

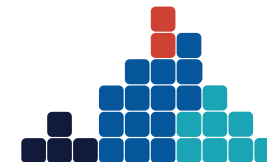
Infectious disease modelling

Quantifying the superspreading potential

Prof. Sukhyun Ryu



1 October 2025

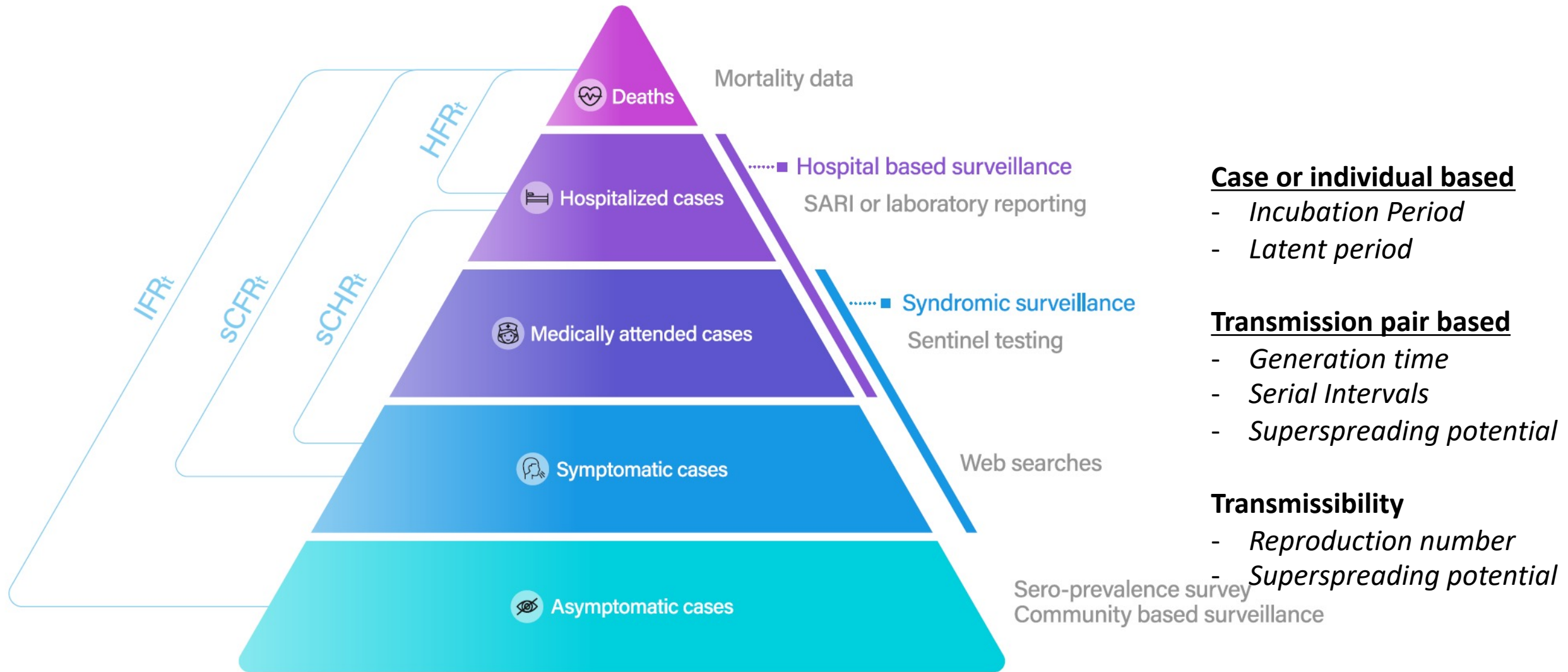


ONE HEALTH Lab
with Computational Epidemiology

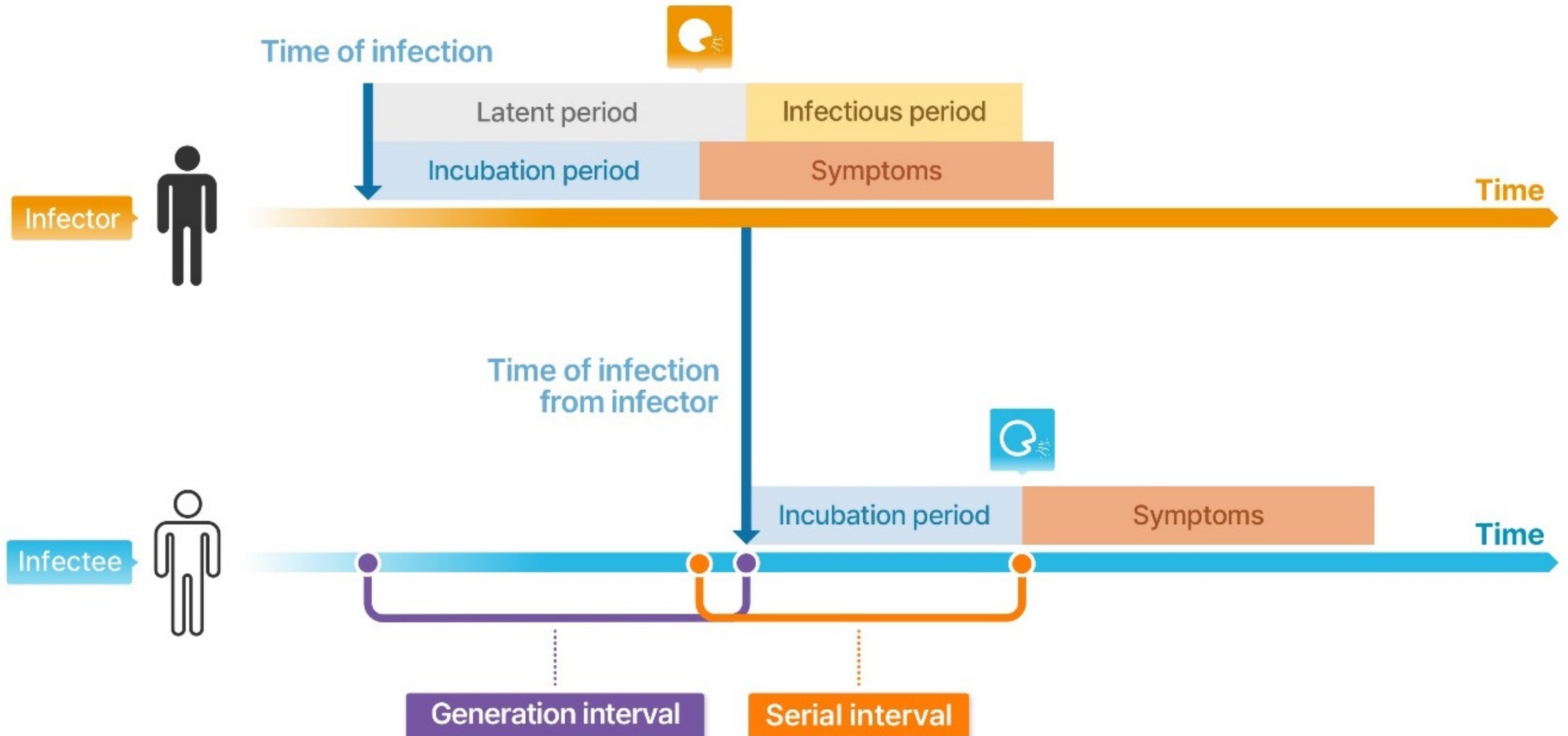
After this course,

- In community epidemics, superspreading can be quantified and described.

Epidemiological parameters

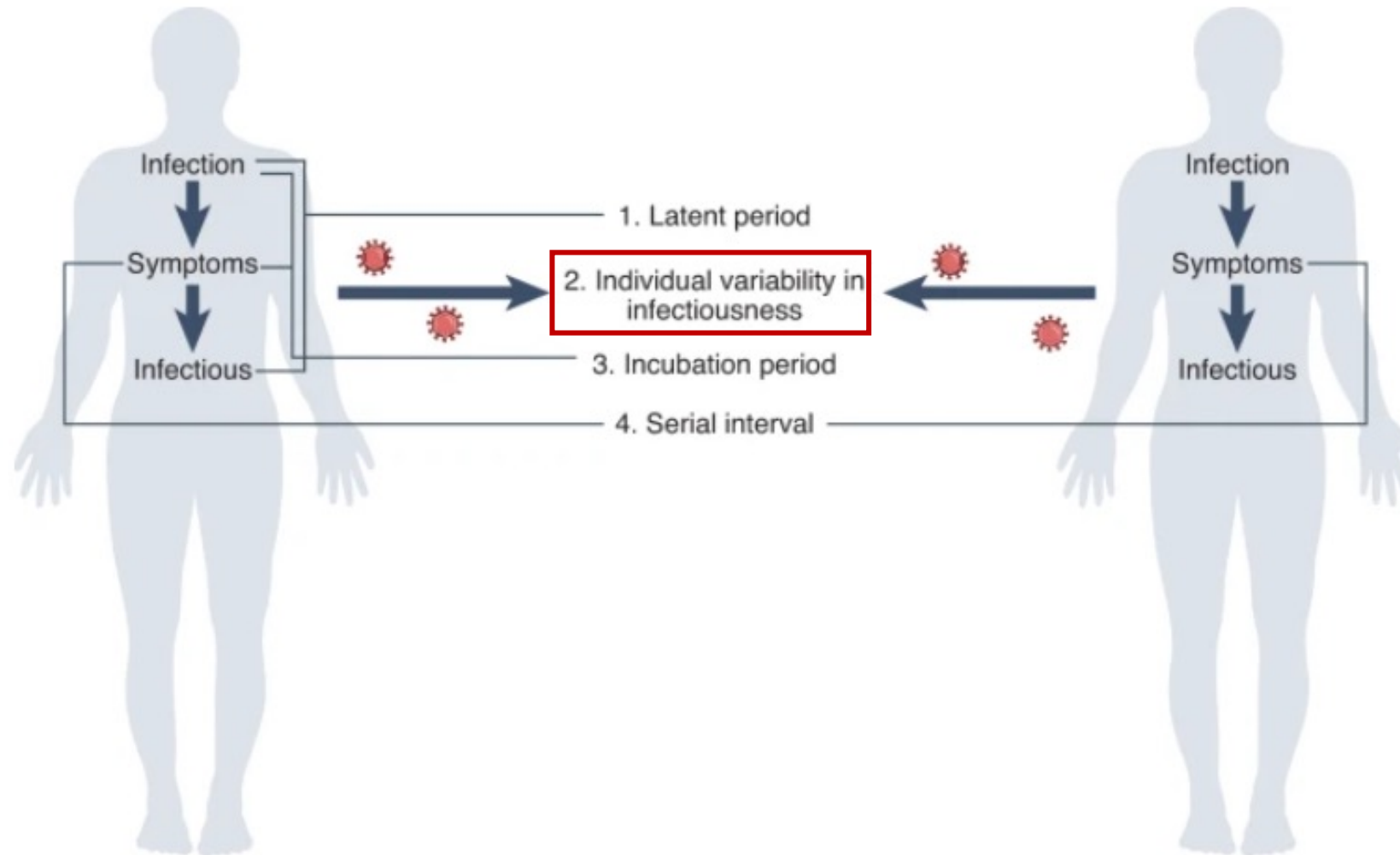


Epidemiological parameters



Epidemiological parameters

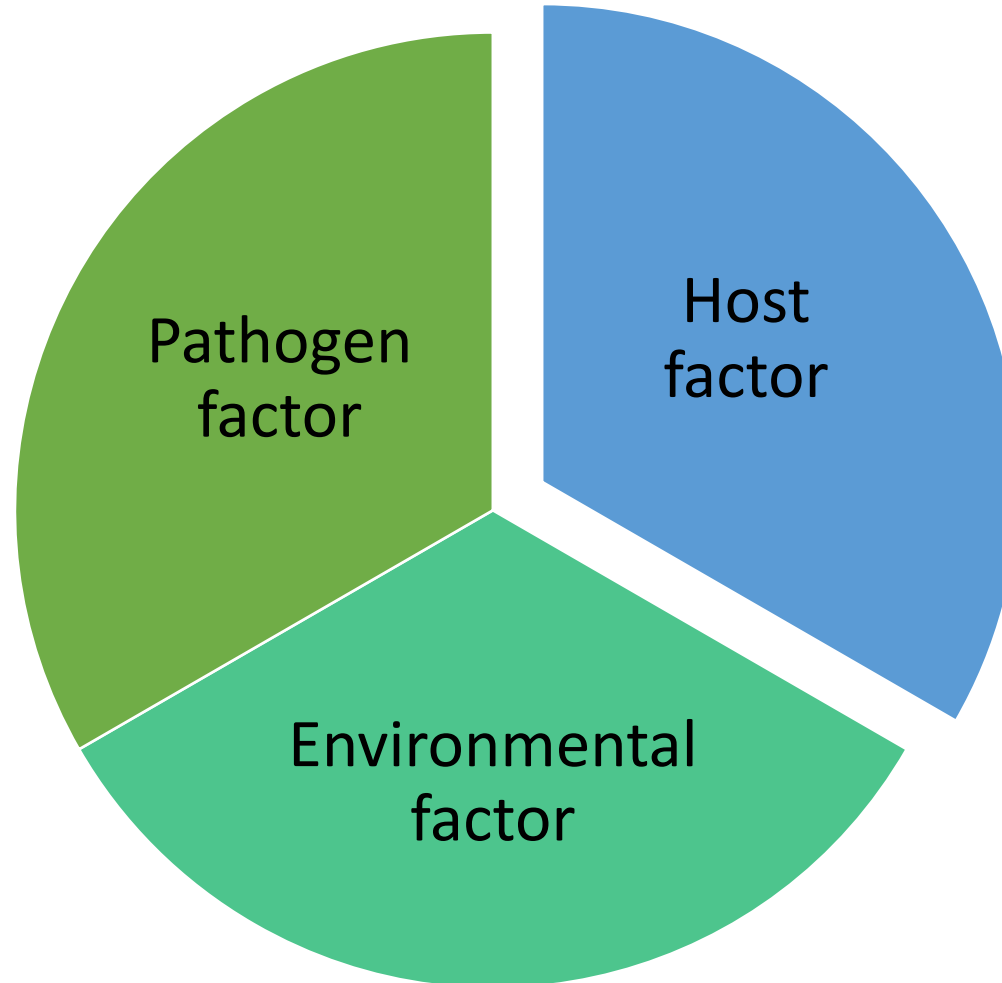
Fig. 1: Epidemiological parameters of SARS-CoV-2 transmission.



Epidemiological parameters

Individual variability of infectiousness

- Virulence
- Co-infection
with another pathogen
- Crowding
- Inter-hospital transfer



- Physiological factors
- Behavioural factors
- Immunological factors

Contents

1. Measures for transmissibility

2. Basic reproduction number

3. Probability distribution

4. Superspreading potential

5. Practice

Measures for transmissibility

- **Infection attack rate**

- The proportion of a population (subgroup) infected over the course of an epidemic

$$\text{attack rate} = \frac{\text{number of new cases in the population at risk}}{\text{number of persons at risk in the population}}$$

- **Secondary (infection) attack rate**

- The proportion of individuals infected in a semi-closed setting (e.g., households) in an outbreak caused by an index case (ideally accounting for pre-existing immunity)

Measures for transmissibility

- **Basic reproductive number (R_0)**
 - The average number of secondary cases generated by an index case when an epidemic begins in a completely susceptible population
- **Effective reproduction number (R_t)**
- **Parameter for superspreading potential ?**

Measures for transmissibility

- **Basic reproductive number (R_0)**
 - The average number of secondary cases generated by an index case when an epidemic begins in a completely susceptible population
- **Effective reproduction number (R_t)**
- **Overdispersion parameter (K)**

Contents

1. Measures for transmissibility

2. Basic reproduction number

3. Probability distribution

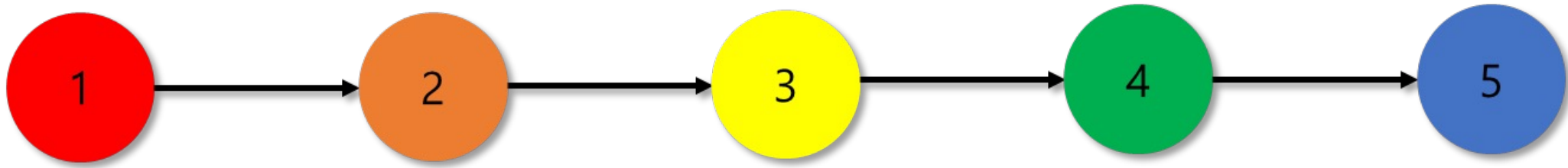
4. Superspreading potential

5. Practice

Basic reproduction number

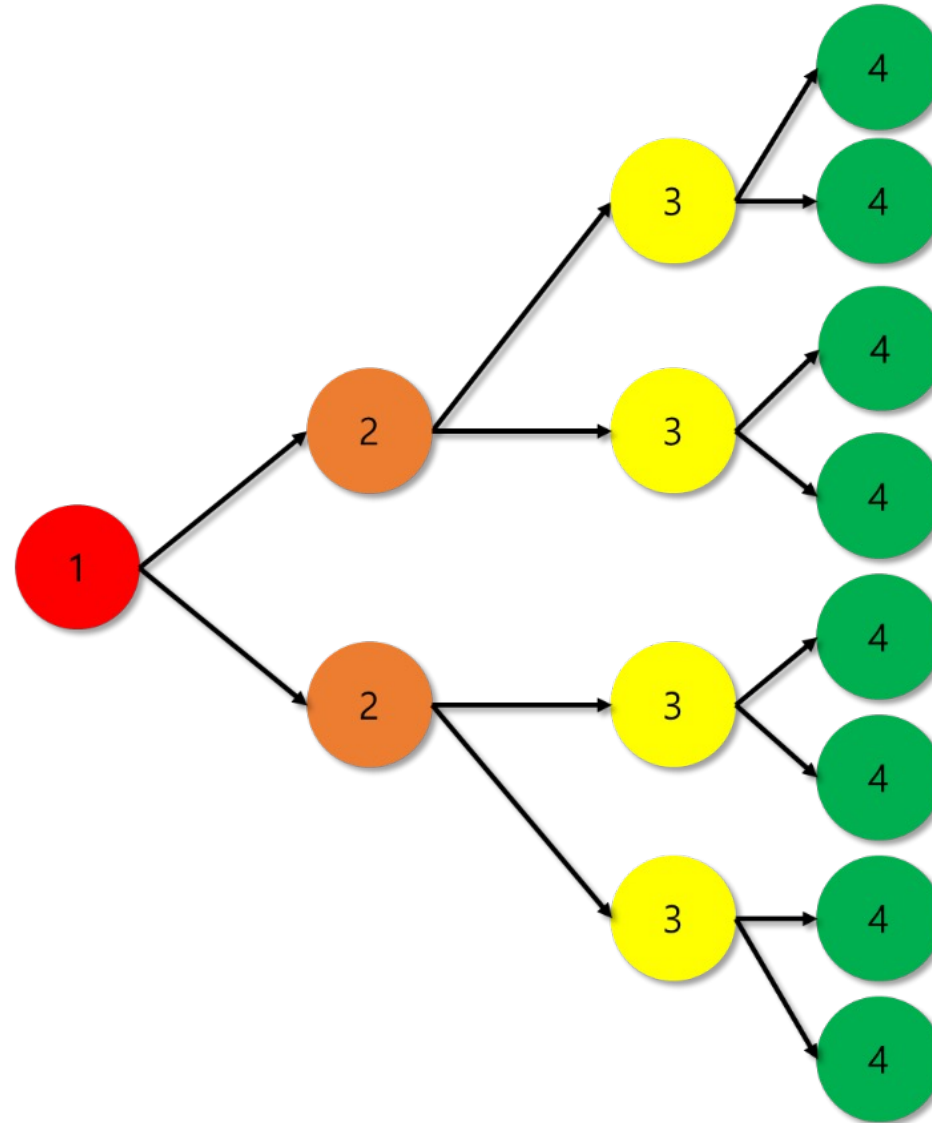
- **Basic reproductive number**

- The average number of secondary cases generated by an index case when an epidemic begins in a completely susceptible population.



$$R_0 = 1$$

Basic reproduction number



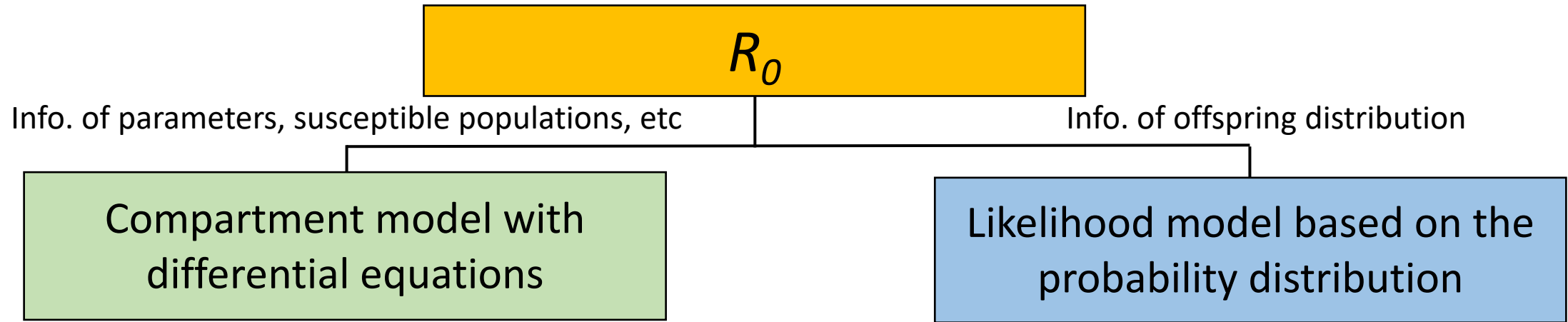
$$R_0 > 1$$

$$\text{Ex) } R_0 = 2$$

Basic reproduction number

- $R_0 < 1 \rightarrow$ the disease will eventually disappear
- $R_0 = 1 \rightarrow$ the disease will become endemic
- $R_0 > 1 \rightarrow$ there will be endemic

Basic reproduction number



미분방정식 (Differential equation) 을 이용한 SIR 모델

The diagram shows the SIR model flow: S (Susceptible) → I (Infected) → R (Recovered). Below S, a box indicates '새로 감염된 사람 수' (Number of newly infected people) with the formula $\beta \left(\frac{S}{N}\right) I$. Below I, a box indicates '감염되었다가 회복된 사람 수' (Number of people who were infected and then recovered) with the formula γI .

$$R_0 = \frac{\beta S(0)}{\gamma N} \approx R_0 = \frac{\beta}{\gamma}$$

$$R = -\frac{\log\left(\frac{1-AR}{S_0}\right)}{AR - (1 - S_0)}$$

$$Y \sim \text{Poisson}(\lambda)$$

Basic reproduction number

Poisson distribution

[Article](#) [Talk](#)

From Wikipedia, the free encyclopedia

In [probability theory](#) and [statistics](#), the **Poisson distribution** (/ˈpwɑːsɒn/) is a [discrete probability distribution](#) that expresses the probability of a given number of [events](#) occurring in a fixed interval of time if these events occur with a known constant mean rate and [independently](#) of the time since the last event.^[1] It can



Siméon Denis Poisson
(1781-1840)

Contents

1. Measures for transmissibility

2. Basic reproduction number

3. Probability distribution

4. Superspreading potential

5. Practice

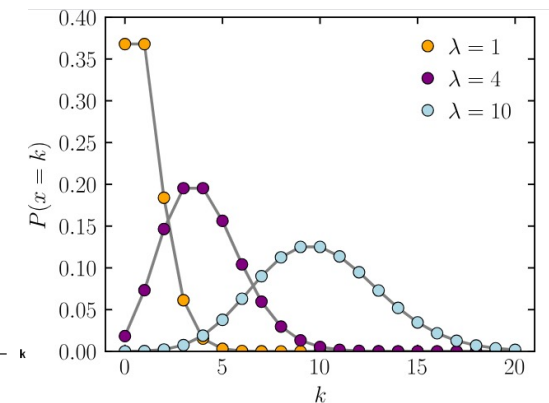
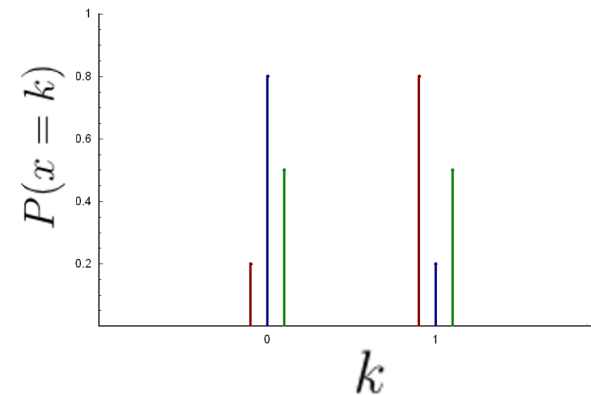
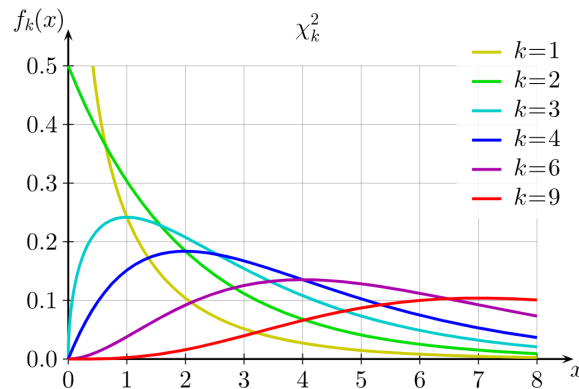
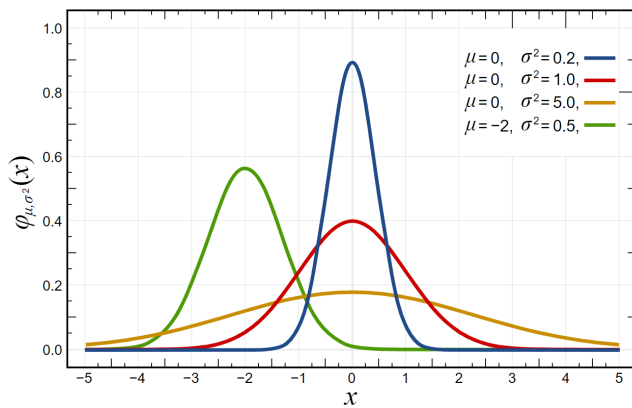
Probability distribution

Probability distribution

확률변수가 특정값을 가질 확률을 나타내는 함수

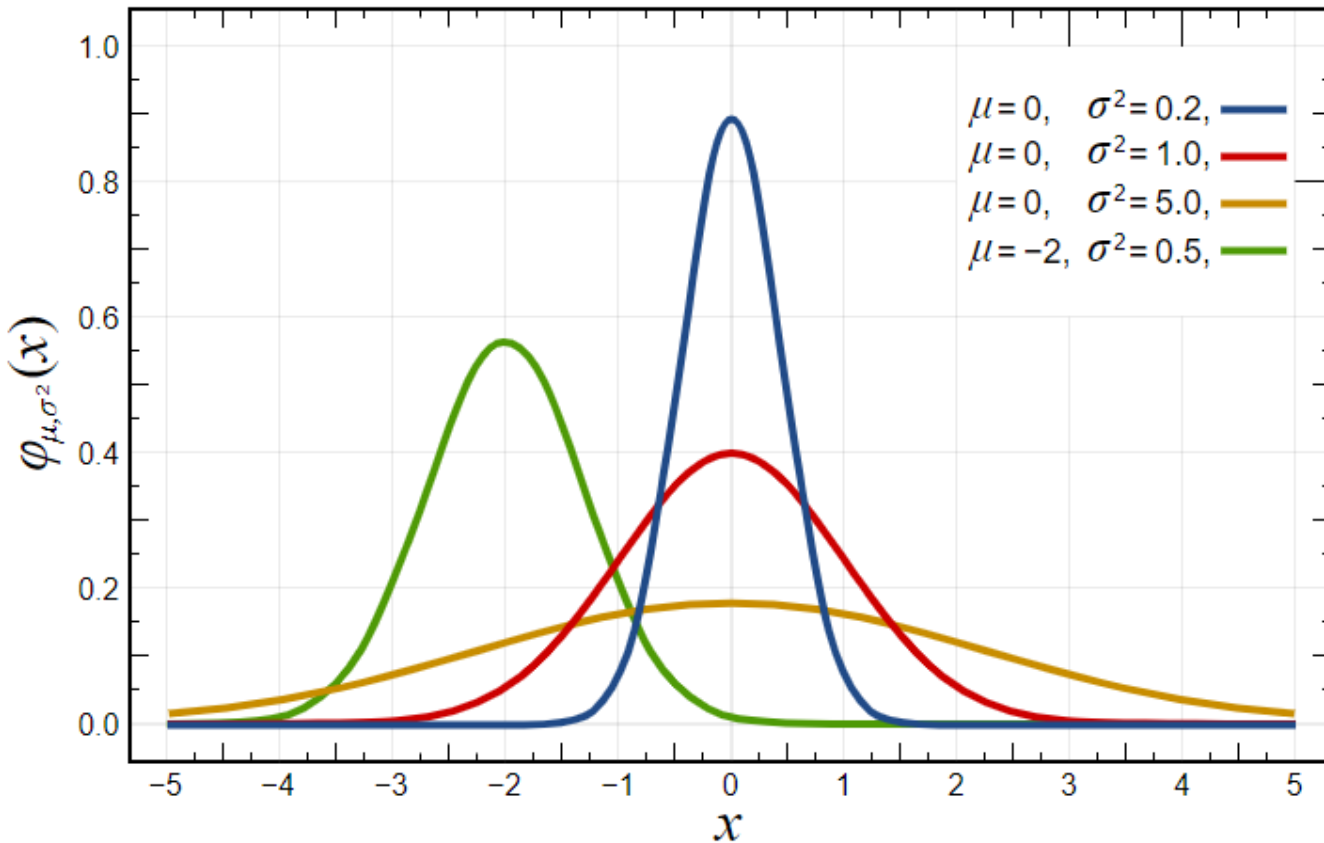
Continuous probability distribution

Discrete probability distribution



Probability distribution

Normal distribution



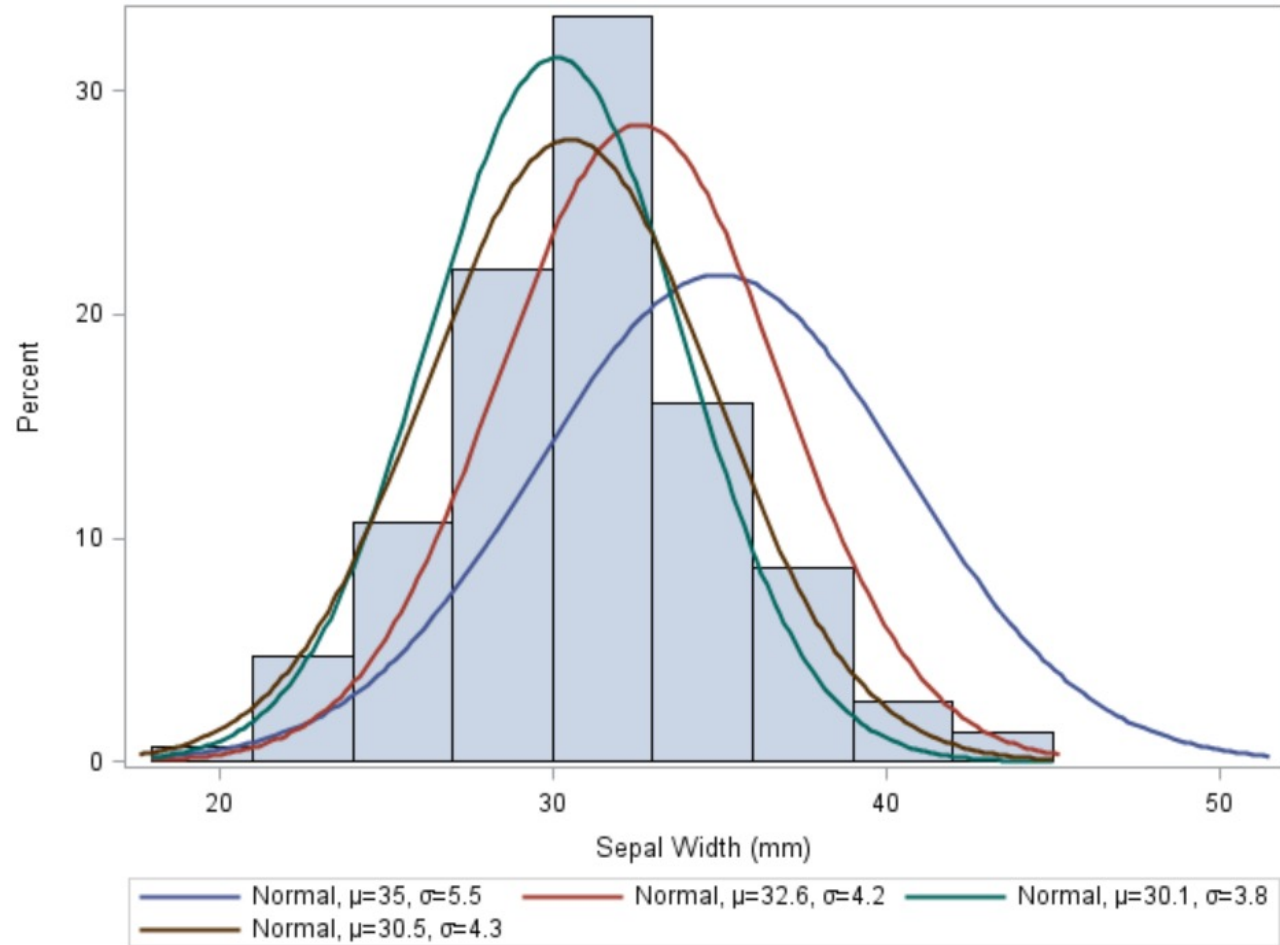
- Continuous probability distribution
- General form of its probability density function(PDF):

$$f(x_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Mean: μ

Standard deviation: σ

Probability distribution



- Maximum likelihood estimation (MLE)
- Estimate the most appropriate parameters among the assumed probability distribution forms.

Probability distribution

$$Y \sim \text{Poisson}(\lambda)$$

Y : Expected value of a specific event occurring (*i. e.*, *Infected Cases*)

λ : Mean no. of times a specific event occurs within a time unit (*i. e.*, R_0)

➡ *No. secondary infected cases from a primary case* $\sim \text{Poisson}(R_0)$

Probability distribution

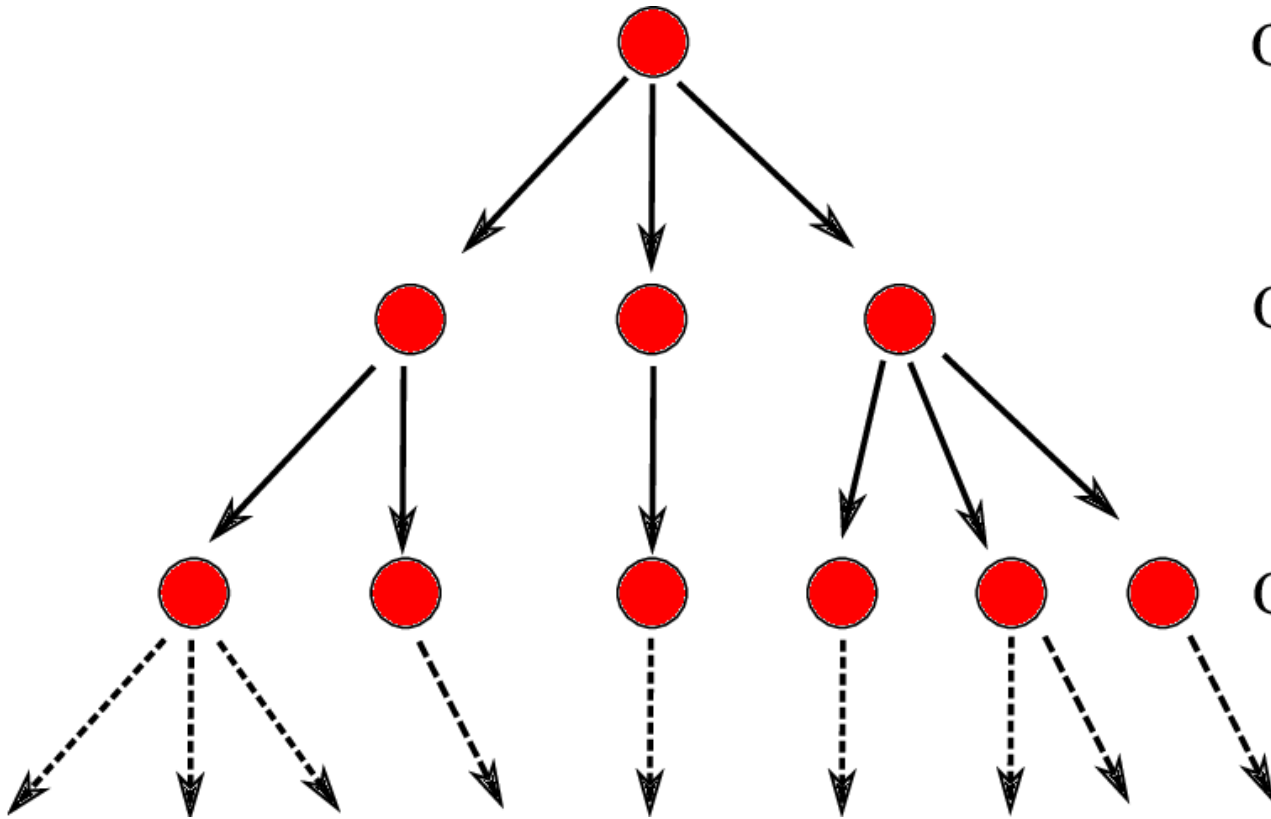
Branching Processes

Generation 0

Generation 1

Generation 2

$$Y \sim \text{Poisson}(\lambda)$$



Probability distribution

$$f(k; \lambda) \sim \text{Pois}(\lambda)$$

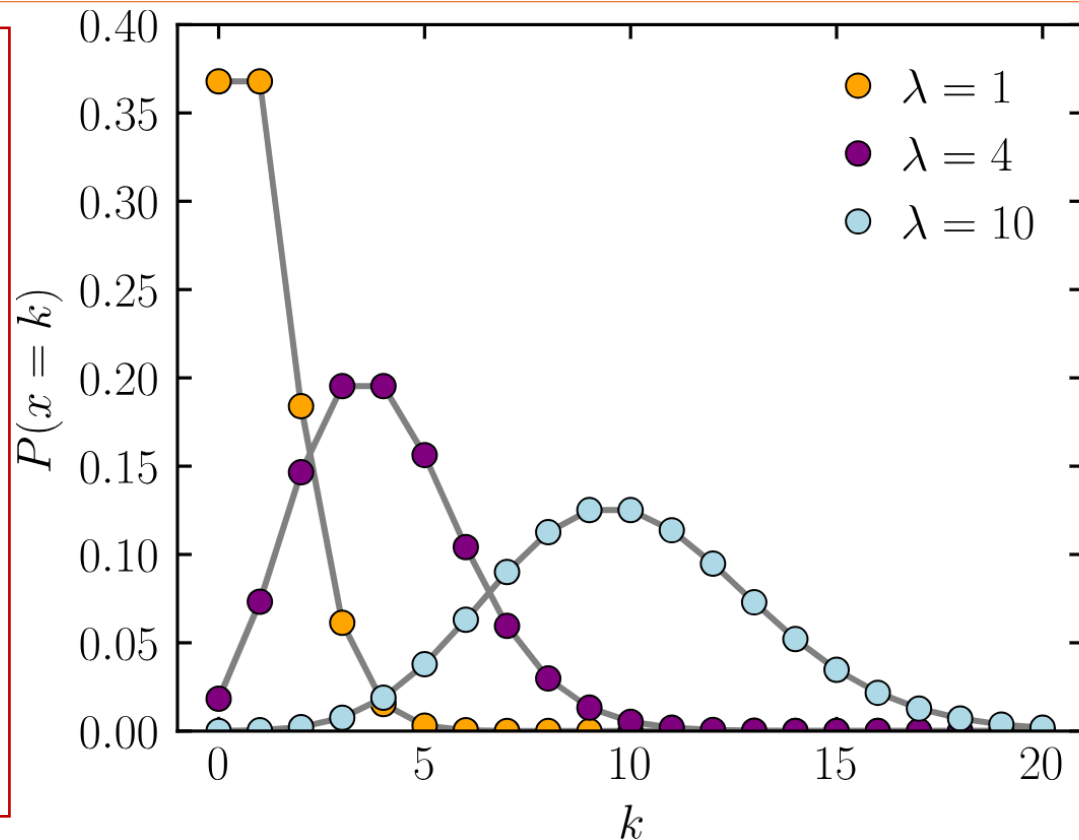
- Probability density function (PDF):

$$f(k; \lambda) = \underline{\text{Pr}(X=k)} = \frac{\lambda^{\underline{k}} e^{-\lambda}}{\underline{k!}},$$

where

- k is the number of occurrences ($k = 0, 1, 2, \dots$)
- e is Euler's number ($e = 2.71828 \dots$)
- $!$ is the factorial function.

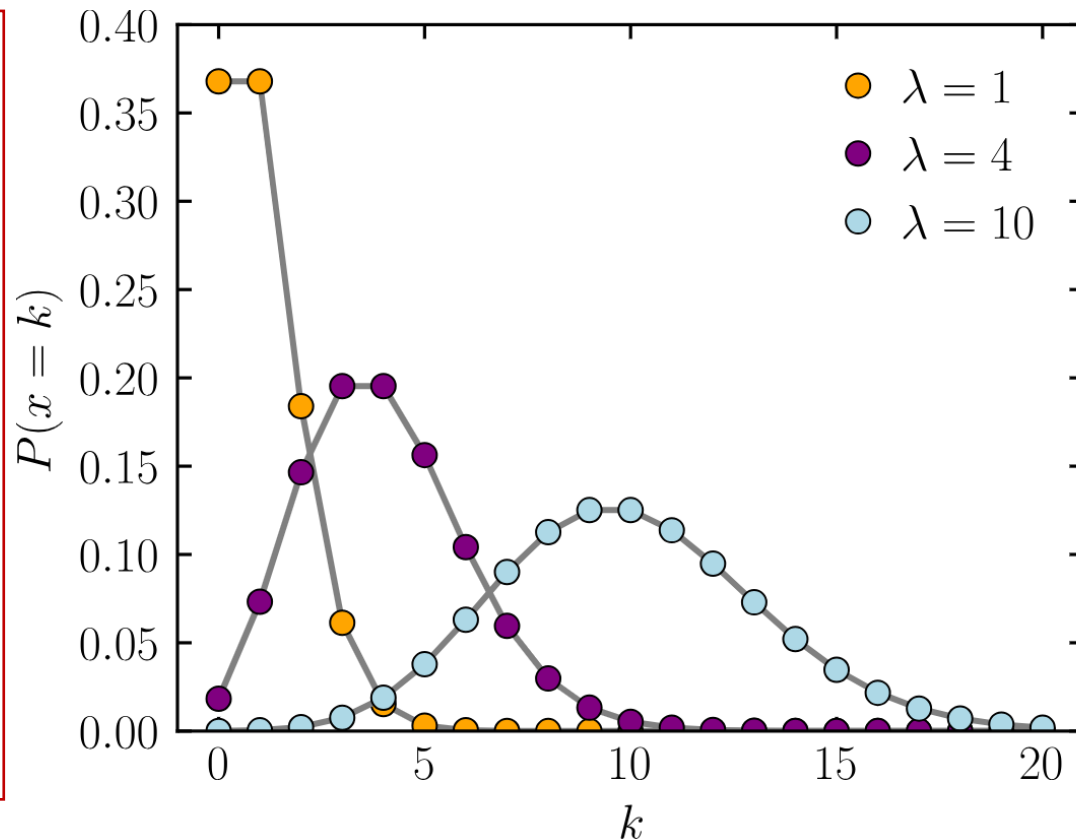
- $E(X) = \lambda$
- $\text{Var}(X) = \lambda$



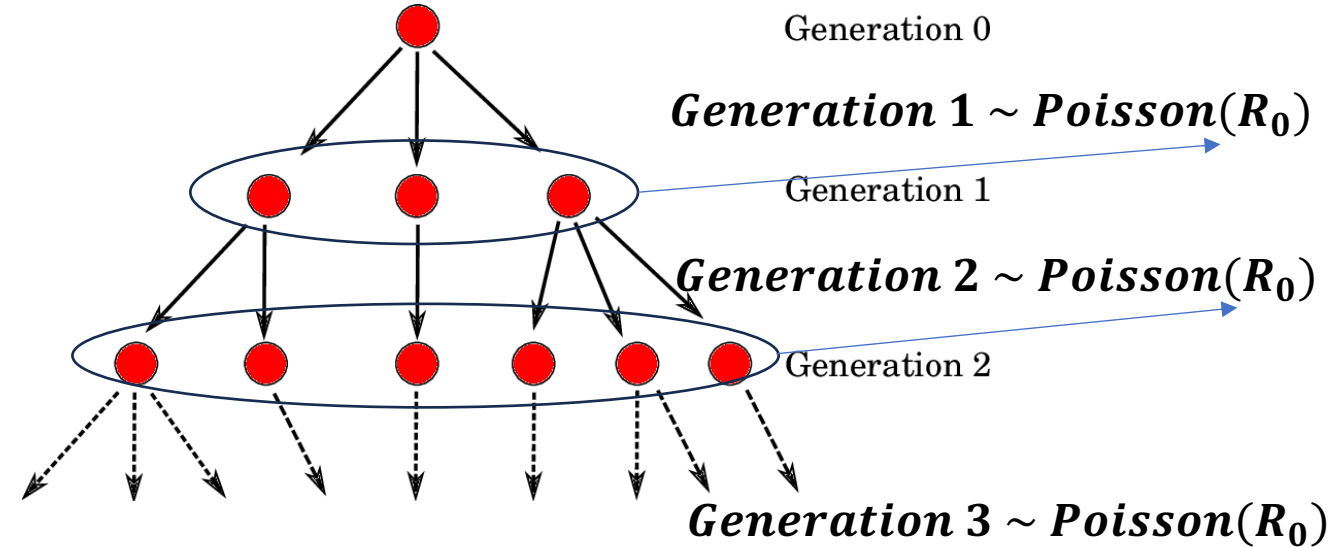
No. of occurrences

Probability distribution

$$f(k; \lambda) \sim \text{Pois}(\lambda)$$



$$Y \sim \text{Poisson}(\lambda)$$



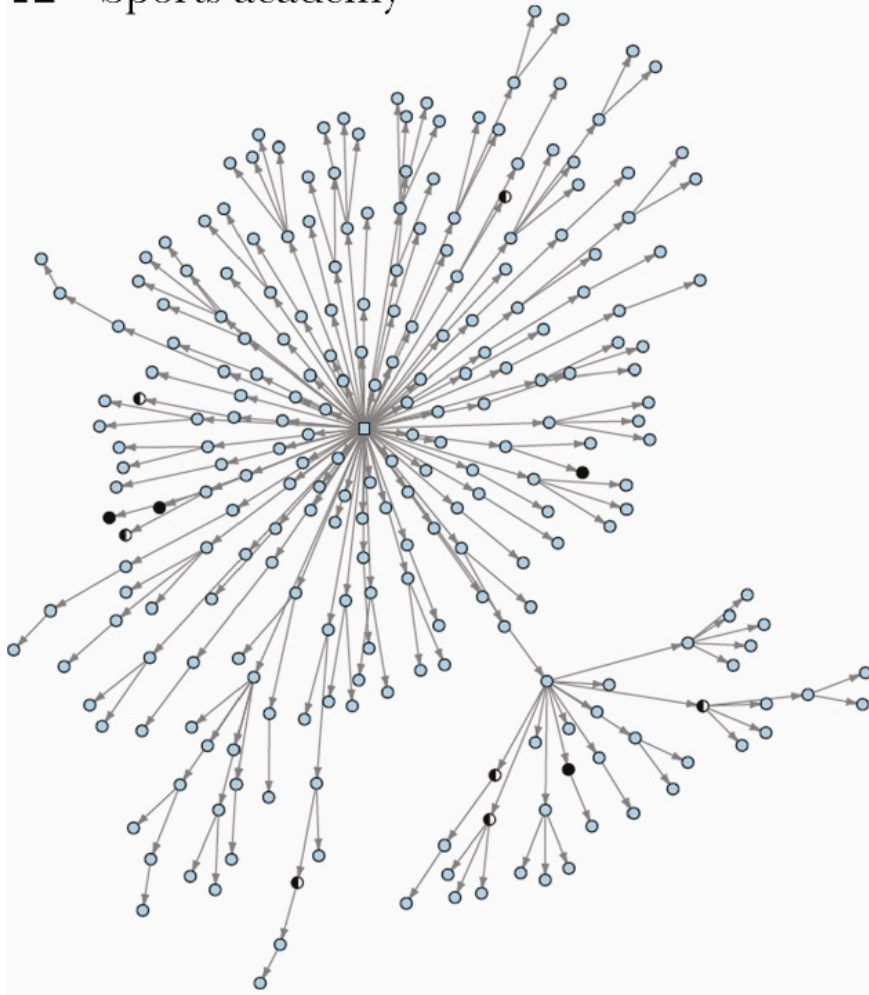
- $E(X) = \lambda$
- $Var(X) = \lambda$



Probability distribution

More accurate model than the Poisson distribution by allowing the mean and variance to be different is needed.

A Sports academy



B Karaoke center



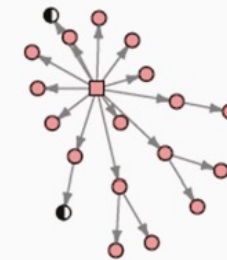
C High school



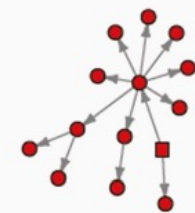
D Other settings
Welfare facility for disabled



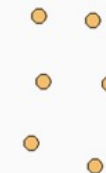
Reading room



Acting school



Sporadic cases



Probability distribution

$$f(k; \lambda) \sim \text{Pois}(\lambda)$$



Siméon Denis Poisson (1781-1840)

https://en.wikipedia.org/wiki/Probability_distribution

$$Y \sim \text{NB}(r, p)$$



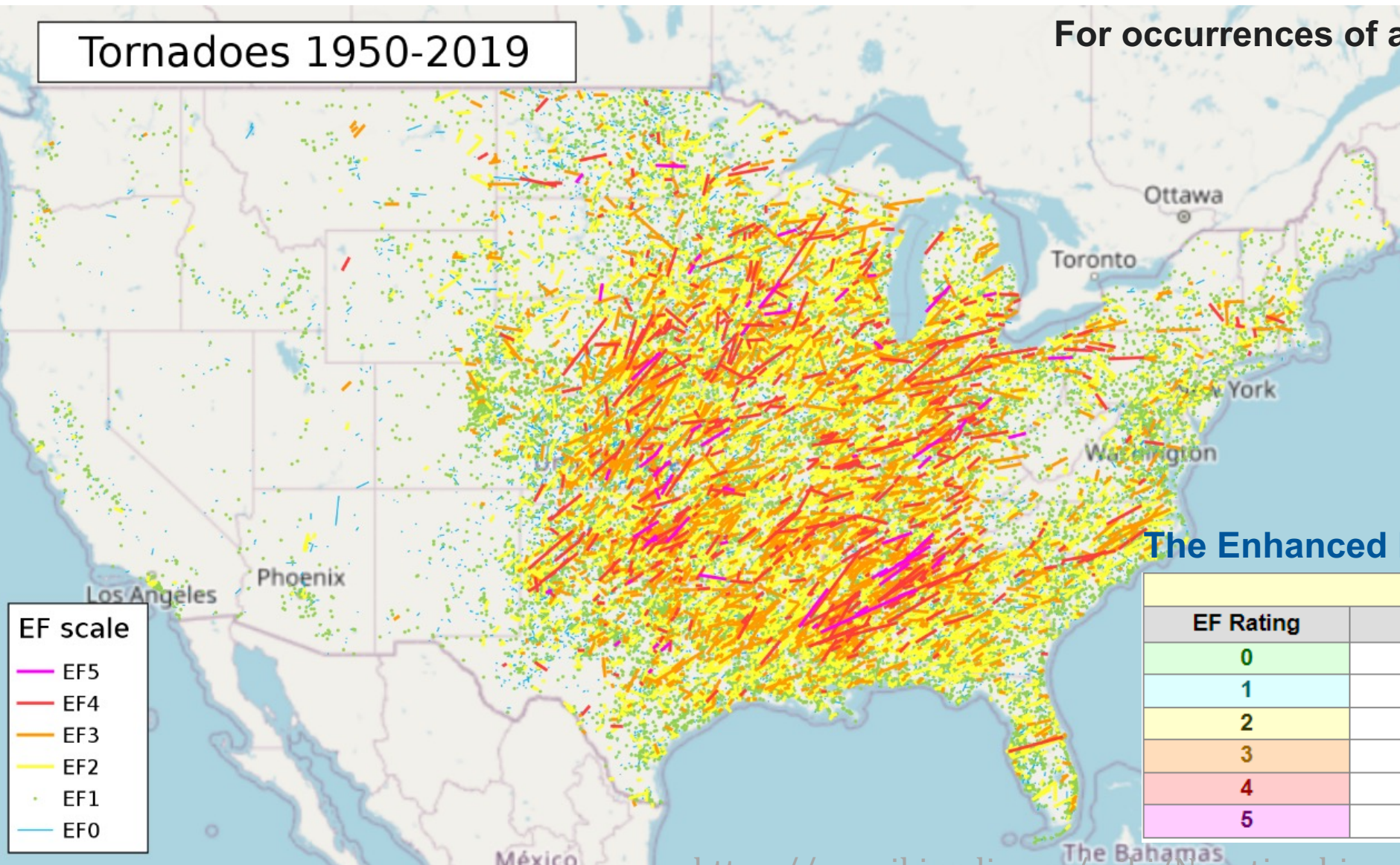
Blaise Pascal (1623-1662)

https://en.wikipedia.org/wiki/Blaise_Pascal

Probability distribution

Tornadoes 1950-2019

For occurrences of associated discrete events



The Enhanced Fujita Scale (EF Scale)

EF SCALE	
EF Rating	3 Second Gust (mph)
0	65-85
1	86-110
2	111-135
3	136-165
4	166-200
5	Over 200

Contents

1. Measures for transmissibility

2. Basic reproduction number

3. Probability distribution

4. Superspreading potential

5. Practice

Superspreading potential

☰ Superspreading event

Article [Talk](#)

From Wikipedia, the free encyclopedia

A **superspreading event** (**SSEV**) is an event in which an [infectious disease](#) is spread much more than usual, while an unusually contagious organism [infected](#) with a [disease](#) is known as a **superspreader**. In the [context of a human-borne illness](#), a superspreader is an individual who is more likely to infect others, compared with a typical infected person. Such superspreaders are of particular concern in [epidemiology](#).

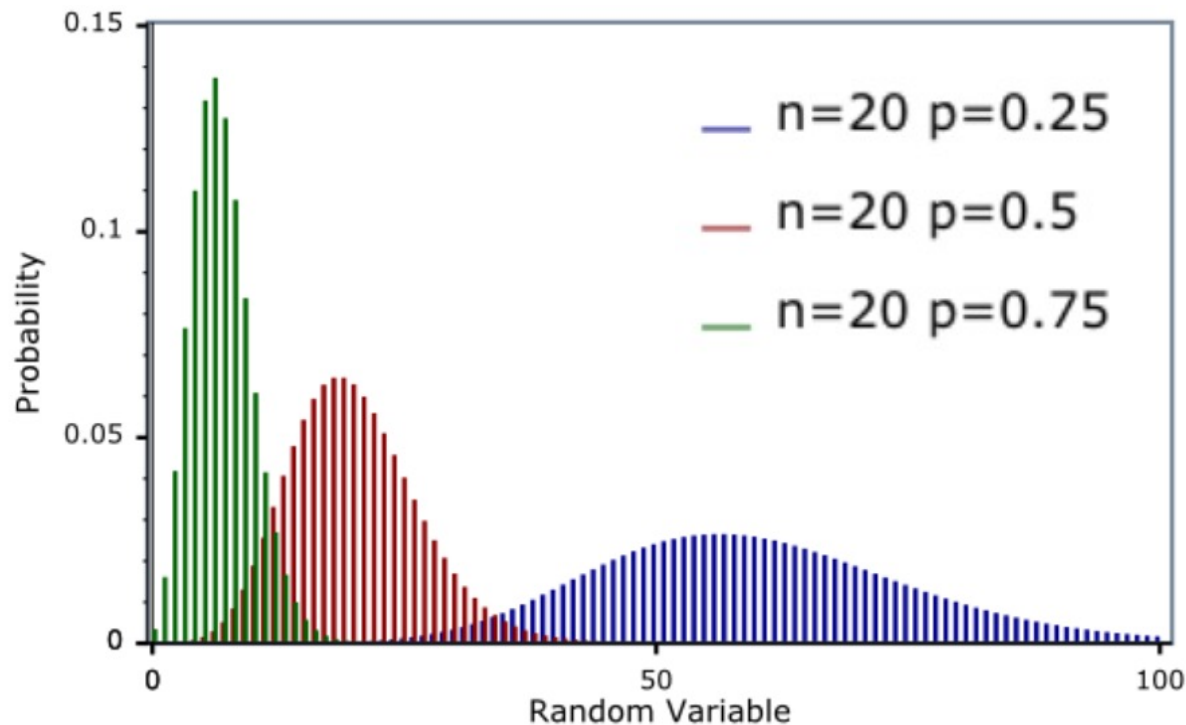
Superspreading potential

- Negative binomial distribution
 - Distribution reflecting variation in individual infectiousness in the Poisson distribution
- Parameters
 - Mean (R_0)
 - Variance ($kappa$) : Variation in individual infectiousness

Smaller K , $\rightarrow$ Greater the variation (i.e., Highly likelihood SSE)

Superspreading potential

$$Y \sim NB(r, p)$$



Negative binomial distribution

- Probability density function

$$NB(r, p) = \binom{k + r - 1}{k} \cdot (1 - p)^k pr$$

- r No. successes by the trial is stopped
- p Probability of success in each trial
- k No. of failures

- $E(X) = \frac{r(1-p)}{p}$
- $Var(X) = \frac{r(1-p)}{p^2}$

Superspreading potential

$$Y \sim \text{Poisson}(\lambda)$$

➡ *No. secondary infected cases from a primary case* $\sim \text{Poisson}(R_0)$

$$Z \sim \text{Poisson}(R_0)$$

Offspring distribution with mean R_0 : Not applicable when the variance is greater or less than the mean

(Failed to show heterogeneity)

➡ Alternative : Negative binomial distribution (Poisson–Gamma mixture)

Superspreading potential

$$Z \sim \text{Poisson}(R_0)$$



Assuming *Gamma*(Γ) distribution

$$R_0 \sim \Gamma\left(\alpha, \frac{p}{1-p}\right)$$

α : shape parameter
p/1-p : scale parameter
p : probability



**Negative binomial distribution
(Poisson–Gamma mixture)**

$$P(Z = z) = \frac{\Gamma(k + z)}{z! \Gamma(k)} \left(\frac{k}{k + R_0}\right)^k \left(\frac{R_0}{k + R_0}\right)^z$$

$$Z \sim \text{NB}(R_0, k)$$

Gamma distribution is used in statistics to model a wide range of processes

Superspreading potential

Gamma-Poisson mixture

$$\begin{aligned} & \int_0^{\infty} f_{Poisson}(\lambda)(z) \times f_{Gamma}(\alpha, \frac{p}{1-p})(\lambda) d\lambda \\ &= \int_0^{\infty} \frac{\lambda^z}{z!} e^{-\lambda} \times \frac{1}{\Gamma(\alpha)} \left(\frac{p}{1-p}\lambda\right)^{\alpha-1} e^{-\frac{p}{1-p}\lambda} \left(\frac{p}{1-p} d\lambda\right) \\ &= \left(\frac{p}{1-p}\lambda\right)^{\alpha} \frac{1}{z!\Gamma(\alpha)} \int_0^{\infty} \lambda^{\alpha+z-1} e^{-\lambda \frac{p+1-p}{1-p}} d\lambda \\ &= \left(\frac{p}{1-p}\lambda\right)^{\alpha} \frac{1}{z!\Gamma(\alpha)} \Gamma(\alpha+z)(1-p)^{z+\alpha} \int_0^{\infty} f_{Gamma}(\alpha, \frac{p}{1-p})(\lambda) d\lambda \\ &= \boxed{\frac{\Gamma(\alpha+z)}{z!\Gamma(\alpha)}} (1-p)^z p^{\alpha} \\ & \quad \binom{z+r-1}{z} \end{aligned}$$

Superspreading potential

Gamma–Poisson Mixture

$$f(z; \alpha, p) = \frac{\Gamma(\alpha+z)}{z! \Gamma(\alpha)} (1-p)^z p^\alpha$$

$$E(Z) = \mu$$

$$* \quad \alpha = k, \quad p = \frac{k}{k+\mu} \quad (\mu = R_0)$$

$$Var(Z) = \mu + \frac{\mu^2}{k} = R_0 + \frac{R_0^2}{k}$$

$$P(Z = z) = \frac{\Gamma(k+z)}{z! \Gamma(k)} \left(\frac{k}{k+R_0} \right)^k \left(\frac{R_0}{k+R_0} \right)^z$$

Lower the k , larger the variance
 $\Rightarrow$ Increase of *heterogeneity*

$$Z \sim NB(R_0, k)$$

k : shape parameter and dispersion parameter

R_0 : basic reproduction number

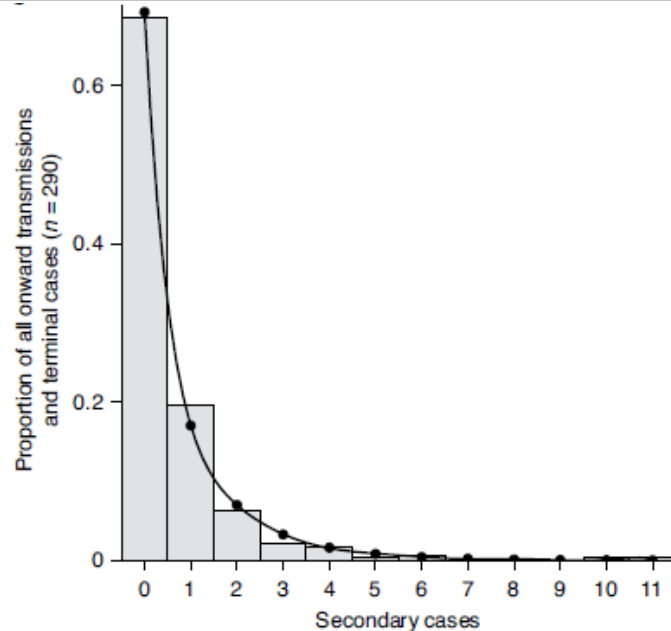
Superspreading potential

Superspreading potential

Based on data type

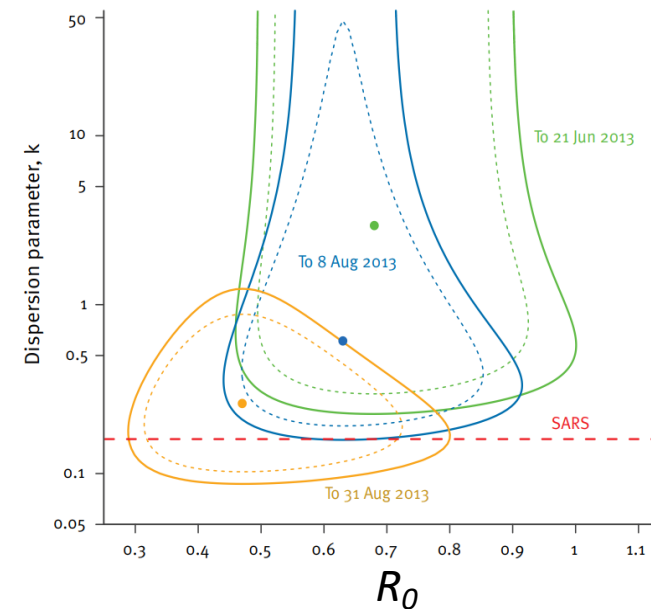
1

Offspring distribution fitted with negative binomial distribution

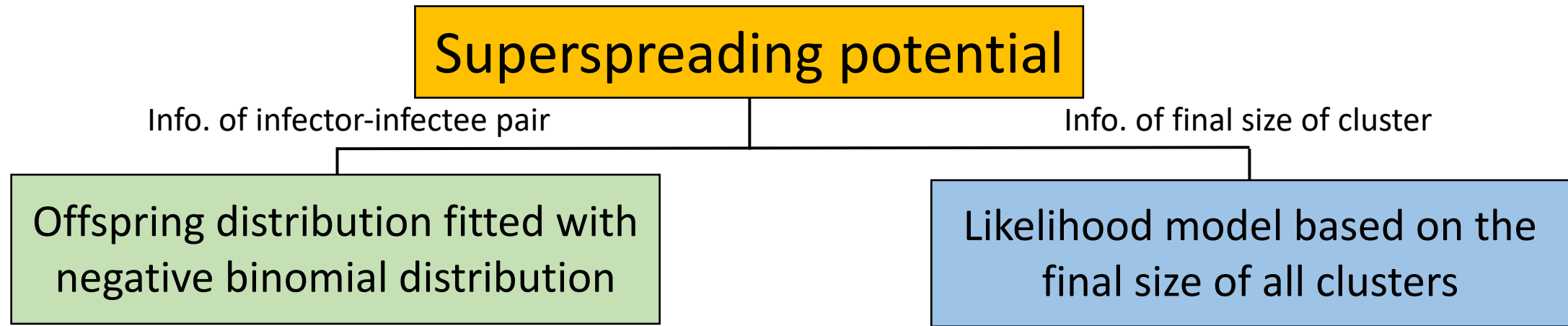


2

Likelihood model based on the final size of all clusters

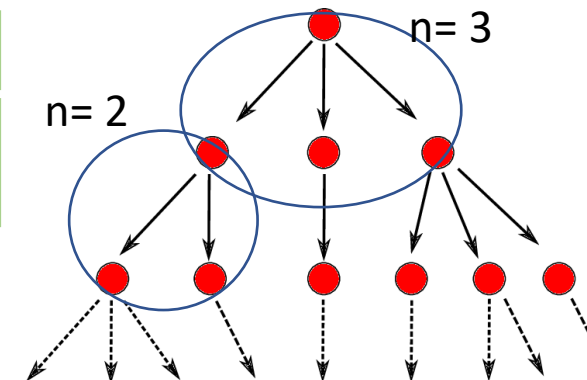


Superspreading potential



No. of secondary cases (i.e., infectee) from each infectors

Infector	A	B	C	D	E	F	G	H
No. offspring	11	10	6	6	5	4	4	...



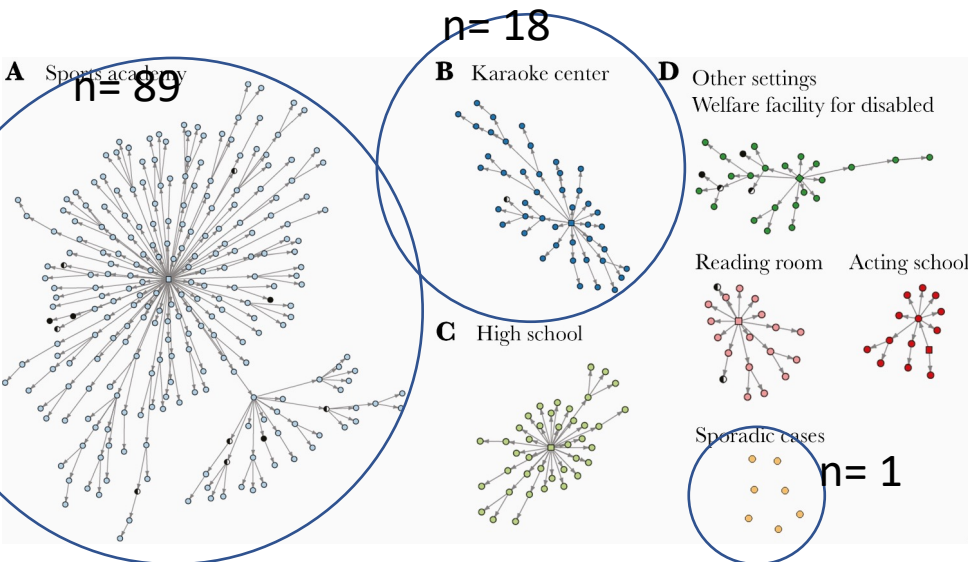
Superspreading potential

Superspreading potential

Offspring distribution fitted with negative binomial distribution

Likelihood model based on the final size of all clusters

Final No. of cluster size →

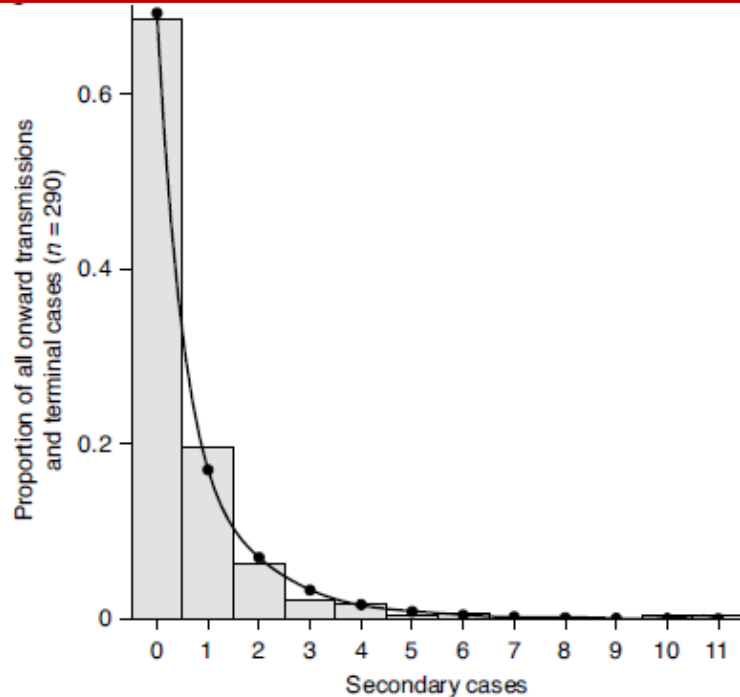


Cluster size	Number of MERS clusters of a given size		
	Breban et al. [6] ^c	Cauchemez et al. [7]	Poletto et al. [5] ^c
1	11	27	42
2	2	2	7
3	3	4	2
4	1	3	–
5	2	2	2
7	–	1	–
10	–	–	1
13	–	1	–
22	–	–	1
24	1	–	–
26			

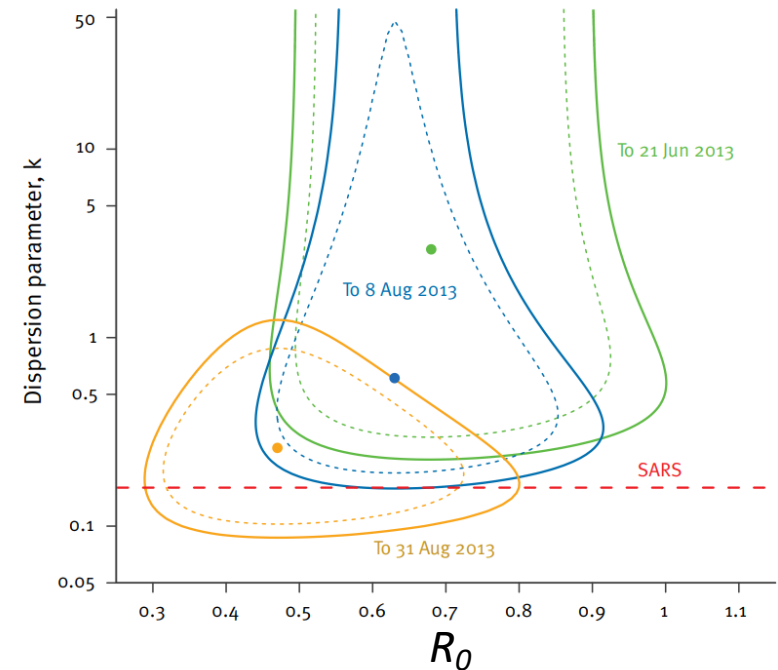
Superspreading events

Superspreading potential

Offspring distribution fitted with negative binomial distribution



Likelihood model based on the final size of all clusters



Contents

1. Measures for transmissibility

2. Basic reproduction number

3. Probability distribution

4. Superspreading potential

5. Practice

R package for Fitting Distributions

Practice

fitdistr

Maximum-likelihood Fitting of Univariate Distributions

Description

Maximum-likelihood fitting of univariate distributions, allowing parameters to be held fixed if desired.

Usage

```
fitdistr(x, densfun, start, ...)
```

Arguments

x	A numeric vector of length at least one containing only finite values.
densfun	Either a character string or a function returning a density evaluated at its first argument. Distributions "beta", "cauchy", "chi-squared", "exponential", "gamma", "geometric", "log-normal", "lognormal", "logistic", <u>"negative binomial"</u> , "normal", "Poisson", "t" and "weibull" are recognised, case being ignored.
start	A named list giving the parameters to be optimized with initial values. This can be omitted for some of the named distributions and must be for others (see Details).
...	Additional parameters, either for densfun or for optim. In particular, it can be used to specify bounds via lower or upper or both. If arguments of densfun (or the density function corresponding to a character-string specification) are included they will be held fixed.

library(MASS)

"fitdistr" function estimates the parameters ' μ ' and 'size' of the log normal distribution

Practice

- ① Add data
- ② Plot histogram
 - for distribution assumption
- ③ Negative binomial distribution fitting by MLE fitting
 - Estimates the R_0 and k
- ④ Plot PDF
- ⑤ 95% confidence interval for R_0 and k

Practice

① Add data

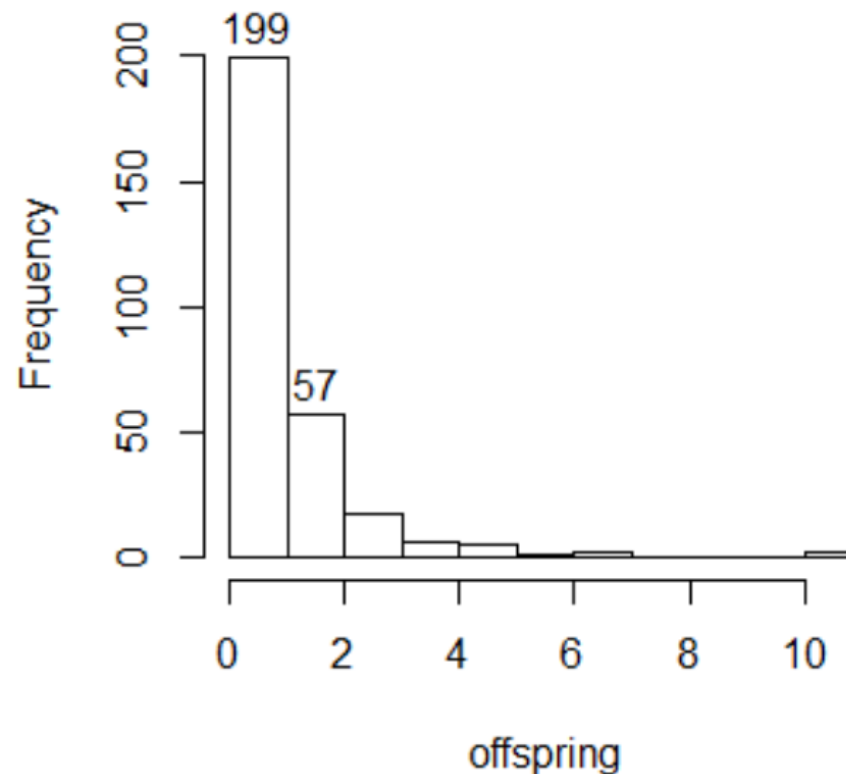
```
1 # offspring distribution
2 offspring <- c(rep(0,199),rep(1,57),rep(2,18),rep(3,6),rep(4,5),5,rep(6,2),10,11)
3 offspring
4 # frequency table
5 table(offspring)
```

② Plot histogram

- for distribution assumption

```
7 # plot histogram
8 hist(offspring)
9 hist(offspring, right = F)
```

생략 또는 right = T: ~이상
Right = F: ~미만



Practice

③ Negative binomial distribution fitting by MLE

- Estimates the R_0 and k

```
11 # fitting negative binomial distribution
12 library(MASS)
13 fit <- fitdistr(offspring, "negative binomial")
14 fit
```

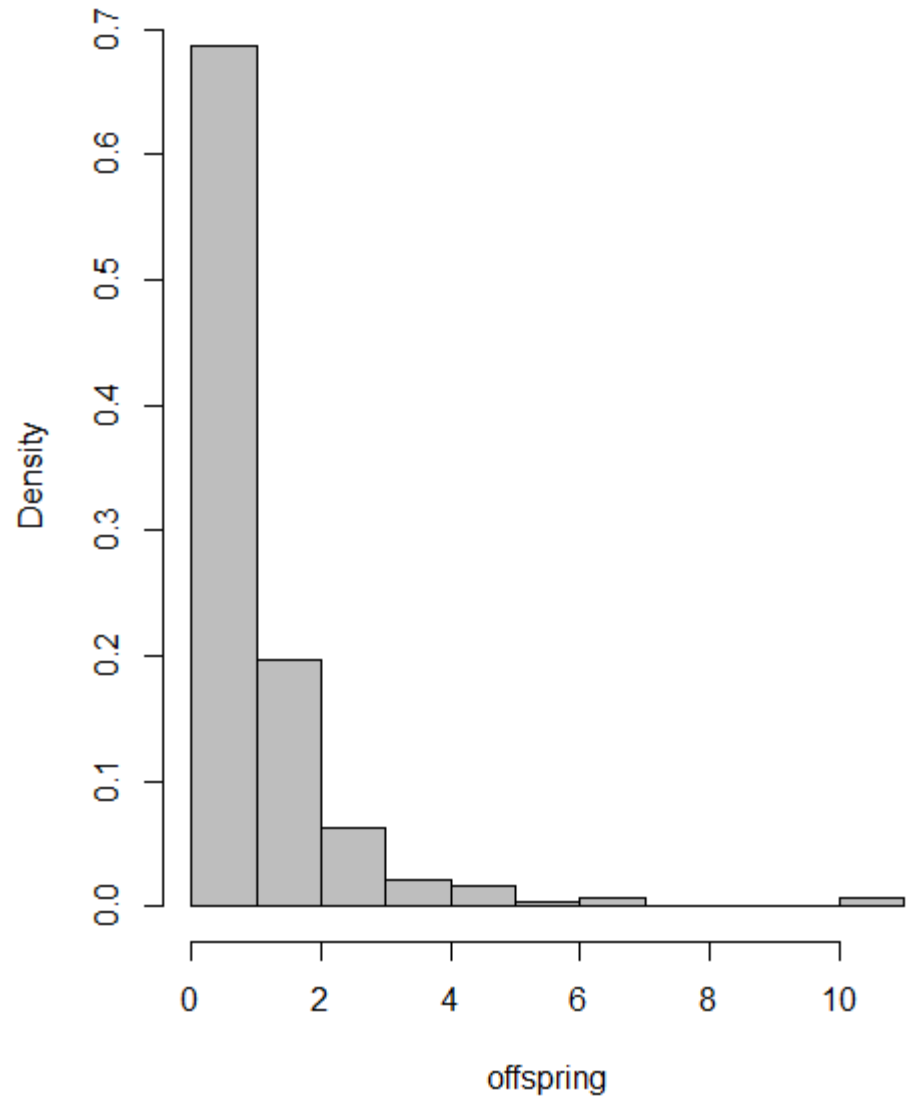
Fitting NB distribution by MLE

```
> fit
       $\hat{k}$  size       $\hat{R}_0$  mu
0.42580272 0.58275962
(0.08797182) (0.06899078)
```

Practice

④ Plot PDF

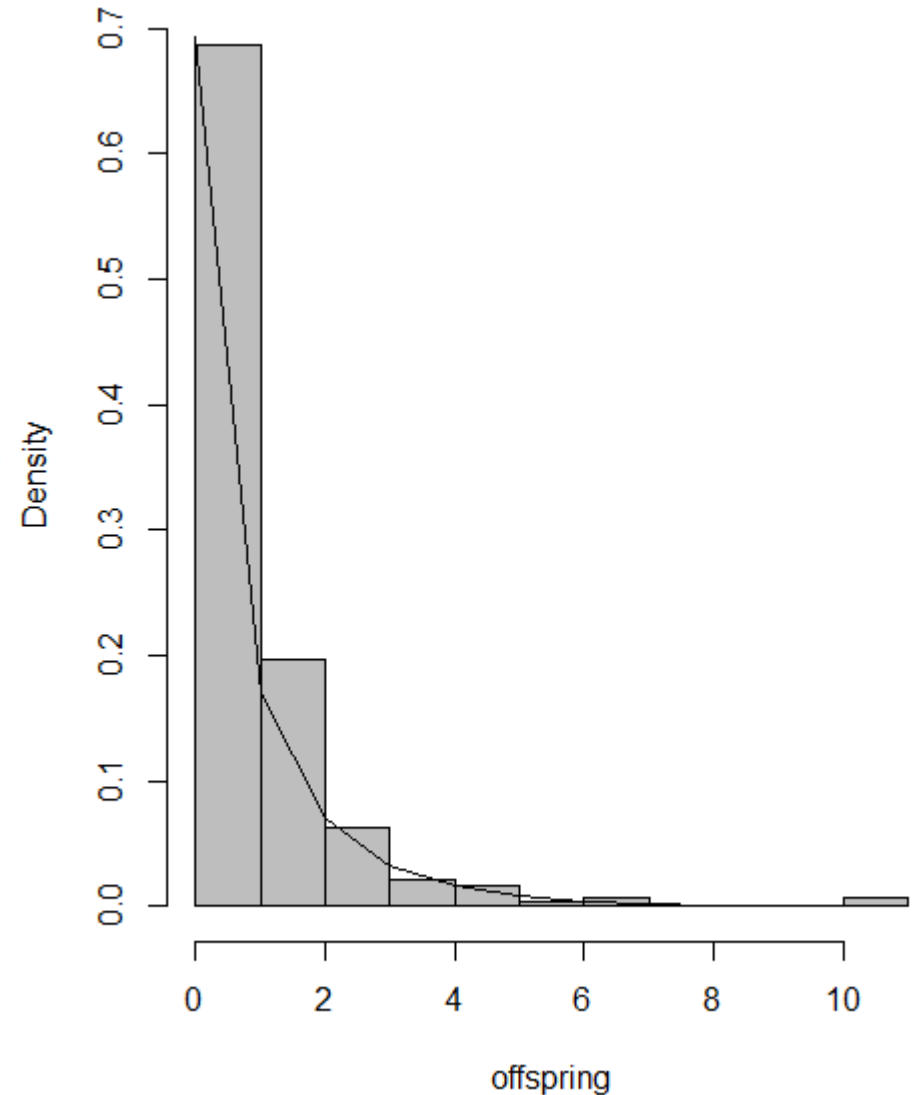
```
16 # plot PDF
17 hist(offspring, prob=T, col="gray",right = F)
```



Practice

④ Plot PDF

```
16 # plot PDF
17 hist(offspring, prob=T, col="gray",right = F)
18 xvalue <- 0:max(offspring)
19 yvalue <- dnbinom(xvalue,
20                  size = fit$estimate["size"],
21                  mu = fit$estimate["mu"])
22 lines(xvalue,yvalue)
```



Practice

⑤ 95% confidence interval of R_0 and k

```
33 # Estimate 95th percentile of size and mu
34 conf_interval <- confint(fit, level = 0.95)
35 print(conf_interval)
```

```
> # Estimate 95th percentile of size and mu
> conf_interval <- confint(fit_nb, level = 0.95)
> print(conf_interval)
```

	2.5 %	97.5 %
size	0.2533811	0.5982243
mu	0.4475402	0.7179791

Summary

1. Measures for transmissibility
2. Basic reproduction number
3. Probability distribution
4. Superspreading potential
5. Practice

References

- Adam CD, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. ***Nature Medicine*** 2020
- Kucharaichi AJ, et al. The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission. ***Eurosurveillance*** 2015
- Ryu S, et al. Transmission dynamics and control of two epidemic waves of SARS-CoV-2 in South Korea. ***BMC Infectious Diseases*** 2021
- Hwang H, et al. Transmission dynamics of the Delta variant of SARS-CoV-2 infections in South Korea. ***Journal of Infectious Diseases*** 2022
- Ryu S, et al. Serial interval and transmission dynamics during the SARS-CoV-2 Delta variant predominance, South Korea. ***Emerging Infectious Diseases*** 2022
- Ryu S, et al. Effect of Nonpharmaceutical interventions on Transmission of Severe Acute Respiratory Syndrome Coronavirus 2, South Korea, 2020. ***Emerging Infectious Diseases*** 2020