Sriram Pemmaraju and Alberto Maria Segre Department of Computer Science The University of Iowa

Spring 2020 https://homepage.cs.uiowa.edu/~sriram/4980/spring20/





Epidemiology [CDC]: the study of the distribution and determinants of healthrelated states or events in specified populations, and the application of this study to the control of health problems.



Epidemiology [CDC]: the study of the distribution and determinants of healthrelated states or events in specified populations, and the application of this study to the control of health problems.

Epidemiology [Webster's]: the science which investigates the causes and control of epidemic diseases.



Epidemiology [CDC]: the study of the distribution and determinants of healthrelated states or events in specified populations, and the application of this study to the control of health problems.

Epidemiology [Webster's]: the science which investigates the causes and control of epidemic diseases.

Epidemic [Webster's]: common to or affecting many people in a community at the same time; prevalent; widespread; said of contagious diseases.



Epidemiology [CDC]: the study of the distribution and determinants of healthrelated states or events in specified populations, and the application of this study to the control of health problems.

Epidemiology [Webster's]: the science which investigates the causes and control of epidemic diseases.

Epidemic [Webster's]: common to or affecting many people in a community at the same time; prevalent; widespread; said of contagious diseases.

Epidemiology is not restricted to the study of contagion, nor should it be confused with immunology (the study of an agent's contagion defense system).





University of Iowa Epidemiology Department (College of Public Health): cancer epidemiology...

... causes, prevention, detection, treatment and quality of life.



University of Iowa Epidemiology Department (College of Public Health): cancer epidemiology...

... causes, prevention, detection, treatment and quality of life. clinical health services epidemiology...

...evaluate performance of clinical and preventative health care practices.



University of Iowa Epidemiology Department (College of Public Health): cancer epidemiology...

... causes, prevention, detection, treatment and quality of life. clinical health services epidemiology...

...evaluate performance of clinical and preventative health care practices. chronic disease epidemiology...

...role of genetics, nutrition, behavior and environment on chronic disease.



University of Iowa Epidemiology Department (College of Public Health): cancer epidemiology...

... causes, prevention, detection, treatment and quality of life. clinical health services epidemiology...

...evaluate performance of clinical and preventative health care practices. chronic disease epidemiology...

...role of genetics, nutrition, behavior and environment on chronic disease. injury epidemiology...

...quantify, prioritize and mitigate risk factors for injury in a population.



University of Iowa Epidemiology Department (College of Public Health): cancer epidemiology...

... causes, prevention, detection, treatment and quality of life. clinical health services epidemiology...

...evaluate performance of clinical and preventative health care practices. chronic disease epidemiology...

...role of genetics, nutrition, behavior and environment on chronic disease. injury epidemiology...

...quantify, prioritize and mitigate risk factors for injury in a population. molecular and genetic epidemiology...

... understand the impact of genetic variation on disease.



University of Iowa Epidemiology Department (College of Public Health): cancer epidemiology...

... causes, prevention, detection, treatment and quality of life. clinical health services epidemiology...

...evaluate performance of clinical and preventative health care practices. chronic disease epidemiology...

...role of genetics, nutrition, behavior and environment on chronic disease. injury epidemiology...

...quantify, prioritize and mitigate risk factors for injury in a population. molecular and genetic epidemiology...

... understand the impact of genetic variation on disease.

infectious disease epidemiology...

...surveillance, risk factors, prediction and mitigation of disease.

University of Iowa COMP CON computational epidemiology research In 1854, a cholera epidemic hit the modern-day Soho district in London, killing 616 people.



In 1854, a cholera epidemic hit the modern-day Soho district in London, killing 616 people.

Cholera is a bacterial infection of the small intestine, characterized by vomiting and diarrhea, that can kill up to 50% of those infected.



In 1854, a cholera epidemic hit the modern-day Soho district in London, killing 616 people.

Cholera is a bacterial infection of the small intestine, characterized by vomiting and diarrhea, that can kill up to 50% of those infected.

At the time, the primary theory of disease was the *miasma theory*, where breathing "bad air" (Italian: "mal aria") made you sick (and there was plenty of bad air in 1854 London).



John Snow



John Inow

John Snow (1813-1858), a London physician, was skeptical of the miasma theory of infection which was prevalent at the time, believing that cholera was water borne (the *germ theory* of disease).



John Snow



John Inow

John Snow (1813-1858), a London physician, was skeptical of the miasma theory of infection which was prevalent at the time, believing that cholera was water borne (the *germ theory* of disease).

His analysis of the 1854 cholera outbreak in his neighborhood was published in his 1856 report *On the Mode of the Communication of Cholera*.

https://youtu.be/lNjrAXGRda4



The Broad Street Pump





Voronoi Diagram in "Step Space"





To validate his ideas, Snow noted that different neighborhoods drew water from different sources.



To validate his ideas, Snow noted that different neighborhoods drew water from different sources.

One company, Lambeth, drew water from the Thames upstream from where London sewage entered the river, while the other, S&V, drew water from downstream.



To validate his ideas, Snow noted that different neighborhoods drew water from different sources.

One company, Lambeth, drew water from the Thames upstream from where London sewage entered the river, while the other, S&V, drew water from downstream.

Snow then compared cholera counts among these two very similar populations served by the different companies to support his theory that cholera was water borne.

Supplier	Number of houses	Cholera deaths	Deaths per 10,000 houses
S&V	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59



Snow's "Grand Experiment" of 1854







He used simple statistics to explore the correlation between water source and disease.



He used simple statistics to explore the correlation between water source and disease.

He used geometric properties of the underlying problem to find support for his theory.



He used simple statistics to explore the correlation between water source and disease.

He used geometric properties of the underlying problem to find support for his theory.

He considered both positive and negative counter examples (brewery workers and the woman from Hampton) to inform his theory.



He used simple statistics to explore the correlation between water source and disease.

He used geometric properties of the underlying problem to find support for his theory.

He considered both positive and negative counter examples (brewery workers and the woman from Hampton) to inform his theory.

He sought to test his theory with an intervention (pump handle removal).



He used simple statistics to explore the correlation between water source and disease.

He used geometric properties of the underlying problem to find support for his theory.

He considered both positive and negative counter examples (brewery workers and the woman from Hampton) to inform his theory.

He sought to test his theory with an intervention (pump handle removal).

He followed up with a (natural) "grand experiment," using statistics to compare outcomes across two similar populations.



He used simple statistics to explore the correlation between water source and disease.

He used geometric properties of the underlying problem to find support for his theory.

He considered both positive and negative counter examples (brewery workers and the woman from Hampton) to inform his theory.

He sought to test his theory with an intervention (pump handle removal).

He followed up with a (natural) "grand experiment," using statistics to compare outcomes across two similar populations.

He was ultimately interested in making recommendations to public officials.



"To understand the behavior of complex biological systems, it is useful to devise computer based models by approximating the interactions, via biomathematical expressions. Without doubt, these models could be over simplifications of complex interactions but they would be useful in comparison to classical laboratory experimental approaches which may not be practical or feasible." [Habtemariam *et al.*, 1988]



"To understand the behavior of complex biological systems, it is useful to devise computer based models by approximating the interactions, via biomathematical expressions. Without doubt, these models could be over simplifications of complex interactions but they would be useful in comparison to classical laboratory experimental approaches which may not be practical or feasible." [Habtemariam *et al.*, 1988]

The schema in this paper is: [mathematical model] + [data] + [implementation] + [testing, validation, sensitivity analysis].



"To understand the behavior of complex biological systems, it is useful to devise computer based models by approximating the interactions, via biomathematical expressions. Without doubt, these models could be over simplifications of complex interactions but they would be useful in comparison to classical laboratory experimental approaches which may not be practical or feasible." [Habtemariam *et al.*, 1988]

The schema in this paper is: [mathematical model] + [data] + [implementation] + [testing, validation, sensitivity analysis].

Here, they let the computer do the work of revealing the behavior of a complex dynamical system (again, not necessarily limited to the study of disease).



"To understand the behavior of complex biological systems, it is useful to devise computer based models by approximating the interactions, via biomathematical expressions. Without doubt, these models could be over simplifications of complex interactions but they would be useful in comparison to classical laboratory experimental approaches which may not be practical or feasible." [Habtemariam *et al.*, 1988]

The schema in this paper is: [mathematical model] + [data] + [implementation] + [testing, validation, sensitivity analysis].

Here, they let the computer do the work of revealing the behavior of a complex dynamical system (again, not necessarily limited to the study of disease).

Simulation is just one of the "new ideas" that distinguish computational epidemiology from traditional epidemiology.



The immediate post World War II period (*i.e.*, the advent of computing) saw the application of simulation studies to nuclear physics and meteorology.


This use of simulation represents a fundamental paradigm shift, an "alternative way" of doing science.



This use of simulation represents a fundamental paradigm shift, an "alternative way" of doing science.

Philosophers of science have traditionally had trouble with the use of simulations, because they move "down the chain" from model to observation rather than "up the chain." ["Computer Simulations in Science," *The Stanford Encyclopedia of Philosophy*, E. Zalta, ed.]



This use of simulation represents a fundamental paradigm shift, an "alternative way" of doing science.

Philosophers of science have traditionally had trouble with the use of simulations, because they move "down the chain" from model to observation rather than "up the chain." ["Computer Simulations in Science," *The Stanford Encyclopedia of Philosophy*, E. Zalta, ed.]

Yet in fields where experiments are not possible, mathematical models and computer simulations can yield insight into how a system behaves under pressure from external forces (e.g., an intervention).



This use of simulation represents a fundamental paradigm shift, an "alternative way" of doing science.

Philosophers of science have traditionally had trouble with the use of simulations, because they move "down the chain" from model to observation rather than "up the chain." ["Computer Simulations in Science," *The Stanford Encyclopedia of Philosophy*, E. Zalta, ed.]

Yet in fields where experiments are not possible, mathematical models and computer simulations can yield insight into how a system behaves under pressure from external forces (e.g., an intervention).

Of course, the value of a simulation is limited by the quality of the underlying model and the values of any necessary parameters.

University of Iowa COMP CON computational epidemiology research



Agent-based simulations (esp. social and behavioral sciences) are based on a mathematical model where each agent is described by its own set of rules for interaction, and overall system behaviors are emergent.



Agent-based simulations (esp. social and behavioral sciences) are based on a mathematical model where each agent is described by its own set of rules for interaction, and overall system behaviors are emergent.

Multiscale simulations are hybrids based on more than one model, where each model operates on a different level of abstraction, and simulation proceeds from general to specific (serial multiscale) or is performed simultaneously at multiple scales (parallel multiscale).



Agent-based simulations (esp. social and behavioral sciences) are based on a mathematical model where each agent is described by its own set of rules for interaction, and overall system behaviors are emergent.

Multiscale simulations are hybrids based on more than one model, where each model operates on a different level of abstraction, and simulation proceeds from general to specific (serial multiscale) or is performed simultaneously at multiple scales (parallel multiscale).

Note: *Monte Carlo simulations* use randomness to estimate the solution of a mathematical model: here, randomness of the algorithm is not a feature of the model itself. The original post-war simulation is now considered a calculational tool, and not really a "simulation."

University of Iowa COMP COPI computational epidemiology research



We'll spend a good amount of time this term looking at various of disease diffusion models, how they are constructed, and how their parameters are tuned.



We'll spend a good amount of time this term looking at various of disease diffusion models, how they are constructed, and how their parameters are tuned.

We'll also look at algorithms on these models for, *e.g.*, constructing models from data, or making predictions on the basis of these models.



We'll spend a good amount of time this term looking at various of disease diffusion models, how they are constructed, and how their parameters are tuned.

We'll also look at algorithms on these models for, *e.g.*, constructing models from data, or making predictions on the basis of these models.

We'll also talk about surveillance and interventions; how does one detect the presence of disease? How does one control its spread, and how effective are the various interventions to do so likely to be (according to the model)?



Daniel Bernoulli and a Model for Smallpox



The first known example of an explicit mathematical model intended to inform public health is due to Daniel Bernoulli (1700-1782) in 1766 almost 100 years before John Snow.



Daniel Bernoulli and a Model for Smallpox



The first known example of an explicit mathematical model intended to inform public health is due to Daniel Bernoulli (1700-1782) in 1766 almost 100 years before John Snow.

Bernoulli was a Swiss mathematician famous for the kinetic theory of gasses, the Bernoulli effect in fluid flow, and early work on the statistical characterization of risk.

University of Iowa COMP CONF computational epidemiology research

Smallpox is a viral disease that kills about 30% of those infected; by 1700, it was a leading cause of death in England.



Smallpox is a viral disease that kills about 30% of those infected; by 1700, it was a leading cause of death in England.

Acquired by inhaling the virus (*variola major* or *variola minor*), direct contact, or through the placenta (rare).



Smallpox is a viral disease that kills about 30% of those infected; by 1700, it was a leading cause of death in England.

Acquired by inhaling the virus (*variola major* or *variola minor*), direct contact, or through the placenta (rare).

Usually, fever and vomiting start 12 days after infection, followed 2-3 days later by lesions first in the mouth and then a characteristic skin rash.



Smallpox is a viral disease that kills about 30% of those infected; by 1700, it was a leading cause of death in England.

Acquired by inhaling the virus (*variola major* or *variola minor*), direct contact, or through the placenta (rare).

Usually, fever and vomiting start 12 days after infection, followed 2-3 days later by lesions first in the mouth and then a characteristic skin rash.

Macules (pimples) become papules (raised) become vesicles (clear fluid) become leaking pustules (opaque fluid) become scabs by day 20.



Smallpox is a viral disease that kills about 30% of those infected; by 1700, it was a leading cause of death in England.

Acquired by inhaling the virus (*variola major* or *variola minor*), direct contact, or through the placenta (rare).

Usually, fever and vomiting start 12 days after infection, followed 2-3 days later by lesions first in the mouth and then a characteristic skin rash.

Macules (pimples) become papules (raised) become vesicles (clear fluid) become leaking pustules (opaque fluid) become scabs by day 20.

Prior infection confers lifetime immunity; inoculation with *variola minor* (less fatal than *variola major*) first documented in China during the 10th century.





Bangladeshi child infected with smallpox in 1973. Freedom from smallpox was declared in Bangladesh in December, 1977 when a WHO International Commission officially certified that smallpox had been eradicated from that country. The CDC declared smallpox eradicated worldwide in 1980 [Wikipedia; photo source CDC].





Variolation is the practice of inoculation with the *variola* virus. Physicians would select source patients with mild cases of smallpox (likely *variola minor*), and then scratch the target patient and introduce a small bit of fluid or ground scab.



Variolation is the practice of inoculation with the *variola* virus. Physicians would select source patients with mild cases of smallpox (likely *variola minor*), and then scratch the target patient and introduce a small bit of fluid or ground scab.

Variolated patients did get smallpox and were infectious, but the disease acquired (via localized direct contact, hopefully from *variola minor* virus) was likely less severe than that you acquire naturally (via inhalation, often from *variola major*). Variolation had a roughly 2-3% fatality rate.



Variolation is the practice of inoculation with the *variola* virus. Physicians would select source patients with mild cases of smallpox (likely *variola minor*), and then scratch the target patient and introduce a small bit of fluid or ground scab.

Variolated patients did get smallpox and were infectious, but the disease acquired (via localized direct contact, hopefully from *variola minor* virus) was likely less severe than that you acquire naturally (via inhalation, often from *variola major*). Variolation had a roughly 2-3% fatality rate.

Technique varied in how the target patient was prepared, what other treatments (many bogus) were combined, and how the target was exposed (scratches, deep cuts, inhalation of powdered scab, etc.).

University of Iowa COMP COPI computational epidemiology research

Variolation was commonplace in England starting circa 1720; Cotton Mather used variolation in Boston as early as 1706 (learned from a West African slave), becoming commonplace after the smallpox outbreak of 1721.



Variolation was commonplace in England starting circa 1720; Cotton Mather used variolation in Boston as early as 1706 (learned from a West African slave), becoming commonplace after the smallpox outbreak of 1721.

In 1796, Edward Jenner (1749-1823) discovered that immunity to smallpox could be conferred via *vaccination* (Latin: *vacca*, or "cow") with cowpox (a zoonotic virus), reducing the risk to the individual while still inducing an immune response.



Variolation was commonplace in England starting circa 1720; Cotton Mather used variolation in Boston as early as 1706 (learned from a West African slave), becoming commonplace after the smallpox outbreak of 1721.

In 1796, Edward Jenner (1749-1823) discovered that immunity to smallpox could be conferred via *vaccination* (Latin: *vacca*, or "cow") with cowpox (a zoonotic virus), reducing the risk to the individual while still inducing an immune response.

The initial idea came from the observation that dairy farmers and others who worked with cattle and horses (*variola equina* or horsepox) were often immune to smallpox.



Variolation was commonplace in England starting circa 1720; Cotton Mather used variolation in Boston as early as 1706 (learned from a West African slave), becoming commonplace after the smallpox outbreak of 1721.

In 1796, Edward Jenner (1749-1823) discovered that immunity to smallpox could be conferred via *vaccination* (Latin: *vacca*, or "cow") with cowpox (a zoonotic virus), reducing the risk to the individual while still inducing an immune response.

The initial idea came from the observation that dairy farmers and others who worked with cattle and horses (*variola equina* or horsepox) were often immune to smallpox.

Cowpox is mild in humans, does not pose risk of fatality, and is not easily transmitted between humans.





Bernoulli set out to compare the long-term benefit of variolation to the immediate risk of dying. His explicit goal was to influence public policy.



Bernoulli set out to compare the long-term benefit of variolation to the immediate risk of dying. His explicit goal was to influence public policy.

His model quantifies the value of universal inoculation policies in terms of average life expectancy (a population-level parameter), thus making the tradeoff between the individual risk of inoculation and the resulting population-level benefits explicit.



Bernoulli set out to compare the long-term benefit of variolation to the immediate risk of dying. His explicit goal was to influence public policy.

His model quantifies the value of universal inoculation policies in terms of average life expectancy (a population-level parameter), thus making the tradeoff between the individual risk of inoculation and the resulting population-level benefits explicit.

Bernoulli assumed those infected with smallpox die instantaneously with probability *a*, and that those who recovered obtained lifelong immunity.



Bernoulli set out to compare the long-term benefit of variolation to the immediate risk of dying. His explicit goal was to influence public policy.

His model quantifies the value of universal inoculation policies in terms of average life expectancy (a population-level parameter), thus making the tradeoff between the individual risk of inoculation and the resulting population-level benefits explicit.

Bernoulli assumed those infected with smallpox die instantaneously with probability *a*, and that those who recovered obtained lifelong immunity.

He also assumed a cohort w(t) of age t consisted of the never infected x(t) and those with immunity z(t), thus w(t) = x(t) + z(t), and that the probability of those in x(t) acquiring smallpox at any is always b independent of t.

University of Iowa COMP COPI computational epidemiology research Bernoulli then directly solved the two resulting ordinary differential equations to obtain his model:

$$x(t) = \frac{w(t)}{(1-a)e^{bt} + a}$$



Bernoulli then directly solved the two resulting ordinary differential equations to obtain his model:

$$x(t) = \frac{w(t)}{(1-a)e^{bt} + a}$$

Using a = 0.125 and b = 0.125 (estimated from observational data) he calculated that the population would be 14% larger at age 26, and that life expectancy would increase from 26.58 to 29.75 if all children were variolated at birth.



Bernoulli then directly solved the two resulting ordinary differential equations to obtain his model:

$$x(t) = \frac{w(t)}{(1-a)e^{bt} + a}$$

Using a = 0.125 and b = 0.125 (estimated from observational data) he calculated that the population would be 14% larger at age 26, and that life expectancy would increase from 26.58 to 29.75 if all children were variolated at birth.

Repeating the calculation with the assumption that 2% of variolated children would die reduces the gain in life expectancy by 1 month: still a good deal for society (and the King, who wanted to increase the pool of military recruits).


Bernoulli then directly solved the two resulting ordinary differential equations to obtain his model:

$$x(t) = \frac{w(t)}{(1-a)e^{bt} + a}$$

Using a = 0.125 and b = 0.125 (estimated from observational data) he calculated that the population would be 14% larger at age 26, and that life expectancy would increase from 26.58 to 29.75 if all children were variolated at birth.

Repeating the calculation with the assumption that 2% of variolated children would die reduces the gain in life expectancy by 1 month: still a good deal for society (and the King, who wanted to increase the pool of military recruits).

Repeating the calculation again, adding the effect of secondary "artificial smallpox" infections from variolated children (recall these are likelier to be mild cases by construction) does not appreciably change these results.



18

K. Dietz, J.A.P. Heesterbeek | Mathematical Biosciences 180 (2002) 1-21



Fig. 12. Life expectancy at birth as a function of the proportion immunized p. The dotted line refers to those that are immunized, the broken line corresponds to the non-immunized individuals and the continuous line represents the weighted average of the two values where the weight is given by p.

University of Iowa COMP COL computational epidemiology research Neither Bernoulli nor Snow needed a computer: their models was simple enough that it could be solved by hand (Bernoulli) or by visualization and simple statistics (Snow) while still providing insight in the underlying disease process.



Neither Bernoulli nor Snow needed a computer: their models was simple enough that it could be solved by hand (Bernoulli) or by visualization and simple statistics (Snow) while still providing insight in the underlying disease process.

In the next lecture, we're going to look at a recent paper that is similar but much more data driven.

