – ESTIMATING A PROPORTION

Coin flip example (or infections):

- Assume you observe $X \sim Binomial(n, \theta)$ (e.g. 3 infections out of 10 susceptible individuals).
- $\theta \in [0, 1]$ is the unknown probability of success (infection).
- Bayesian approach:
  - Prior: $\theta \sim Beta(a, b)$, i.e. $\mathrm{p}(\theta) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$.
  - Posterior: ?

EXAMPLE – ESTIMATING A PROPORTION

Coin flip example (or infections):

- Assume you observe $X \sim Binomial(n, \theta)$ (e.g. 3 infections out of 10 susceptible individuals).
- $\theta \in [0, 1]$ is the unknown probability of success (infection).
- Bayesian approach:
  - Prior: $\theta \sim Beta(a, b)$.
  - Posterior: $\theta | X \sim Beta(a + X, b + n - X)$ [conjugate prior].

THE CHALLENGE OF THE POSTERIOR AND MCMC METHOD

Calculating the posterior distribution $p(\theta|Data)$ directly is often mathematically impossible for complex models like our stochastic epidemic models.

Since we can't calculate the posterior directly, we resort to a clever computational trick: sampling from it.

Most common sampling method: Markov Chain Monte Carlo (MCMC)

- Instead of finding the exact mathematical form of the posterior distribution, we generate a large number of samples (parameter values) that are drawn from that posterior distribution.
- If we have enough samples, we can then approximate the shape of the posterior and calculate its properties (mean, median, credible intervals).

ANALOGY: THE HIKER ON THE MOUNTAIN

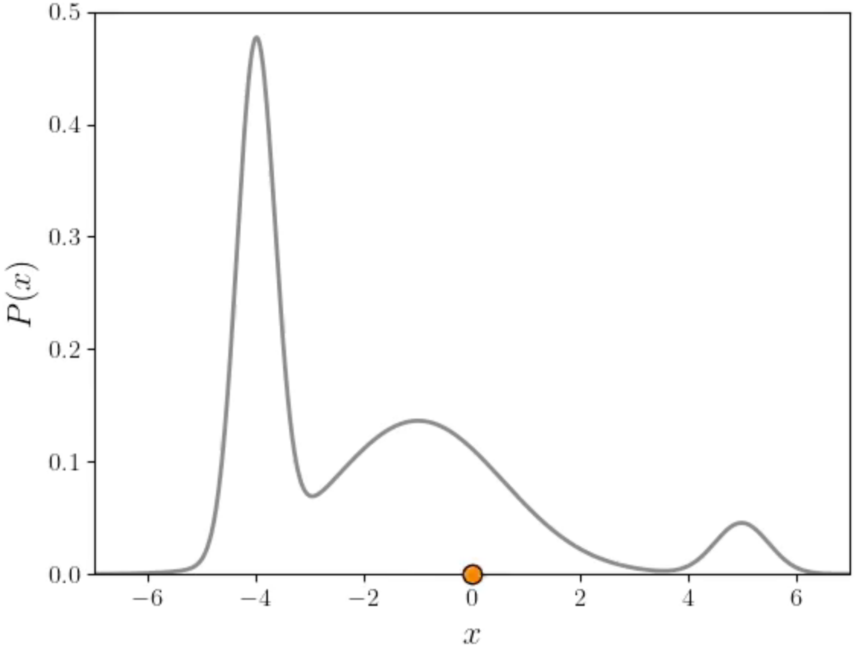
Imagine the posterior distribution as a complex mountain range. We want to map its shape, especially its peaks (high probability regions).

MCMC is like sending out a hiker (a "walker" in the parameter space):

- The hiker takes many, many steps.
- The rules for how the hiker moves ensure that they spend more time in higher elevations (regions of higher posterior probability) and less time in lower elevations.
- After a very long walk, the trail left by the hiker (the sequence of visited locations/parameter values) will effectively trace out the shape of the mountain range.
- The "Markov" part means that the hiker's next step depends only on their current position, not on their entire past history.

ANIMATION

HTTPS://WWW.ALGORITHM-ARCHIVE.ORG/CONTENTS/METROPOLIS/METROPOLIS.HTML



METROPOLIS-HASTINGS (MH) ALGORITHM WITH NORMAL PROPOSALS

1. Initialise: Choose starting value $\theta^{(0)}$.
2. For $i = 1, 2, \dots, I$:
   - Propose $\theta^{*} \sim N(\theta^{(i-1)}, \sigma^2)$ [Add a small noise to the current position] [or $\theta^{*} \sim N(\theta^{(i-1)}, \Sigma)$]
   - Compute acceptance ratio:

$$\alpha = \min\left(1, \frac{p(\theta^{*}|Data)}{p(\theta^{(i-1)}|Data)}\right) = \min\left(1, \frac{L(\theta^{*}|Data)\, p(\theta^{*})}{L(\theta^{(i-1)}|Data)\, p(\theta^{(i-1)})}\right).$$

   - Accept/reject:
     - With probability α: set $\theta^{(i)} = \theta^{*}$ [accept].
     - Otherwise: $\theta^{(i)} = \theta^{(i-1)}$ [reject].
3. Return $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(I)}$.

WHAT DOES "ACCEPT WITH PROBABILITY α" MEAN IN PRACTICE?

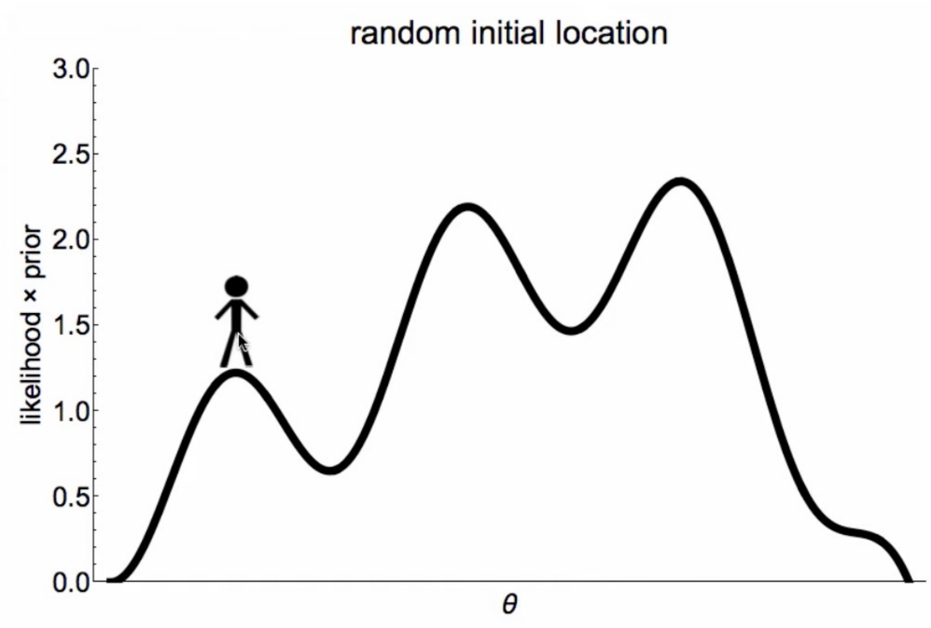
It means you generate a random number from a $Uniform(0,1)$ distribution, say:

$$u \sim Uniform(0,1).$$

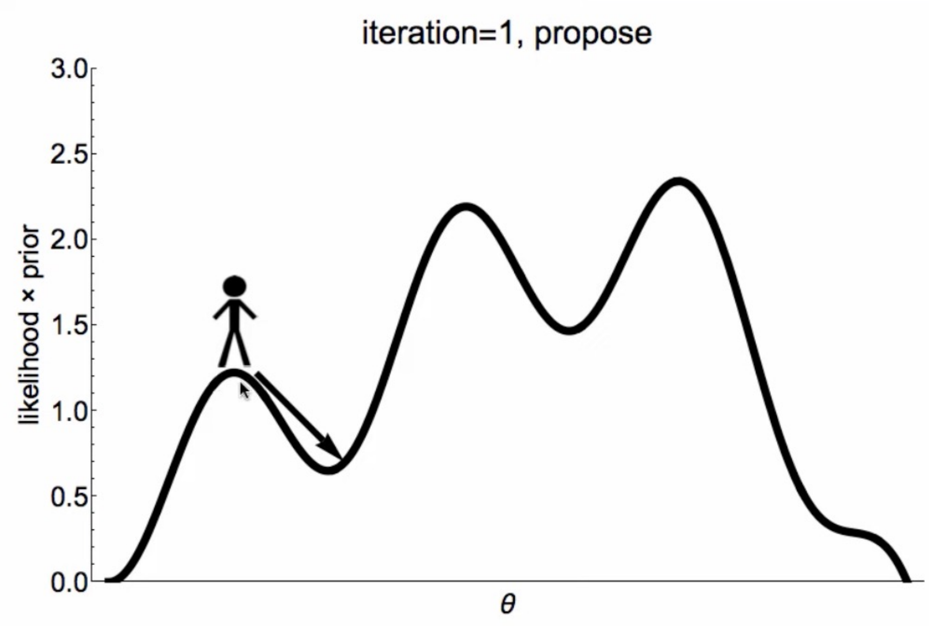
And then:

- If $u < \alpha$, accept the proposal, i.e. set $\theta^{(i)} = \theta^*$.
- If $u \geq \alpha$, reject the proposal, i.e. set $\theta^{(i)} = \theta^{(i-1)}$.

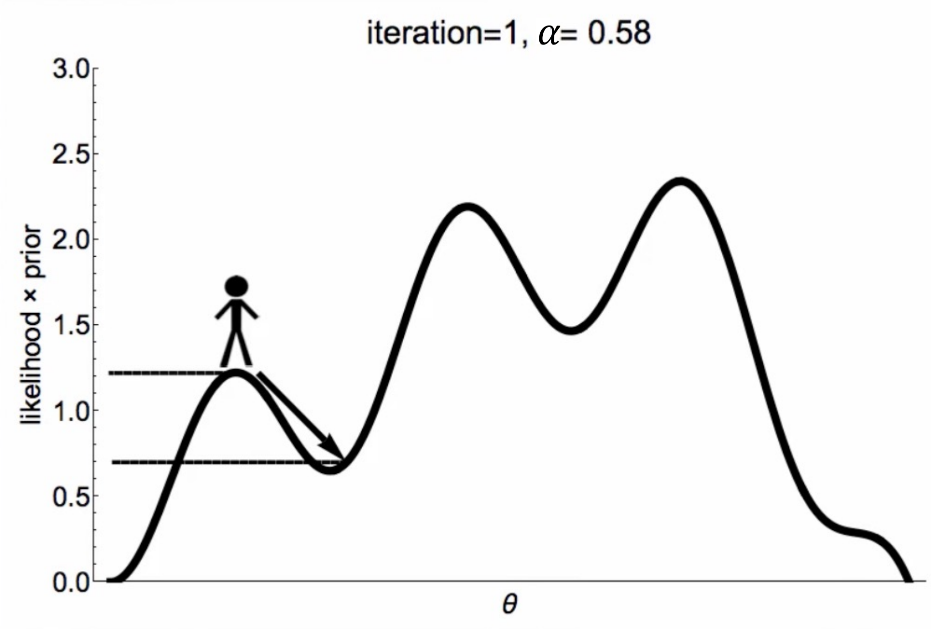
ANIMATION MH ALGORITHM



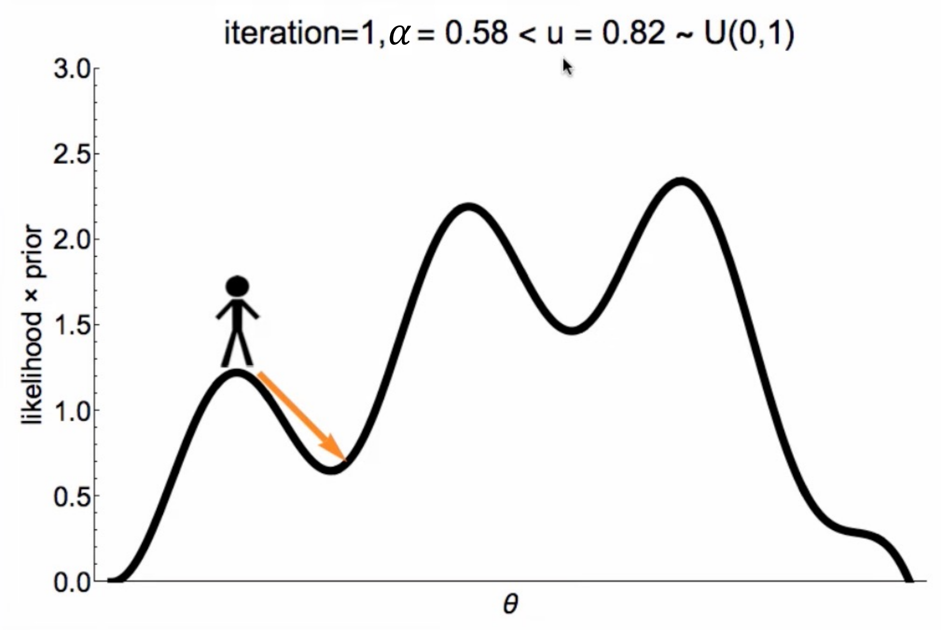
ANIMATION MH ALGORITHM



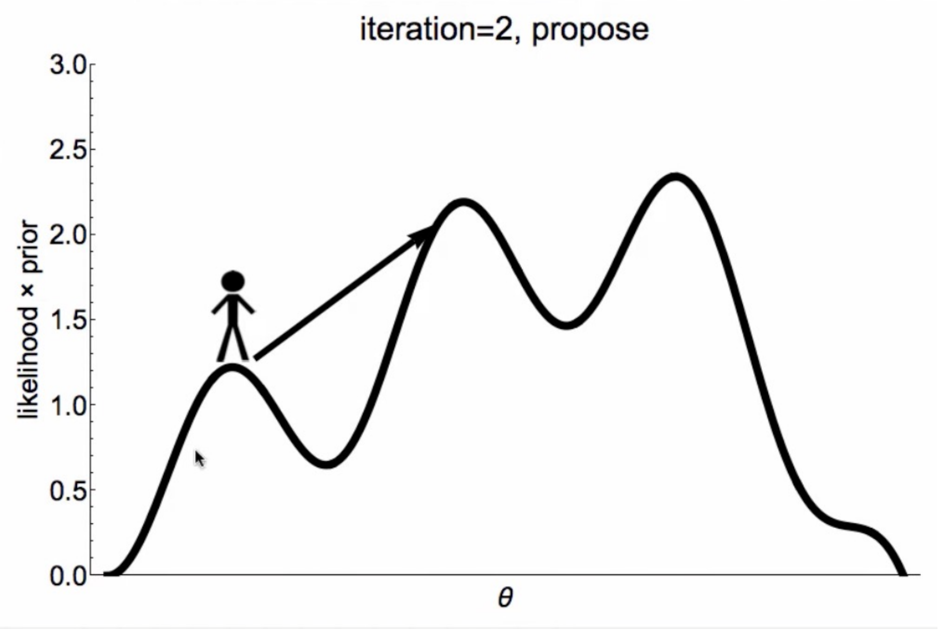
ANIMATION MH ALGORITHM



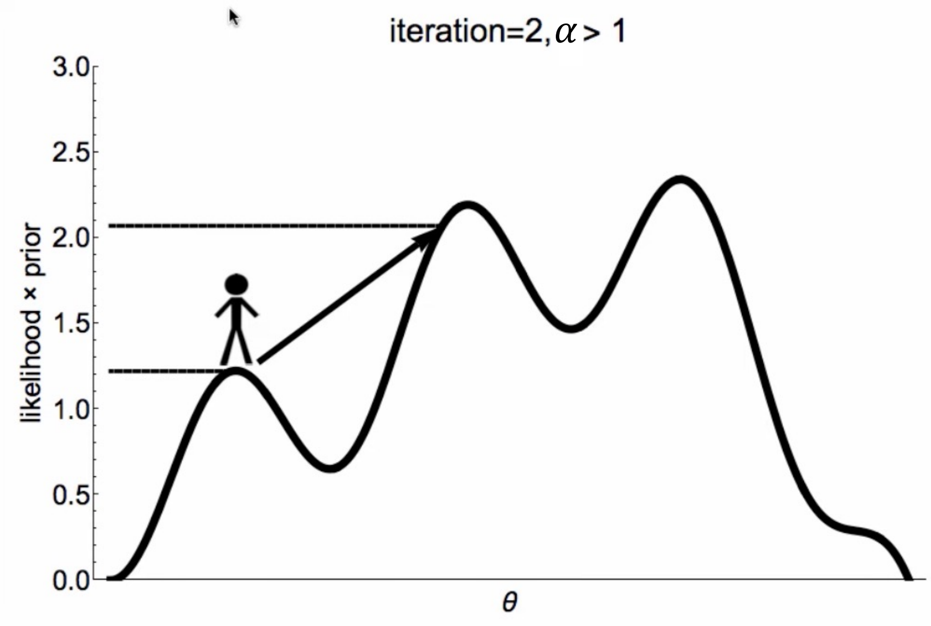
ANIMATION MH ALGORITHM



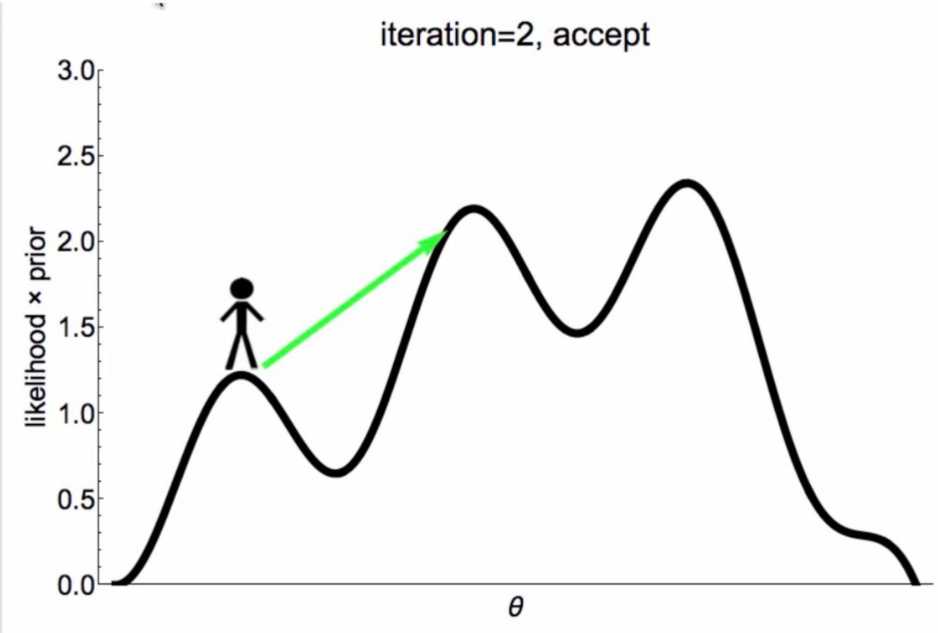
ANIMATION MH ALGORITHM



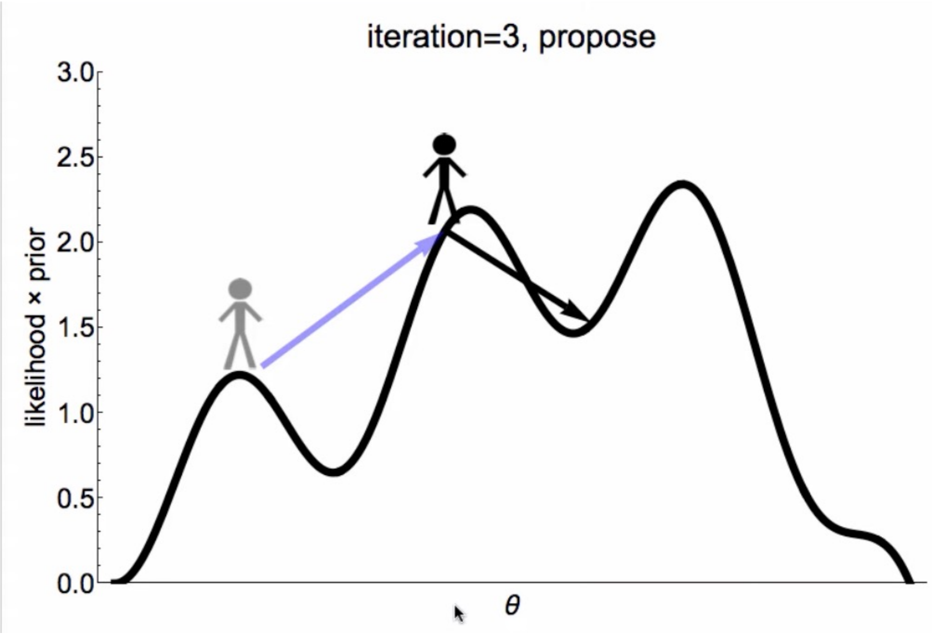
ANIMATION MH ALGORITHM



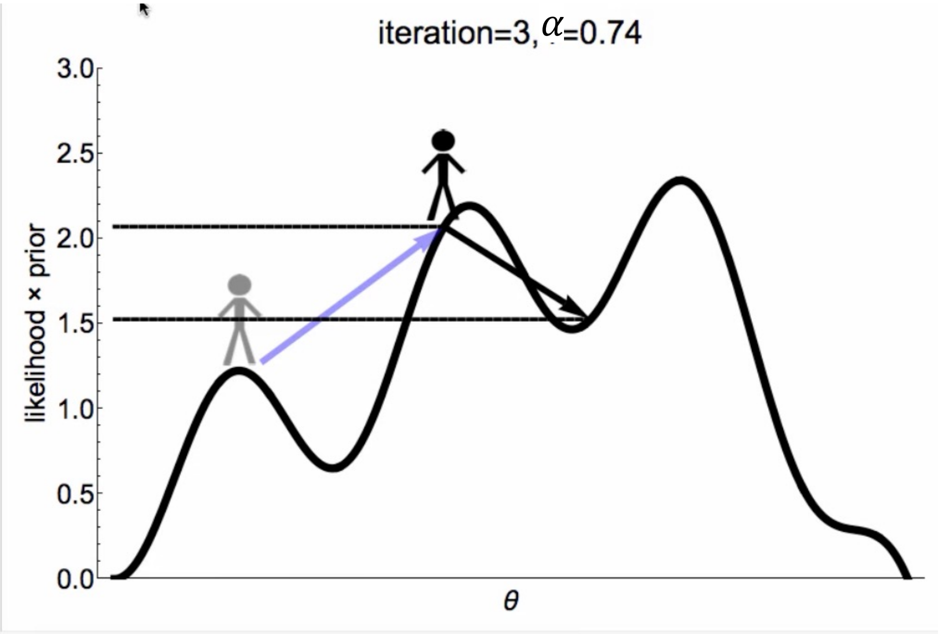
ANIMATION MH ALGORITHM



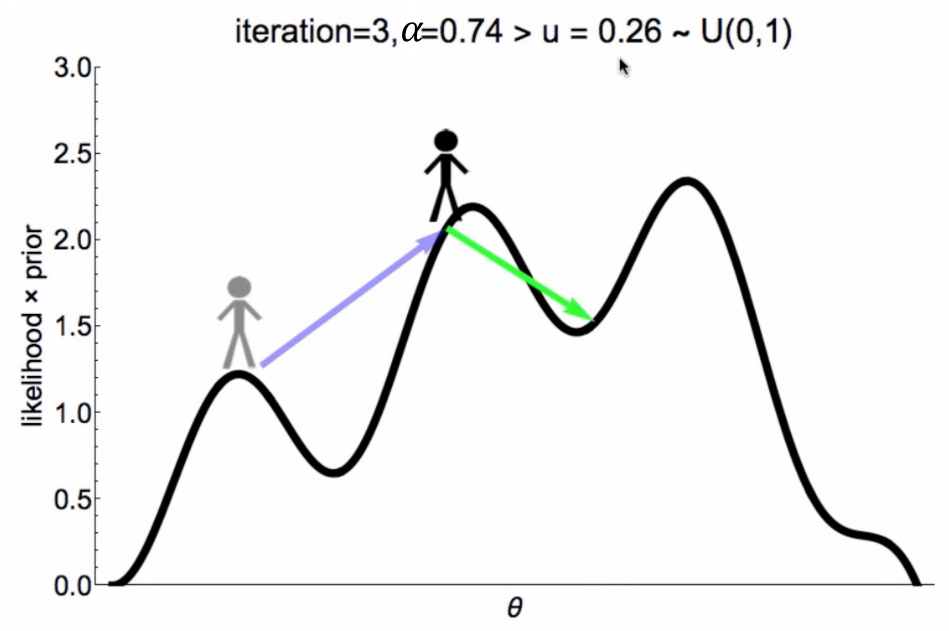
ANIMATION MH ALGORITHM



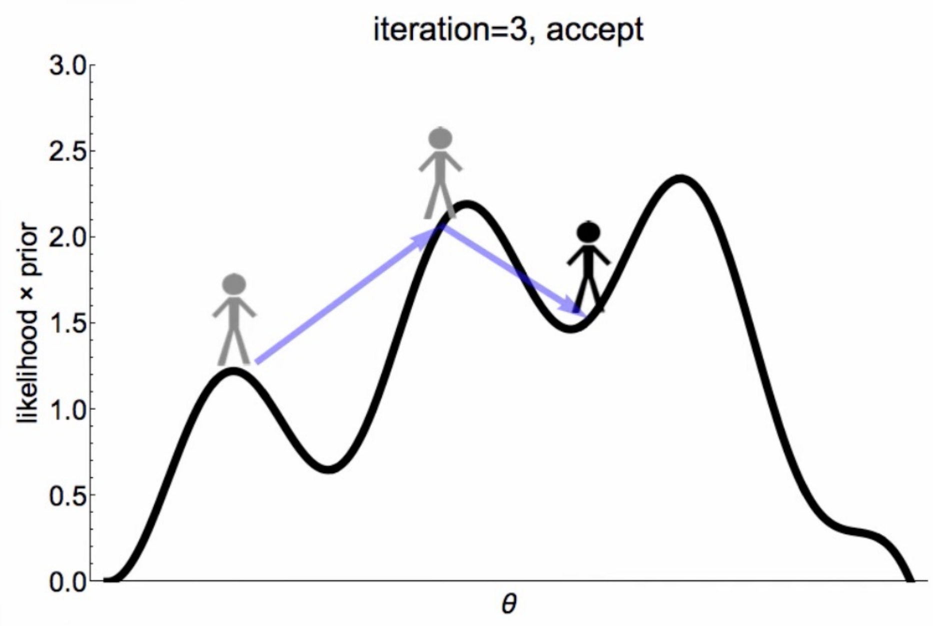
ANIMATION MH ALGORITHM



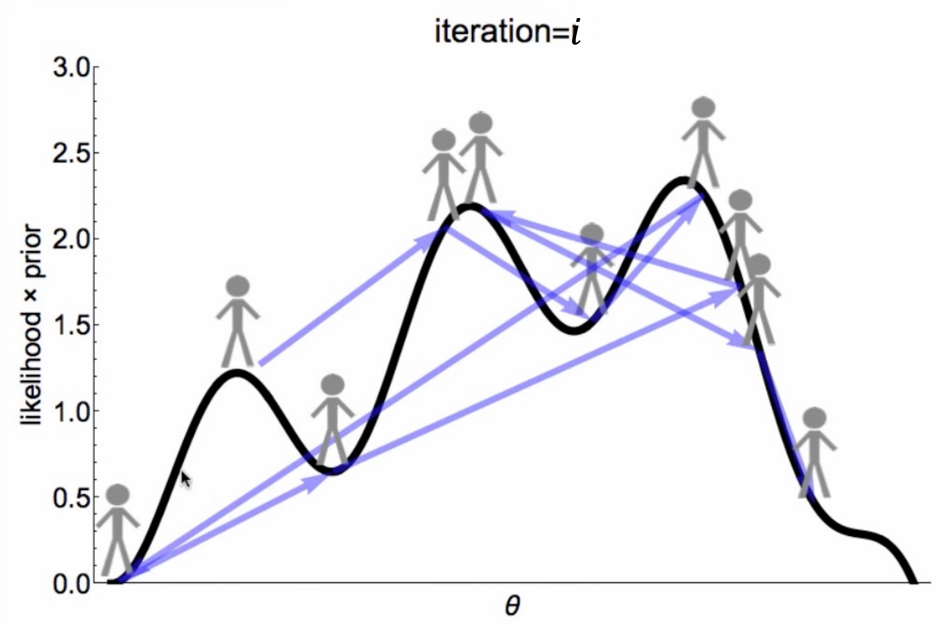
ANIMATION MH ALGORITHM



ANIMATION MH ALGORITHM



ANIMATION MH ALGORITHM



BACK TO THE SIR MODEL: WHAT DO WE NEED?

- Prior distributions for β and γ.
  - Since β and γ are positive rates, we use priors defined on $(0, \infty)$. Common choices:
    1. Exponential prior (simple, non-informative):
       $$\beta \sim Exponential(\lambda_\beta) \text{ and } \gamma \sim Exponential(\lambda_\gamma).$$
    2. Gamma prior (more flexible):
       $$\beta \sim Gamma(a_\beta, b_\beta) \text{ and } \gamma \sim Gamma(a_\gamma, b_\gamma).$$

- Likelihood: same as in MLE.
  - $L\,(\beta, \gamma \mid \mathrm{S_t}, \mathrm{R_t}, \mathrm{I_t}) = \prod_{t=0}^{T-2} L_t(\beta, \gamma \mid \mathrm{S_t}, \mathrm{R_t}, \mathrm{I_t})$.

- Posterior: combine prior and likelihood.
  - $p(\beta, \gamma \mid \mathrm{S_t}, \mathrm{R_t}, \mathrm{I_t}) \propto L\,(\beta, \gamma \mid \mathrm{S_t}, \mathrm{R_t}, \mathrm{I_t})\; p(\beta)\; p(\gamma)$.

RECAP - MH ALGORITHM (PROPOSAL)

- Propose $\theta^* \sim N\left(\theta^{(i-1)}, \Sigma\right)$.
  - We typically use a Normal distribution.
  - In SIR model $\theta = (\beta, \gamma)$ → Multivariate Normal distribution.
  - Σ is the proposal covariance matrix.
    - Influences how efficiently the algorithm explores the parameter space.
    - Controls the step size and correlation between β and γ.
    - For illustration, we assume Σ is diagonal, meaning that β and γ are proposed independently from normal distributions. Specifically:

$$\beta^* \sim N(\beta^{(i-1)}, \sigma^2),$$

$$\gamma^* \sim N\left(\gamma^{(i-1)}, \sigma^2\right).$$

RECAP - MH ALGORITHM (USING LOGS)

- Compute acceptance ratio:

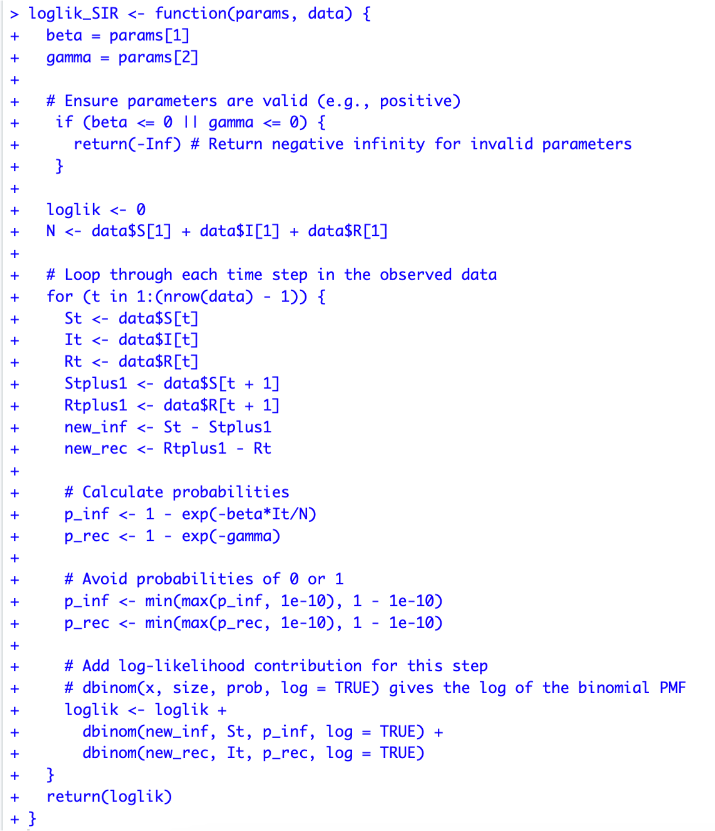
$$\alpha = \min\left(1, \frac{p(\theta^*|Data)}{p(\theta^{(i-1)}|Data)}\right) = \min\left(1, \frac{L(\theta^*|Data)\, p(\theta^*)}{L(\theta^{(i-1)}|Data)\, p(\theta^{(i-1)})}\right).$$

  - When computing the acceptance ratio α, it's common practice to work with the logarithm of the posterior, i.e.

$$\log\big(L(\theta^*|Data)\big) + \log\big(p(\theta^*)\big) - \log\left(L\big(\theta^{(i-1)}|Data\big)\right) - \log(p(\theta^{(i-1)})).$$

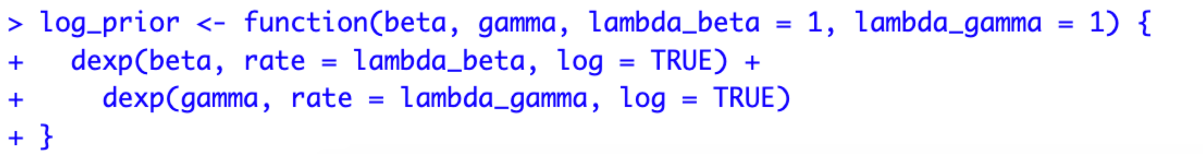
  - Working with logs avoids numerical underflow/overflow issues because likelihoods and priors can be very small numbers, and their product can be even smaller.

R CODE: LOG-LIKELIHOOD FUNCTION (SAME AS IN MLE)



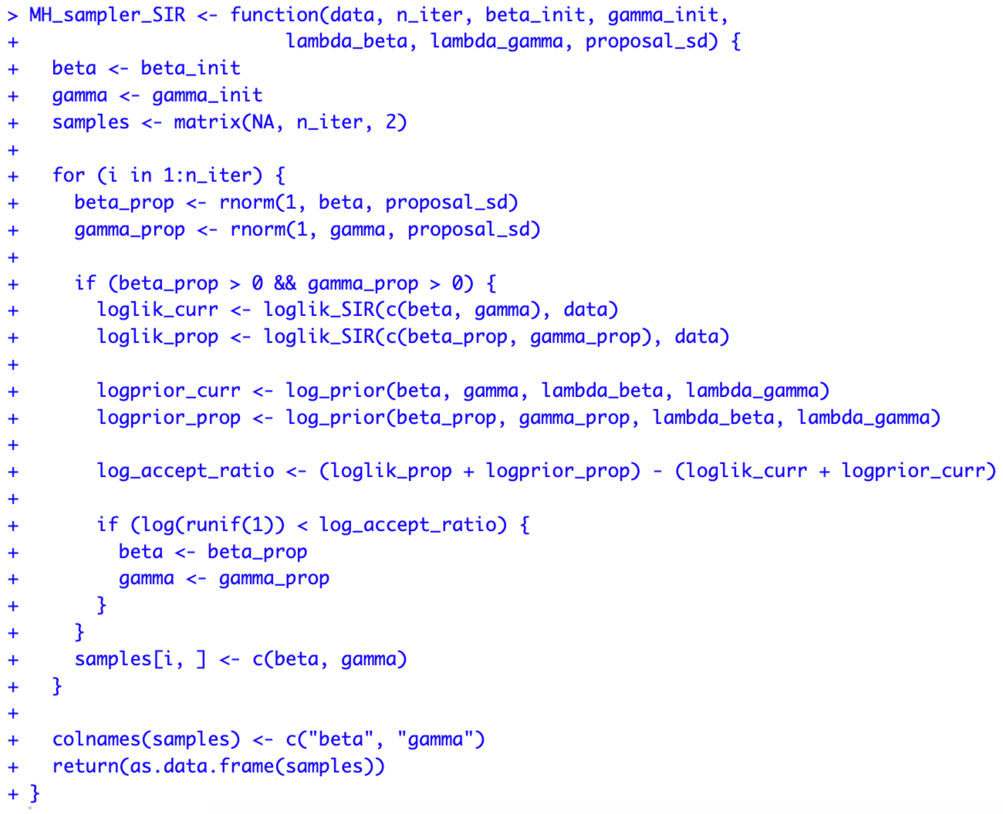
```
> loglik_SIR <- function(params, data) {
+   beta = params[1]
+   gamma = params[2]
+ 
+   # Ensure parameters are valid (e.g., positive)
+    if (beta <= 0 || gamma <= 0) {
+      return(-Inf) # Return negative infinity for invalid parameters
+    }
+ 
+   loglik <- 0
+   N <- data$S[1] + data$I[1] + data$R[1]
+ 
+   # Loop through each time step in the observed data
+   for (t in 1:(nrow(data) - 1)) {
+     St <- data$S[t]
+     It <- data$I[t]
+     Rt <- data$R[t]
+     Stplus1 <- data$S[t + 1]
+     Rtplus1 <- data$R[t + 1]
+     new_inf <- St - Stplus1
+     new_rec <- Rtplus1 - Rt
+ 
+     # Calculate probabilities
+     p_inf <- 1 - exp(-beta*It/N)
+     p_rec <- 1 - exp(-gamma)
+ 
+     # Avoid probabilities of 0 or 1
+     p_inf <- min(max(p_inf, 1e-10), 1 - 1e-10)
+     p_rec <- min(max(p_rec, 1e-10), 1 - 1e-10)
+ 
+     # Add log-likelihood contribution for this step
+     # dbinom(x, size, prob, log = TRUE) gives the log of the binomial PMF
+     loglik <- loglik +
+       dbinom(new_inf, St, p_inf, log = TRUE) +
+       dbinom(new_rec, It, p_rec, log = TRUE)
+   }
+   return(loglik)
+ }
```

R CODE: LOG-PRIOR FUNCTION (EXPONENTIAL PRIORS)



```
> log_prior <- function(beta, gamma, lambda_beta = 1, lambda_gamma = 1) {
+   dexp(beta, rate = lambda_beta, log = TRUE) +
+     dexp(gamma, rate = lambda_gamma, log = TRUE)
+ }
```

R CODE: MH SAMPLER

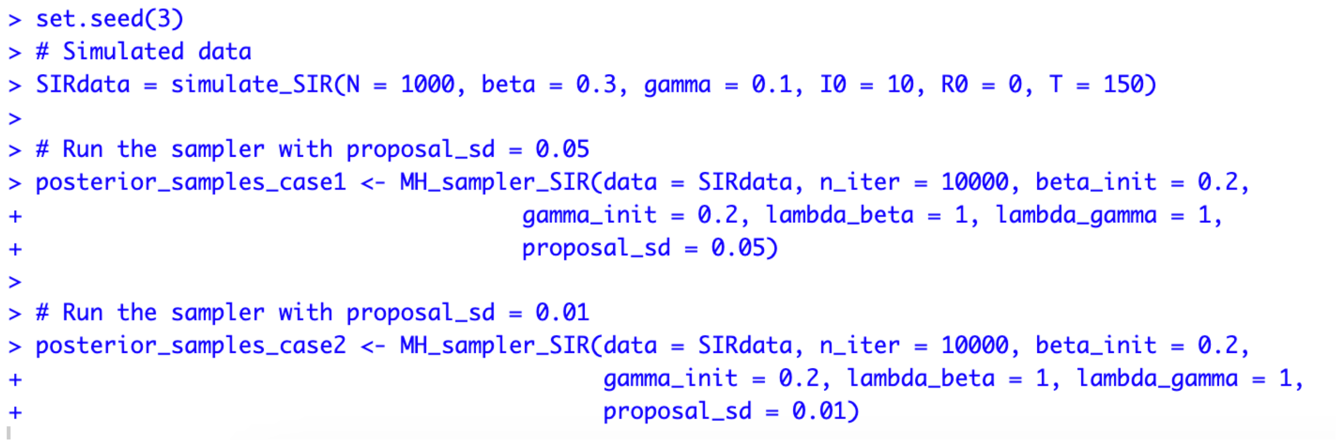


```
> MH_sampler_SIR <- function(data, n_iter, beta_init, gamma_init,
+                            lambda_beta, lambda_gamma, proposal_sd) {
+   beta <- beta_init
+   gamma <- gamma_init
+   samples <- matrix(NA, n_iter, 2)
+
+   for (i in 1:n_iter) {
+     beta_prop <- rnorm(1, beta, proposal_sd)
+     gamma_prop <- rnorm(1, gamma, proposal_sd)
+
+     if (beta_prop > 0 && gamma_prop > 0) {
+       loglik_curr <- loglik_SIR(c(beta, gamma), data)
+       loglik_prop <- loglik_SIR(c(beta_prop, gamma_prop), data)
+
+       logprior_curr <- log_prior(beta, gamma, lambda_beta, lambda_gamma)
+       logprior_prop <- log_prior(beta_prop, gamma_prop, lambda_beta, lambda_gamma)
+
+       log_accept_ratio <- (loglik_prop + logprior_prop) - (loglik_curr + logprior_curr)
+
+       if (log(runif(1)) < log_accept_ratio) {
+         beta <- beta_prop
+         gamma <- gamma_prop
+       }
+     }
+     samples[i, ] <- c(beta, gamma)
+   }
+
+   colnames(samples) <- c("beta", "gamma")
+   return(as.data.frame(samples))
+ }
```

Recap:

1. Initialise: Choose starting value $\theta^{(0)} = (\beta^{(0)}, \gamma^{(0)})$.
2. For $i = 1, 2, \ldots, I$:
   - Propose $\beta^* \sim N\left(\beta^{(i-1)}, \sigma^2\right)$ and $\gamma^* \sim N\left(\gamma^{(i-1)}, \sigma^2\right)$.
   - Compute the log acceptance ratio.
   - Accept/reject: Draw $u \sim Uniform(0, 1)$.
     - If $\log(u) < \log(\alpha)$ set $\theta^{(i)} = (\beta^*, \gamma^*)$ [accept].
     - If $\log(u) \geq \log(\alpha)$ set $\theta^{(i)} = \theta^{(i-1)}$ [reject].
3. Return $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(I)}$.

IMPLEMENTING MH IN R



```
> set.seed(3)
> # Simulated data
> SIRdata = simulate_SIR(N = 1000, beta = 0.3, gamma = 0.1, I0 = 10, R0 = 0, T = 150)
>
> # Run the sampler with proposal_sd = 0.05
> posterior_samples_case1 <- MH_sampler_SIR(data = SIRdata, n_iter = 10000, beta_init = 0.2,
+                                     gamma_init = 0.2, lambda_beta = 1, lambda_gamma = 1,
+                                     proposal_sd = 0.05)
>
> # Run the sampler with proposal_sd = 0.01
> posterior_samples_case2 <- MH_sampler_SIR(data = SIRdata, n_iter = 10000, beta_init = 0.2,
+                                           gamma_init = 0.2, lambda_beta = 1, lambda_gamma = 1,
+                                           proposal_sd = 0.01)
```

Key components of the MH sampler:

- **Number of iterations:** Should be large enough to explore the posterior.
- **Tuning:** Adjust proposal variance to achieve an acceptance rate of ~20-40%.

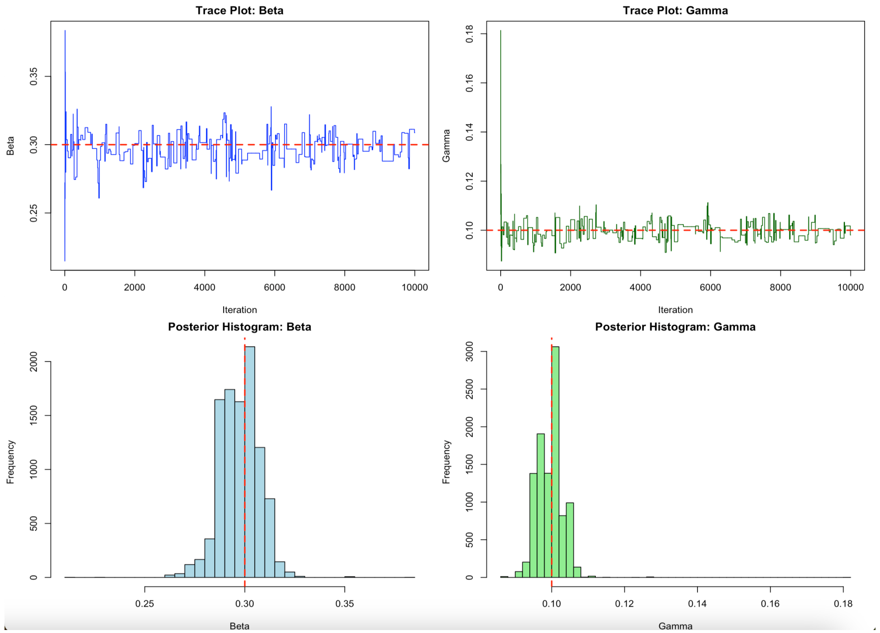
KEY DIAGNOSTICS FOR CONVERGENCE

It's crucial to check if your MCMC sampler has run long enough and is properly sampling from the posterior. This is called convergence diagnostics.

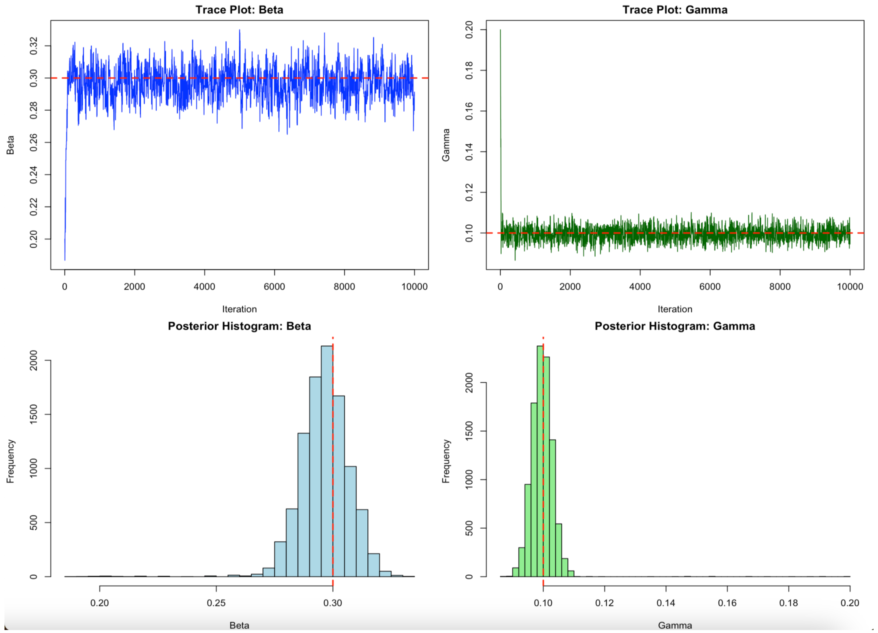
- **Trace plots:** Visualize parameter values over iterations. A good trace plot should look like a 'hairy caterpillar' – dense, random fluctuations around a stable mean, with no clear trends.
- **Autocorrelation plots:** Check how correlated samples are across iterations. We want it to drop quickly.
- **Histogram/density plots:** Assess marginal posterior distributions. They should look smooth and unimodal (unless the posterior is truly multimodal).
- **Multiple chains:** Run several chains from different starting values to check that they all converge to the same distribution.

R OUTPUT: TRACELPLOTS AND HISTOGRAMS

Case 1: proposal SD 0.05



Case 2: proposal SD 0.01



INTERPRETING BAYESIAN RESULTS

Once MCMC has converged, analyse your samples:

- Posterior distributions:
  - Give a full range of likely values for each parameter, not just one estimate.
  - The peak = most likely value (posterior mode).
- Credible intervals (CIs):
  - Bayesian version of confidence intervals.
  - A 95% CI means there's a 95% chance the true value lies in that interval.
  - Compute using the 2.5th and 97.5th percentiles of your samples.
- Posterior mean/median:
  - Point estimates: mean = average, median = middle of the sampled values.

R OUTPUT: SUMMARY

Summary table showing the mean, median, and 95% credible intervals for the posterior samples of β and γ:

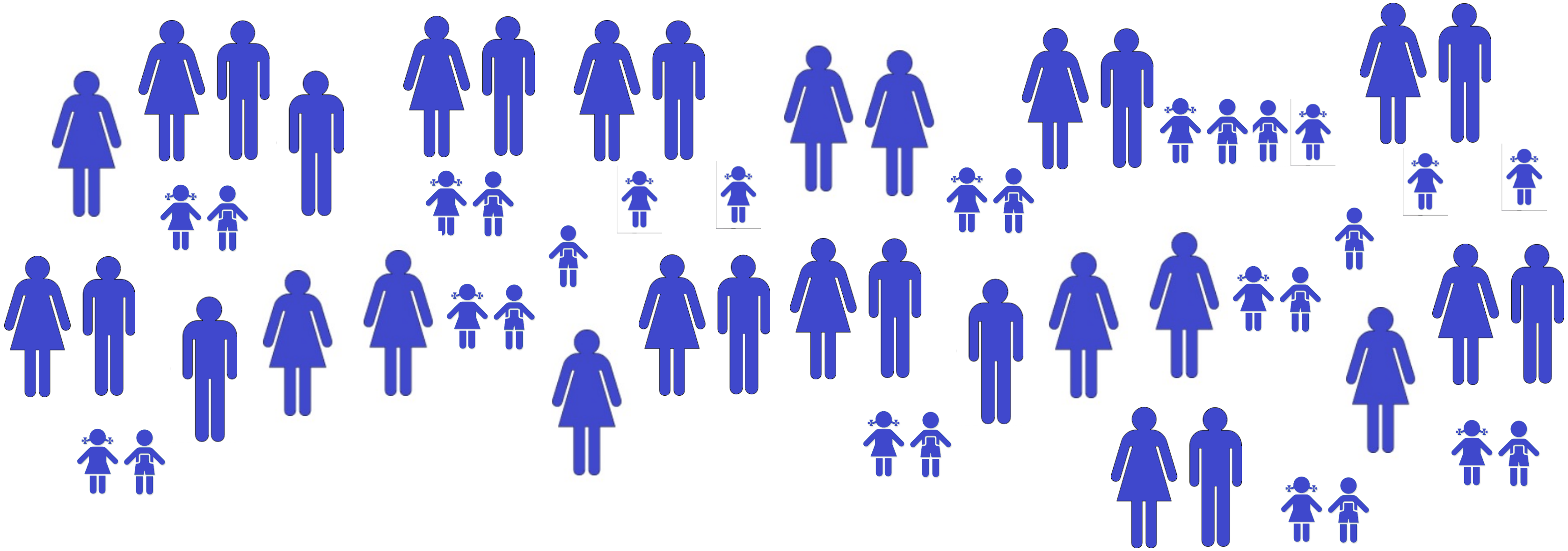
Parameter	Mean	Median	95% CI Lower	95% CI Upper
β	0.2964706	0.2966657	0.2779652	0.3152894
γ	0.09966155	0.0996311	0.09308974	0.1060935

ONCE WE ESTIMATE THE PARAMETERS, WHAT CAN WE DO?

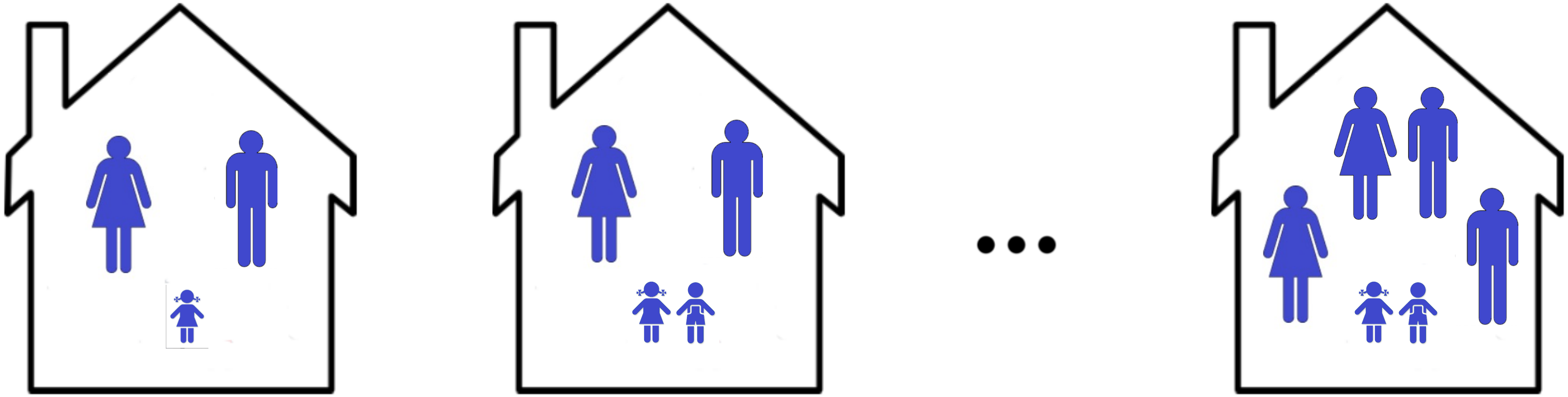
- Understand disease dynamics:
  - β tells us how contagious the disease is.
  - γ tells us how quickly people recover.
- Make predictions about future outbreaks.
  - Simulate future outbreaks.
  - Estimate epidemic duration and size.
- Evaluate interventions.
  - What happens if we reduce β (e.g., masks, distancing)?
  - What if we increase γ (e.g., faster treatment)?
- Inform public health decisions and policy.
  - Guide decisions on lockdowns, vaccination, and resource allocation.

INDIVIDUAL LEVEL EPIDEMIC DATA

POPULATION LEVEL DATA



INDIVIDUAL LEVEL DATA

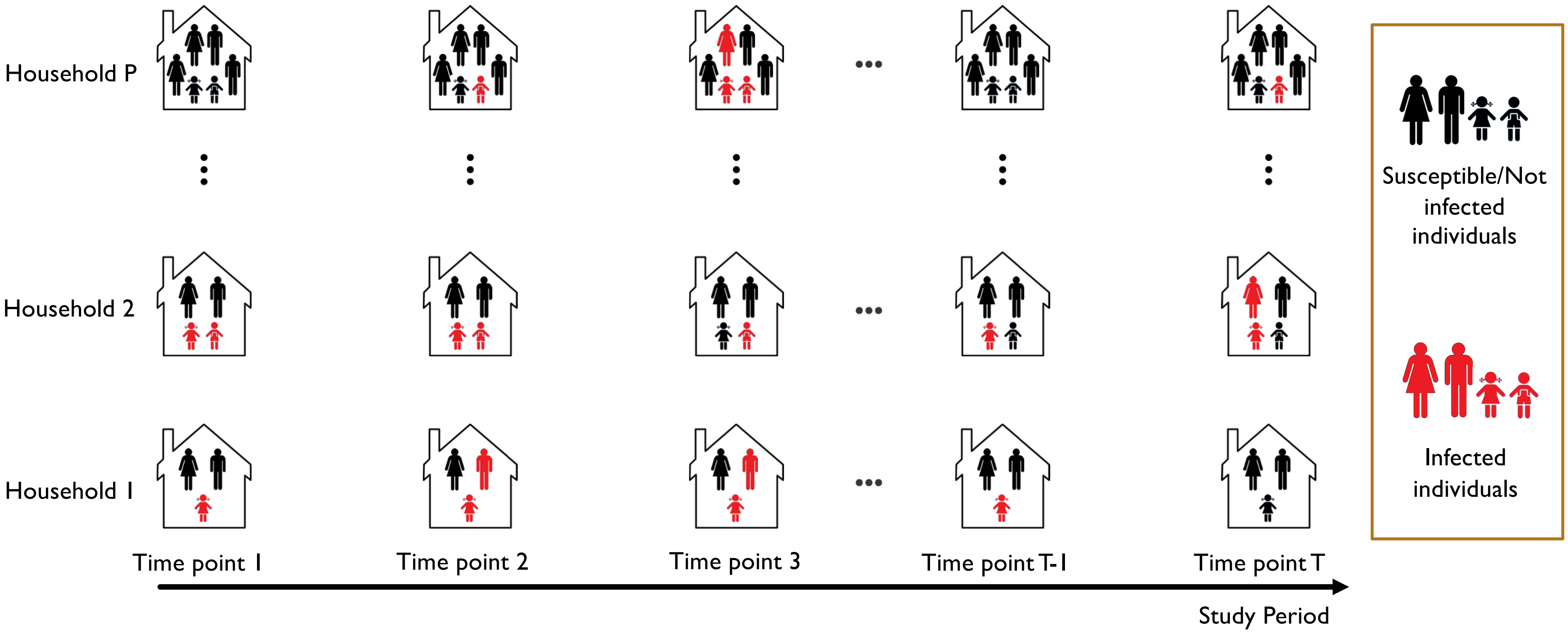


Household 1

Household 2

Household P

OVERVIEW OF THE OBSERVED DATA



THE SIS TRANSMISSION MODEL

BASIC ASSUMPTIONS

At any given time point, each individual is in one of the two following disease states:

- **Susceptible** (**S**): do not have the disease and are able to be infected by it.
- **Infected** (**I**): have the disease and are able to infect susceptible individuals.

The transmission of the disease occurs when infected individuals transmit the disease to healthy susceptible individuals. **Susceptible individuals acquire infection** via two possible routes:

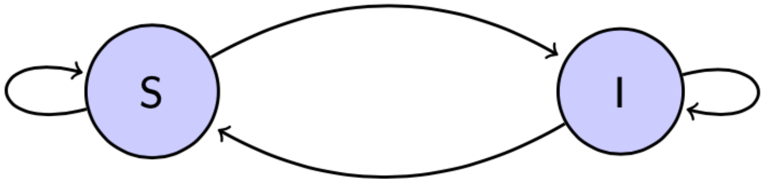
- Direct or indirect transmission from other infected individuals within the same household, with rate β.
- External or community transmission; transmission from other environmental sources from outside the household, with rate α.

Infected individuals recover and move into the susceptible state, with rate γ.

THE SIS TRANSMISSION MODEL

TRANSITION PROBABILITIES BETWEEN THE STATES

- Two health states: **S** and **I**.



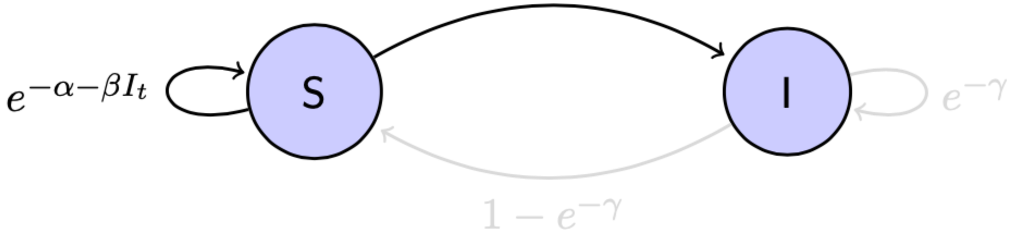
- α: community acquisition rate.
- β: within-household acquisition rate.
- γ: recovery rate.
- I_t: the number of infected individual in the household at time t.

THE SIS TRANSMISSION MODEL

TRANSITION PROBABILITIES BETWEEN THE STATES

- Two health states: **S** and **I**.

P(**S** → **I**) = $1 - e^{-\alpha - \beta I_t}$

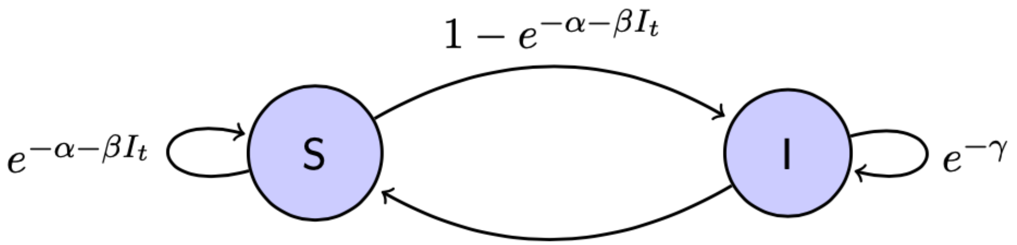


- α: community acquisition rate.
- β: within-household acquisition rate.
- γ: recovery rate.
- I_t: the number of infected individual in the household at time t.

THE SIS TRANSMISSION MODEL

TRANSITION PROBABILITIES BETWEEN THE STATES

- Two health states: **S** and **I**.

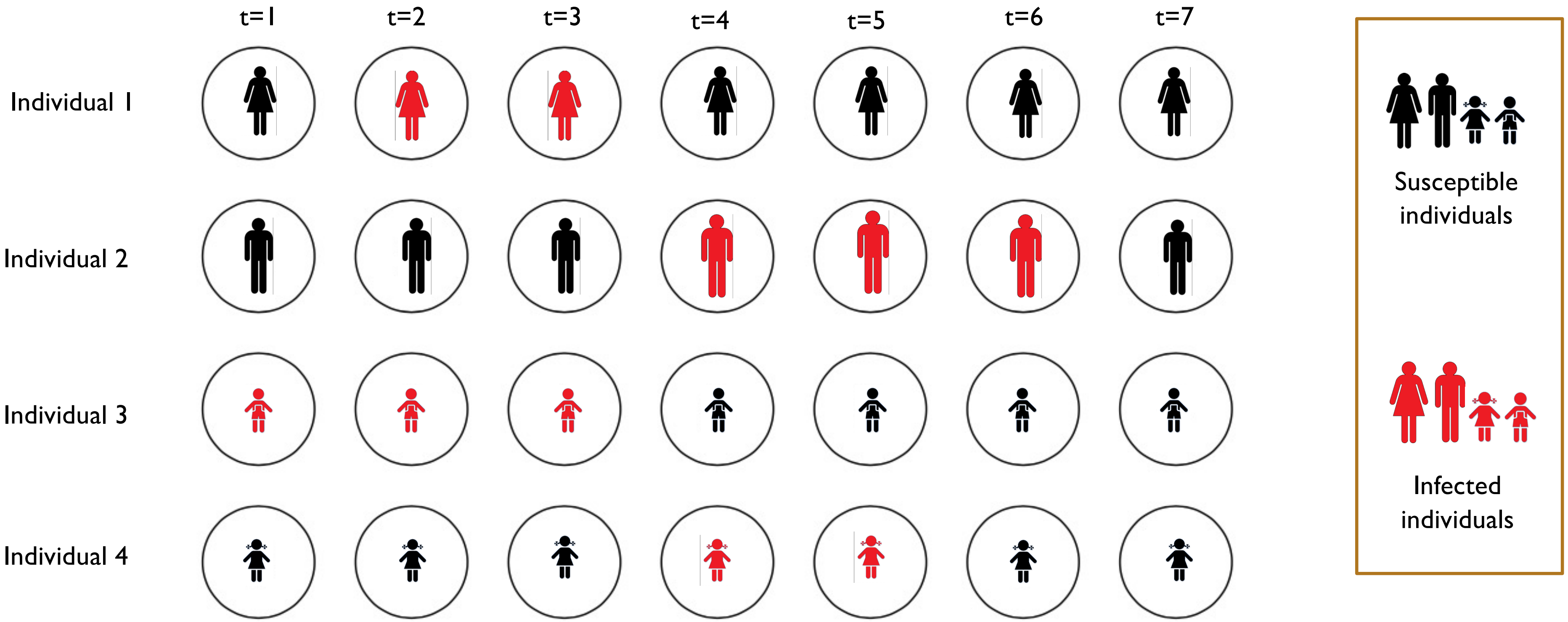


P(**I** → **S**) = $1 - e^{-\gamma}$

- α: community acquisition rate.
- β: within-household acquisition rate.
- γ: recovery rate.
- I_t: the number of infected individual in the household at time t.

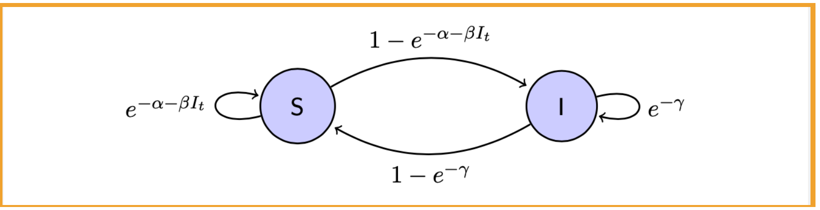
AN EXAMPLE

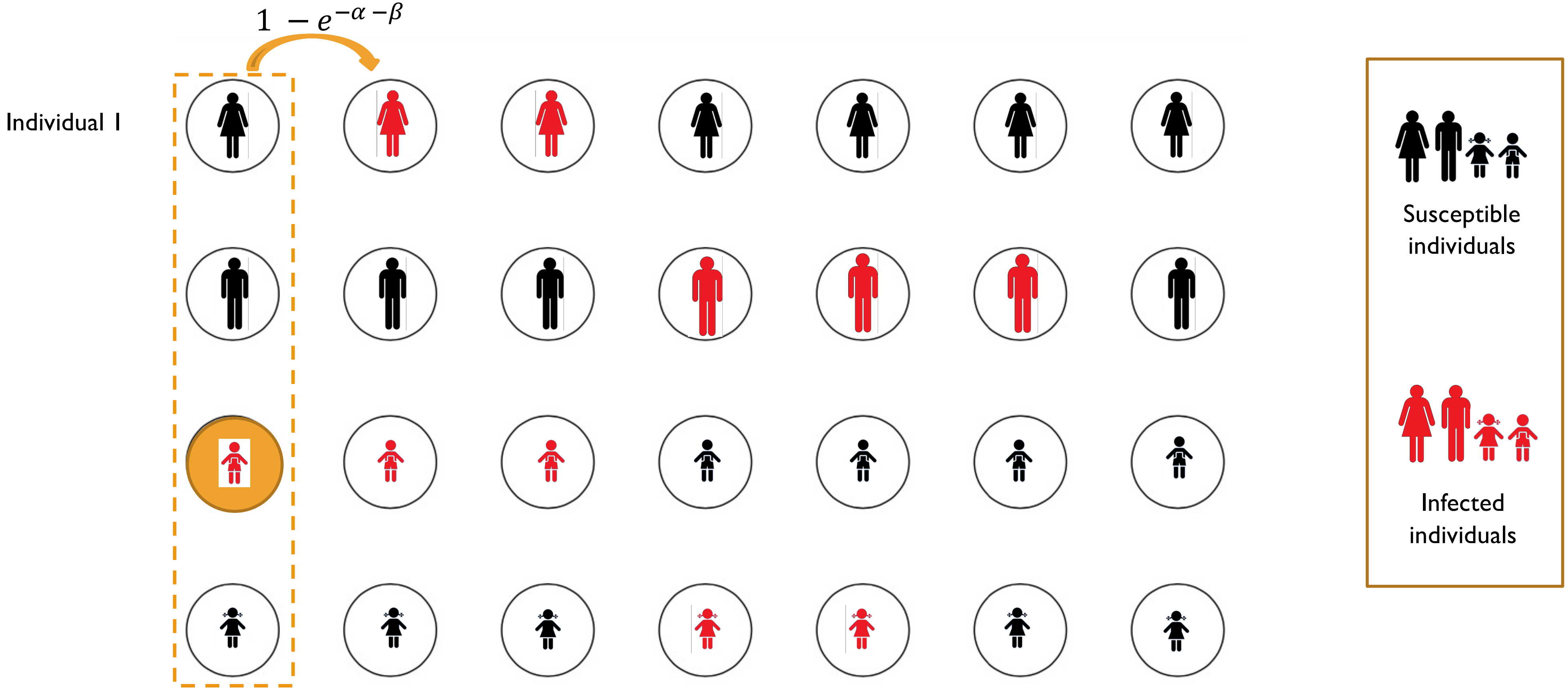
1 HOUSEHOLD, 4 INDIVIDUALS & 7 TIME POINTS



AN EXAMPLE

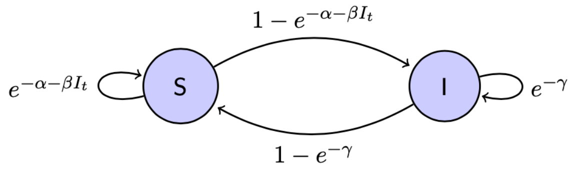
TRANSITION PROBABILITIES

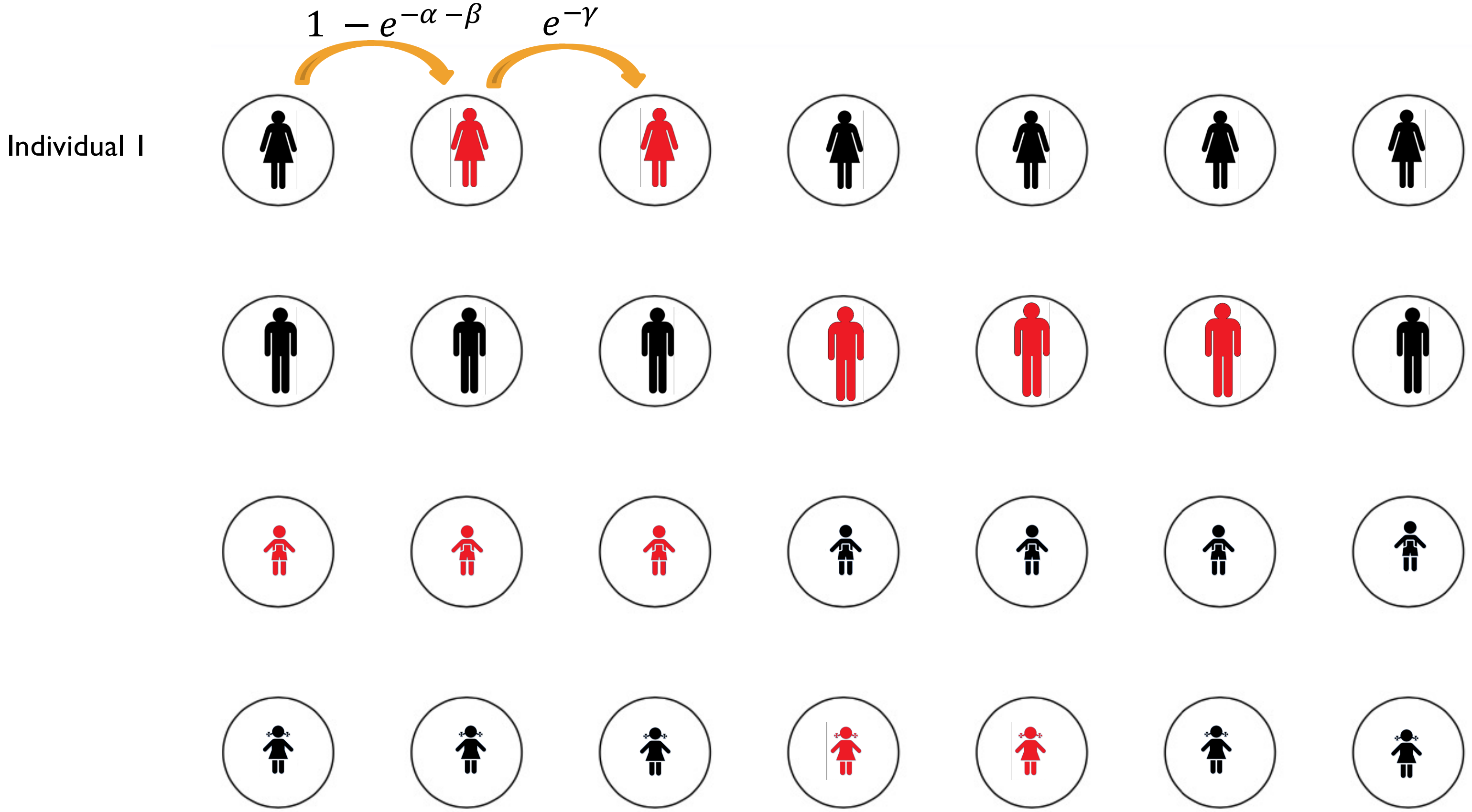


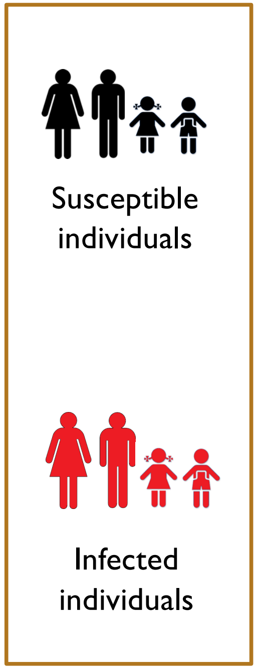


AN EXAMPLE

TRANSITION PROBABILITIES

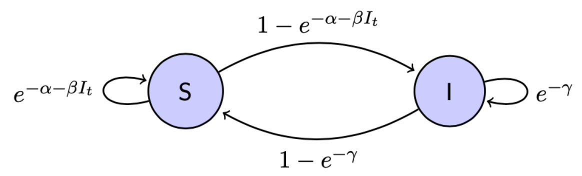


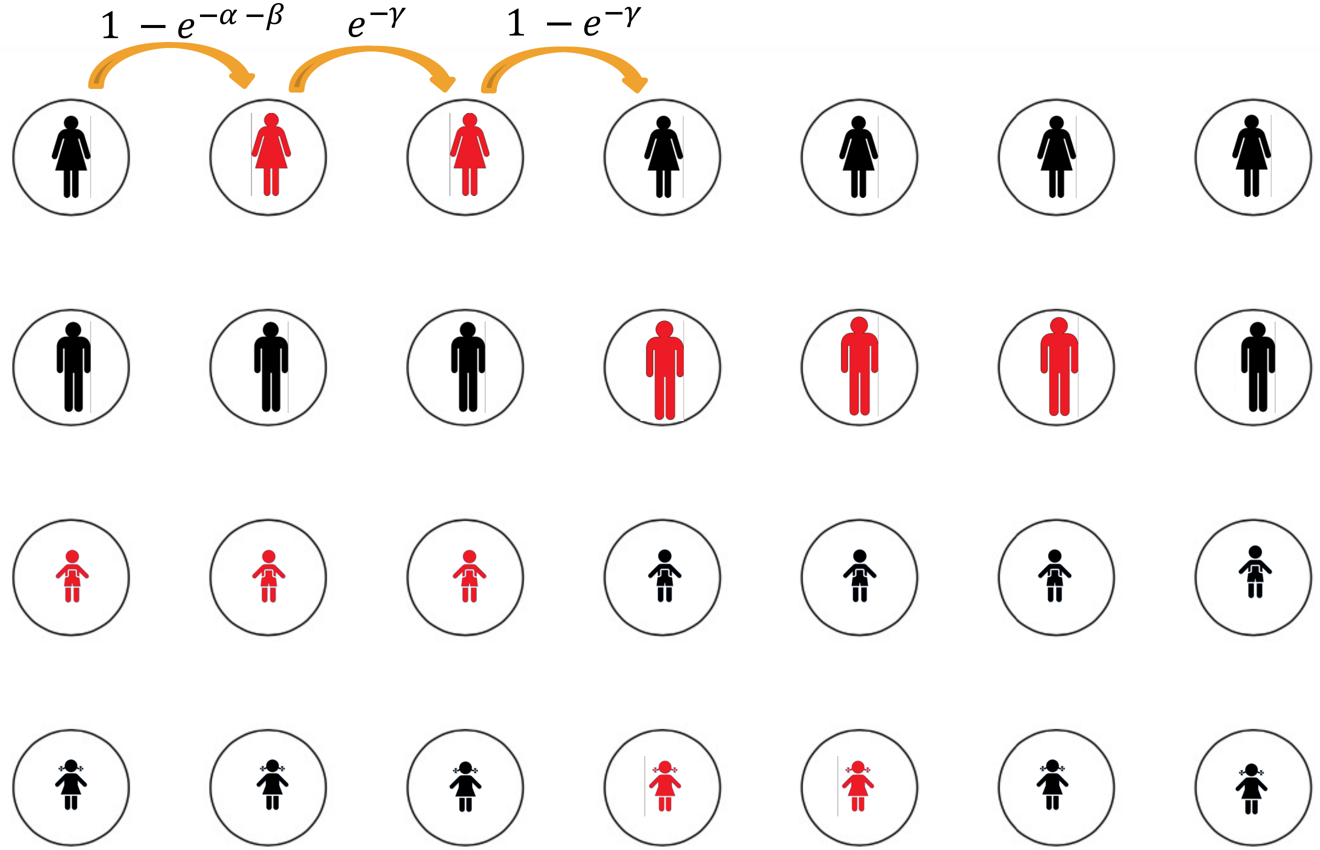




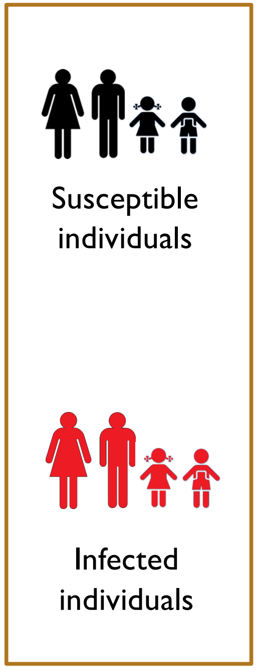
AN EXAMPLE

TRANSITION PROBABILITIES





Individual 1

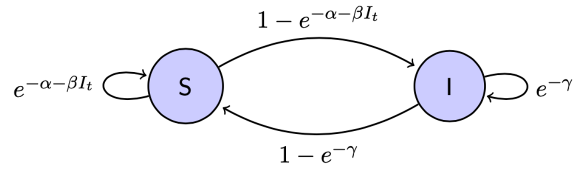


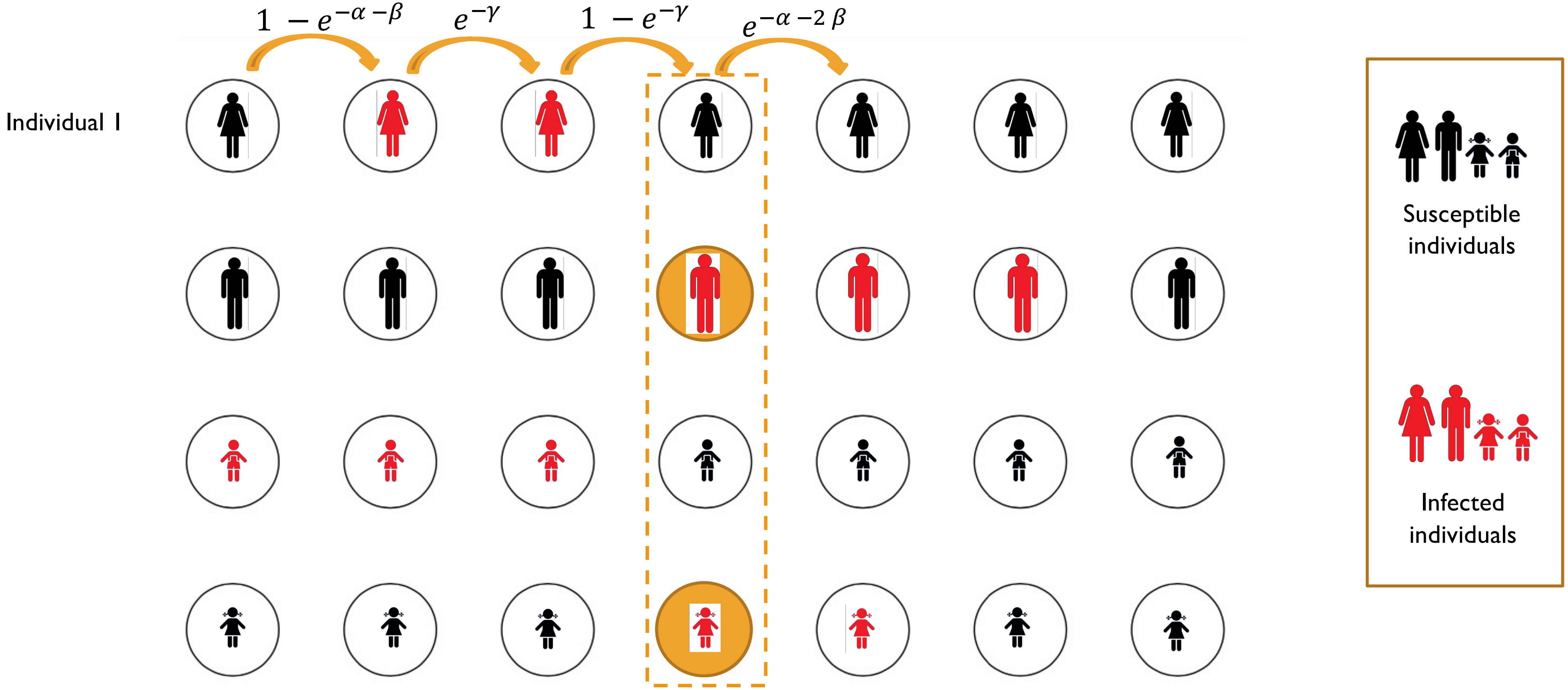
Susceptible individuals

Infected individuals

AN EXAMPLE

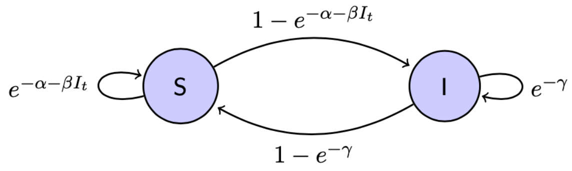
TRANSITION PROBABILITIES

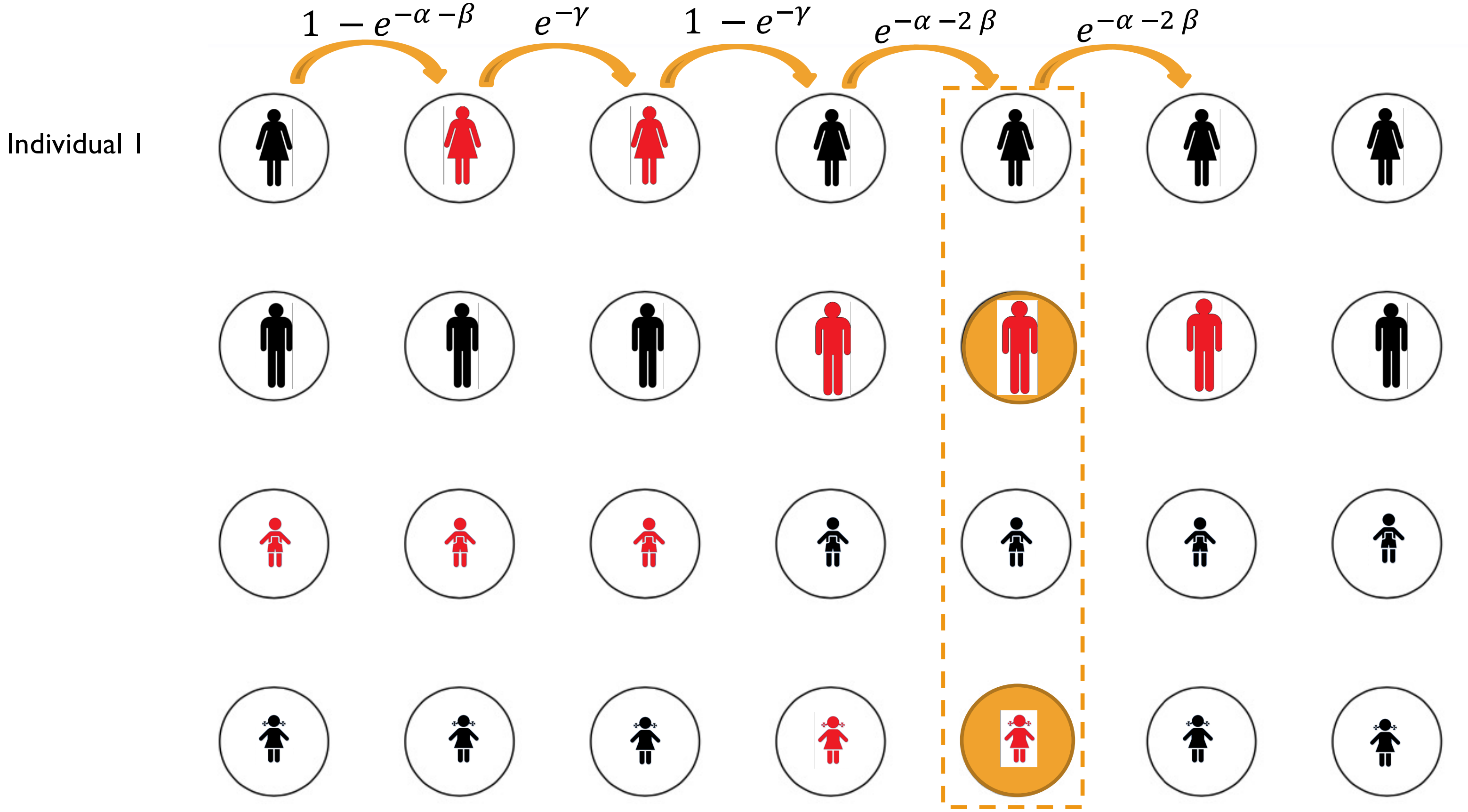




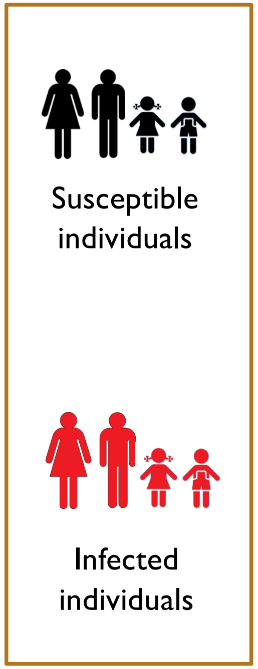
AN EXAMPLE

TRANSITION PROBABILITIES



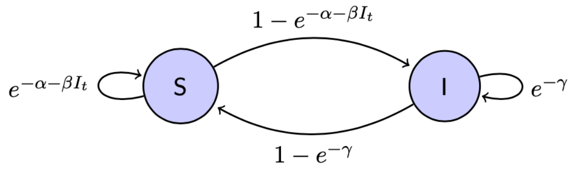


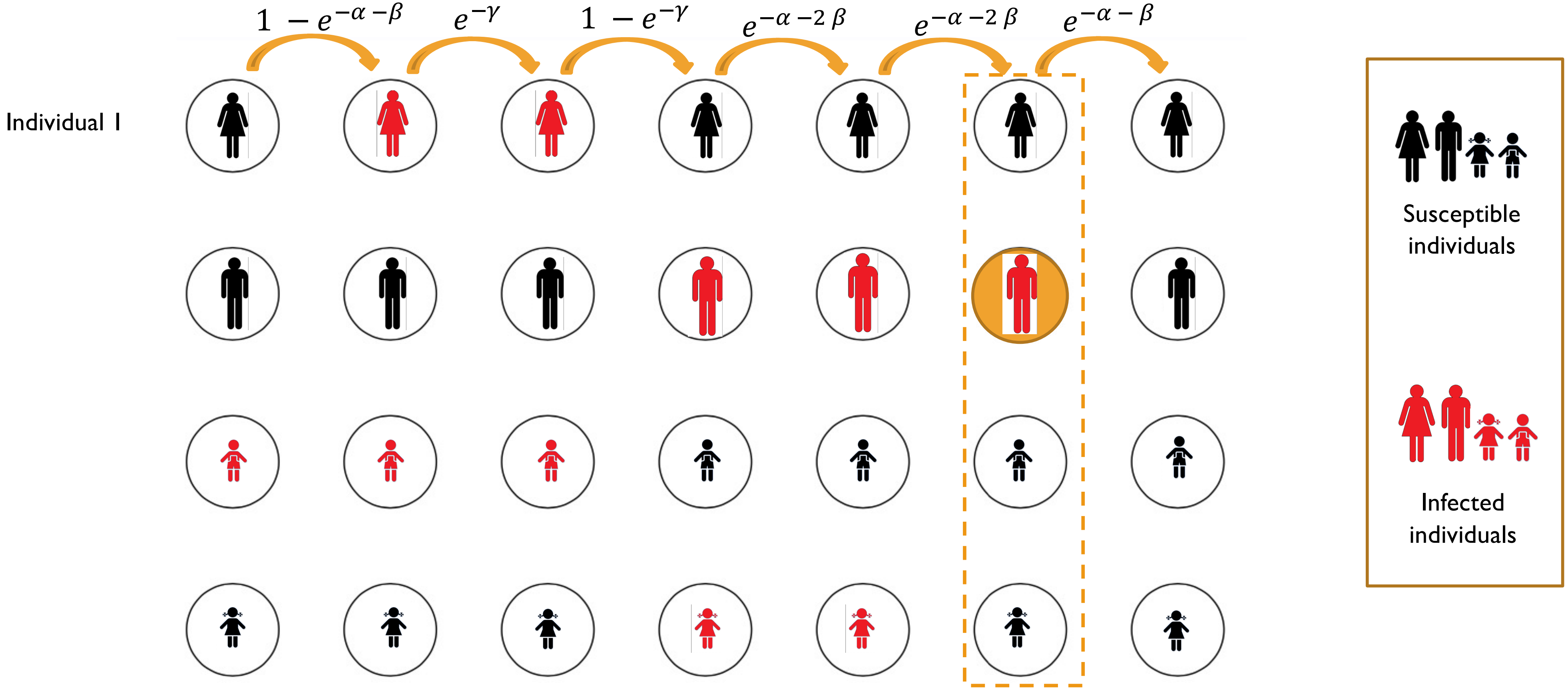
$1 \; - e^{-\alpha\, -\beta}$
$e^{-\gamma}$
$1 \; - e^{-\gamma}$
$e^{-\alpha\, -2\,\beta}$
$e^{-\alpha\, -2\,\beta}$
Individual I



AN EXAMPLE

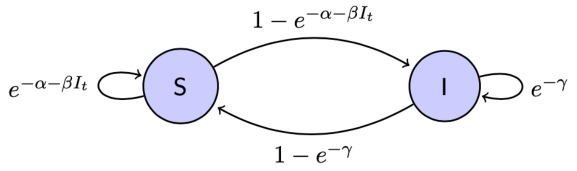
TRANSITION PROBABILITIES

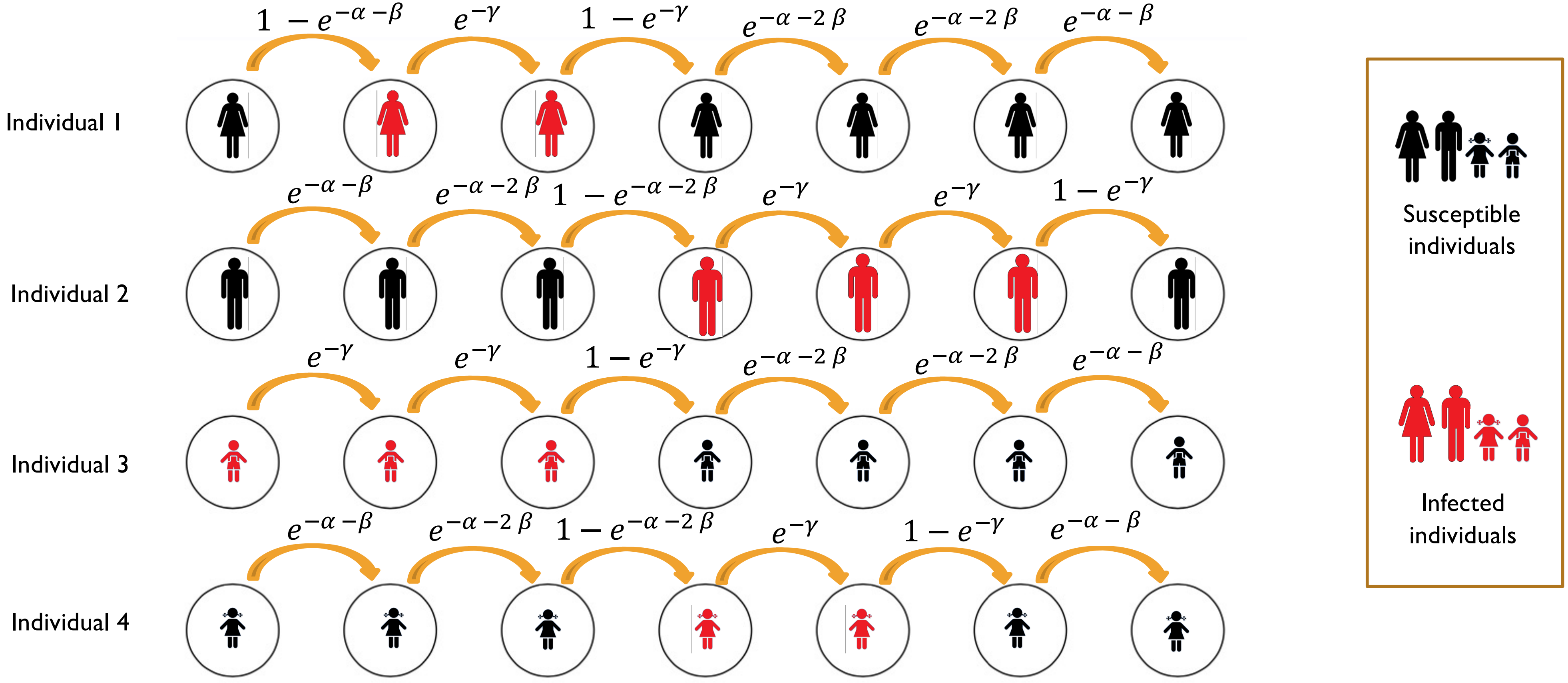




AN EXAMPLE

TRANSITION PROBABILITIES





PARAMETER ESTIMATION

Which are the unknown parameters of the model?

PARAMETER ESTIMATION

Which are the unknown parameters of the model?

- α: community acquisition rate.
- β: within-household acquisition rate.
- γ: recovery rate.

PARAMETER ESTIMATION

Which are the unknown parameters of the model?

- α: community acquisition rate.
- β: within-household acquisition rate.
- γ: recovery rate.

How can we estimate them?

PARAMETER ESTIMATION

Which are the unknown parameters of the model?

- α: community acquisition rate.
- β: within-household acquisition rate.
- γ: recovery rate.

How can we estimate them?

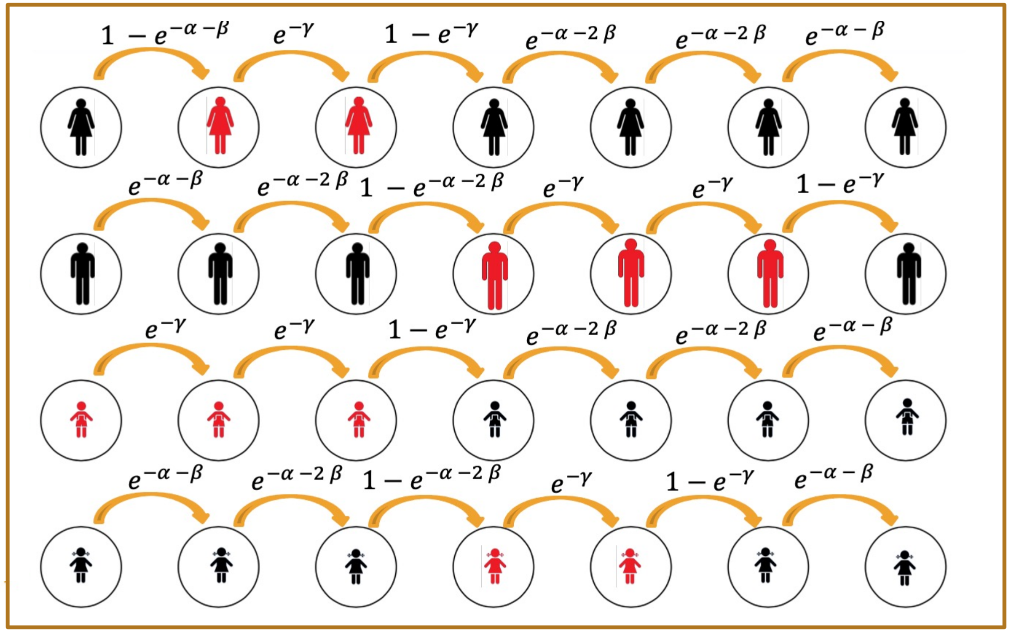
- **Maximum likelihood estimation or**
- **Metropolis Hastings.**

AN EXAMPLE CONTINUED

LIKELIHOOD

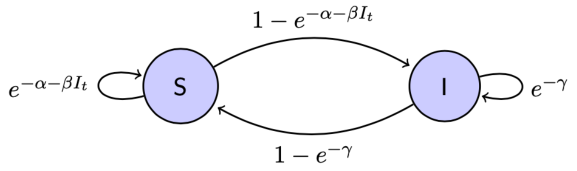
The likelihood function is given by the product of the transition probabilities:

$$L(\alpha,\beta,\gamma) = \left(1 - e^{-\alpha-\beta}\right)\left(e^{-\alpha-\beta}\right)^5\left(1 - e^{-\alpha-2\beta}\right)^2\left(e^{-\alpha-2\beta}\right)^6(e^{-\gamma})^6(1 - e^{-\gamma})^4$$



PARAMETER ESTIMATION

LIKELIHOOD



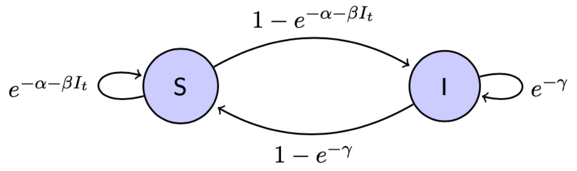
$$L(\alpha,\beta,\gamma) = \underset{i,t:\ X^j_{i,t-1}=1}{\prod_{j=1}^{P}\prod_{t=2}^{T}\prod_{i=1}^{N}} \left[(e^{-\gamma})^{X^j_{i,t}}(1-e^{-\gamma})^{1-X^j_{i,t}}\right] \times \underset{i,t:\ X^j_{i,t-1}=0}{\prod_{j=1}^{P}\prod_{t=2}^{T}\prod_{i=1}^{N}} \left[(1-e^{-\alpha-\beta\sum_{i=1}^{N}X^j_{i,t-1}})^{X^j_{i,t}}(e^{-\alpha-\beta\sum_{i=1}^{N}X^j_{i,t-1}})^{1-X^j_{i,t}}\right]$$

where

- $P =$ Total number of groups.
- $T =$ Total number of time points.
- $N =$ Total number of individuals per group.
- $X^j_{i,t} =$ The health state of individual i at time t in group j.
- $X^j_{i,t} = 1$ if the individual is infected.
- $X^j_{i,t} = 0$ if the individual is susceptible.

PARAMETER ESTIMATION

LIKELIHOOD



$$L(\alpha,\beta,\gamma) = \prod_{j=1}^{P}\prod_{t=2}^{T}\prod_{\substack{i=1 \\ i,t:\ X^j_{i,t-1}=1}}^{N}\left[(e^{-\gamma})^{X^j_{i,t}}(1-e^{-\gamma})^{1-X^j_{i,t}}\right] \times \prod_{j=1}^{P}\prod_{t=2}^{T}\prod_{\substack{i=1 \\ i,t:\ X^j_{i,t-1}=0}}^{N}\left[\left(1-e^{-\alpha-\beta\sum_{i=1}^{N}X^j_{i,t-1}}\right)^{X^j_{i,t}}\left(e^{-\alpha-\beta\sum_{i=1}^{N}X^j_{i,t-1}}\right)^{1-X^j_{i,t}}\right]$$

$$= e^{-\gamma N(1,1)}(1-e^{-\gamma})^{N(1,0)} \times \prod_{j=1}^{P}\prod_{t=2}^{T}\prod_{\substack{i=1 \\ i,t:\ X^j_{i,t-1}=0}}^{N}\left[\left(1-e^{-\alpha-\beta\sum_{i=1}^{N}X^j_{i,t-1}}\right)^{X^j_{i,t}}\left(e^{-\alpha-\beta\sum_{i=1}^{N}X^j_{i,t-1}}\right)^{1-X^j_{i,t}}\right]$$

where

- $N(1,0) =$ Total number of transitions from state 1 to state 0.
- $N(1,1) =$ Total number of transitions from state 1 to state 1.

PARAMETER ESTIMATION

ESTIMATES

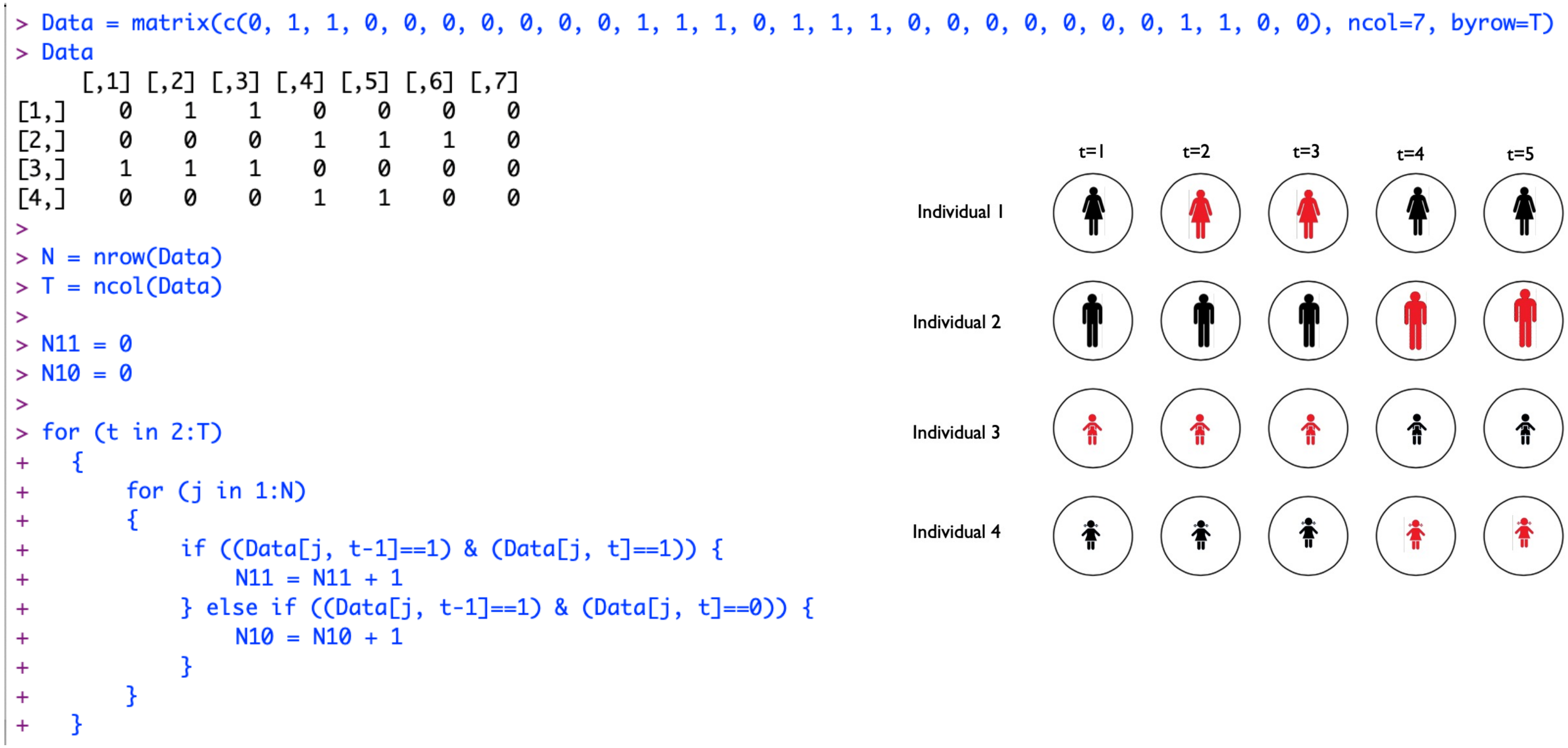
Taking the natural logarithm and differentiating as we did before, we can find the maximum likelihood estimate for γ, given by:

$$\hat{\gamma} = -\log\left(\frac{N(1,1)}{N(1,1)+N(1,0)}\right).$$

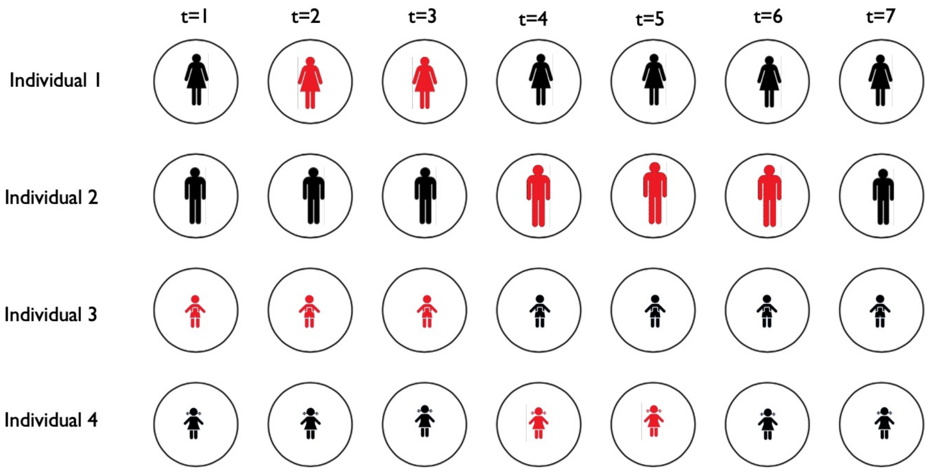
The maximum likelihood estimates for the parameters α and β can be found by maximizing the likelihood function numerically, where γ is replaced by its maximum likelihood estimate.

AN EXAMPLE CONTINUED

ILLUSTRATION IN R



```
> Data = matrix(c(0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0), ncol=7, byrow=T)
> Data
     [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]    0    1    1    0    0    0    0
[2,]    0    0    0    1    1    1    0
[3,]    1    1    1    0    0    0    0
[4,]    0    0    0    1    1    0    0
>
> N = nrow(Data)
> T = ncol(Data)
>
> N11 = 0
> N10 = 0
>
> for (t in 2:T)
+   {
+       for (j in 1:N)
+       {
+           if ((Data[j, t-1]==1) & (Data[j, t]==1)) {
+               N11 = N11 + 1
+           } else if ((Data[j, t-1]==1) & (Data[j, t]==0)) {
+               N10 = N10 + 1
+           }
+       }
+   }
```



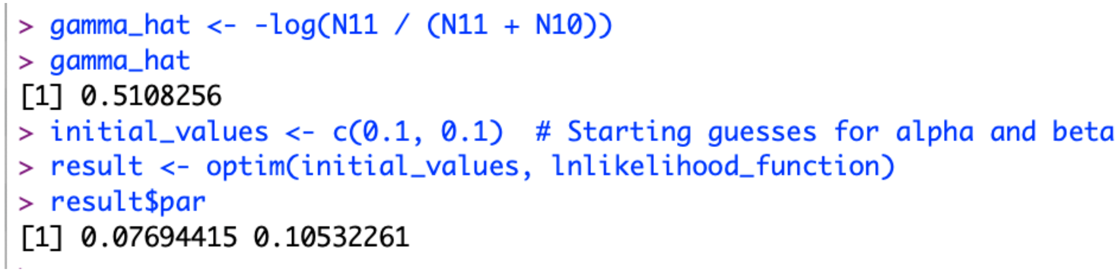
AN EXAMPLE CONTINUED

ILLUSTRATION IN R

```
> lnlikelihood_function <- function(params) {
+     alpha = params[1]
+     beta = params[2]
+
+     N = nrow(Data)
+   T = ncol(Data)
+
+   likelihood = 0
+
+   for (t in 2:T)
+   {
+       It = sum(Data[, t-1])
+       for (j in 1:N)
+       {
+           if ((Data[j, t-1]==0) & (Data[j, t]==0)) {
+               likelihood = likelihood  -alpha - beta*It
+           } else if ((Data[j, t-1]==0) & (Data[j, t]==1)) {
+               likelihood = likelihood + log(1 - exp(-alpha - beta*It))
+           }  else if ((Data[j, t-1]==1) & (Data[j, t]==1)) {
+               likelihood = likelihood -gamma_hat
+           } else {
+               likelihood = likelihood + log(1 - exp(-gamma_hat))
+           }
+       }
+   }
+   return(-likelihood)
+ }
```

AN EXAMPLE CONTINUED

ILLUSTRATION IN R



```
> gamma_hat <- -log(N11 / (N11 + N10))
> gamma_hat
[1] 0.5108256
> initial_values <- c(0.1, 0.1)  # Starting guesses for alpha and beta
> result <- optim(initial_values, lnlikelihood_function)
> result$par
[1] 0.07694415 0.10532261
```

SUMMARY AND RESEARCH DIRECTIONS

CHALLENGES IN REAL-WORLD EPIDEMIC DATA & INFERENCE

- Data quality issues:
  - Under-reporting: Observed cases are often a small fraction of true infections.
  - Missing data: Gaps in time series, incomplete contact tracing.
  - Imperfect diagnostic tests.
  - Reporting delays: The date a case is reported is often not the date of symptom onset or infection.
- Population heterogeneity:
  - Age structure: Different age groups have different contact patterns, susceptibility, and disease severity.
  - Spatial structure: Disease spread varies by location (e.g., cities vs. rural areas).
  - Social networks: Disease spreads along specific contact networks, not randomly.

OPEN PROBLEMS AND RESEARCH DIRECTIONS

- Inference with partial or noisy observations:
  - Problem: In real-world scenarios, we rarely observe the full state (e.g., exact number of infected individuals at each time).
  - Research direction: How can we perform inference when only partial or noisy data is available?
- Model misspecification and robustness:
  - Problem: The model will explore assumes homogeneous mixing and constant parameters.
  - Research direction: What happens when these assumptions are violated.
- Scalability and efficiency of inference algorithms:
  - Problem: MH and MLE can be slow or inefficient for large datasets or complex models.
  - Research direction: How can we scale inference to large populations or networks?

SUMMARY

- The SIR and SIS models are a powerful starting point.
- Real epidemics require more complex models.
- Many open problems remain in modelling, inference, and data integration.
- You now have the tools to explore these questions!

THANK YOU!!!

ANY QUESTIONS?