WARWICK
THE UNIVERSITY OF WARWICK

MA930 Data Analysis & Machine Learning

Lecture 1: Basic Probability

Haoran Ni (haoran.ni@warwick.ac.uk)
4th Year PhD student in MathSys

Outline

- Course Overview
- Rules of Probability
- Discrete and continuous distributions
- Cumulative distributions
- Probability generating functions
- Characteristic functions

Objectives

By the end of today's session, you should:

- Understand the outline of the course
- Know what is required of you
  (Including the deadlines of assignments)
- Believe that data analysis techniques are crucial for answering real world problems
- Have revised background probability
- Be happy to ask questions during lectures of either me or Nathan.

Course Overview

- Lectures and problem classes (32 hours)
- Weeks 1-4: Mondays and Thursdays
- Week 5: Class test (Monday 30th October) and vivas (Thursday 2nd November)
- 2 homework assignments

  Assignment 1. Due by noon on Friday 20th October (week 3)

  Assignment 2. Due by noon on Friday 27th October (week 4)
- Evaluation

  Homework assignments (20%)

  Class test (40%)

  Viva (40%)

Course Overview

- Day 1. Motivation, Basic probability
- Day 2. Probability, Basic statistics. Sample mean and variance, law of large numbers, central limit theorem
- Day 3. Frequentist statistics. Point estimation, confidence intervals, type-I and type-II errors, hypothesis tests
- Day 4. Bayesian statistics. Likelihood, maximum likelihood, Bayes' theorem, conjugate priors, credible intervals
- Day 5 (morning). Time-series analysis. Polynomial fits, auto-regressive models

Course Overview

- Day 5 (afternoon). Advanced Bayesian inference I. Markov chain Monte Carlo (MCMC)
- Day 6 (morning). Advanced Bayesian inference II. Approximate Bayesian Computation (ABC)
- Day 6 (afternoon). Machine learning for data analysis I. General concepts, gradient descent, logistic regression
- Day 7. Machine learning for data analysis II. K-nearest neighbour, K-means clustering, neural networks
- Day 8. Machine learning practical tutorial – non-examinable

Course Overview

Goal for each day: Work through the slides and complete all exercises.

Lectures and problem classes will be combined, providing background material and an opportunity to ensure all exercises are completed. I will try to bring in real-world research examples.

Intro…

- Name
- Why MathSys?
- What do you want to do in your research?

THE UNIVERSITY OF WARWICK

Why do data analysis?

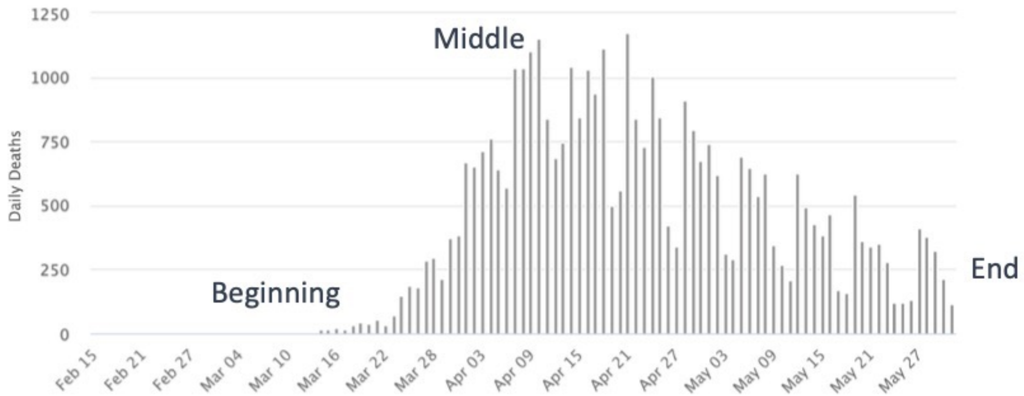
Course Overview

Why do data analysis?

- Investigate answers to real-world questions
- Explore phenomena in many different areas of research
- Good alternative to guessing!

Course Overview

Let us check out a real-world example for data analysis:



Beginning

- Will initial cases lead to a major epidemic?
- Which interventions reduce the epidemic risk?

Middle

- How effective are current interventions?
- Which interventions will minimise numbers of cases?

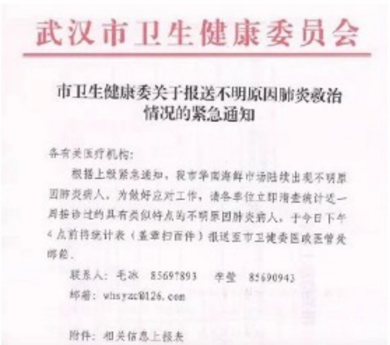
End

- How should interventions be lifted?
- Is the epidemic over?

Course Overview

Beginning

- Will initial cases lead to a major epidemic?
- Which interventions reduce the epidemic risk?



武汉市卫生健康委员会

市卫生健康委关于报送不明原因肺炎救治情况的紧急通知

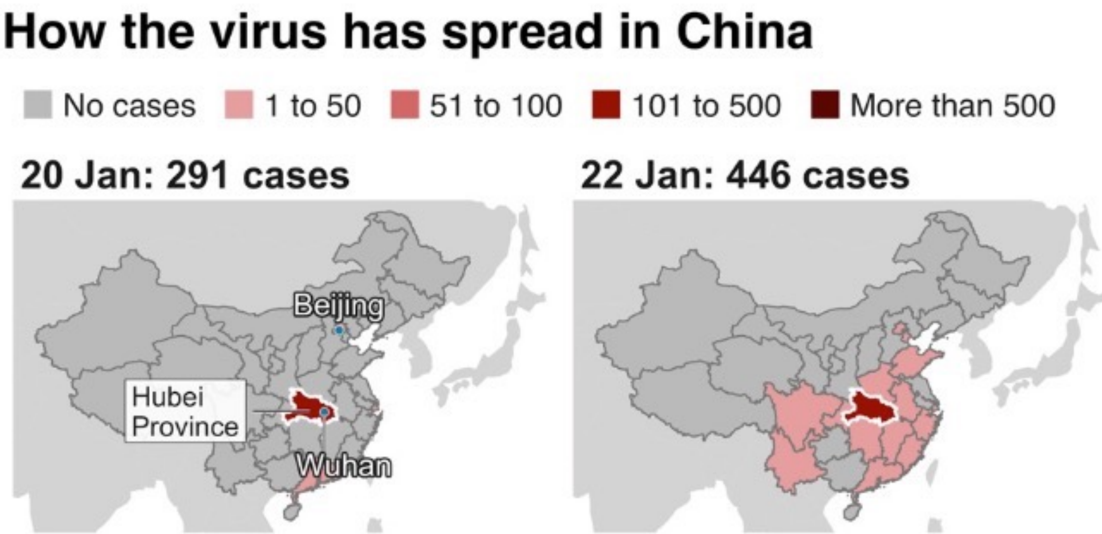
各有关医疗机构：

根据上级紧急通知，我市华南海鲜市场陆续出现不明原因肺炎病人。为做好应对工作，请各单位立即摸查统计近一周接诊过的具有类似特点的不明原因肺炎病人，于今日下午4点前将统计表（盖章扫描件）报送至市卫健委医政医管处邮箱。

联系人：毛冰 85697893 李莹 85690943

邮箱：whsyzc@126.com

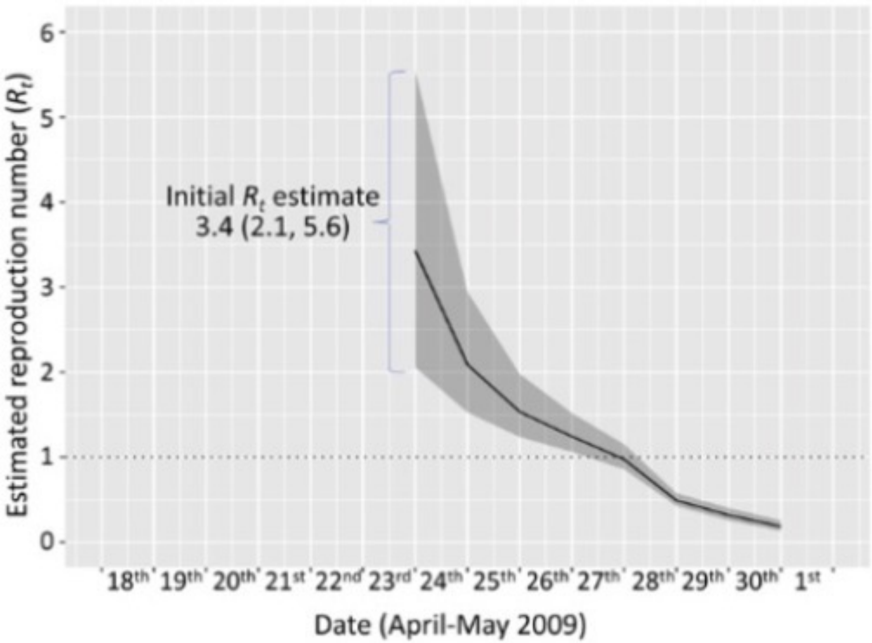
附件：相关信息上报表



23 Jan: Line lists released (approx. 70 patients, incomplete data)

What is the epidemic risk outside of China?

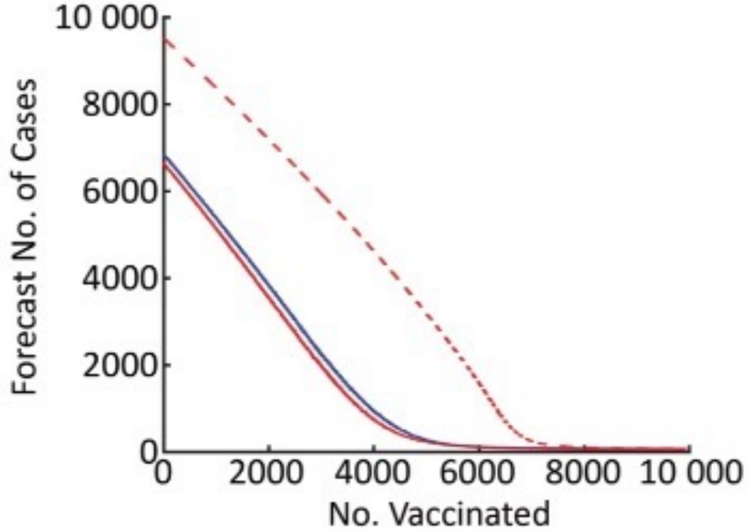
Course Overview



$$\mathbf{P}(I_{t-\tau}^{\text{local}}, I_{t-\tau+1}^{\text{local}}, \ldots, I_t^{\text{local}} \mid I_0, \ \ldots, I_{t-\tau-1}, w_s, R_t)$$
$$= \prod_{k=t-\tau}^{t} \frac{(R_t \Lambda_k(w_s))^{I_k^{\text{local}}} \exp(-R_t \Lambda_k(w_s))}{I_k^{\text{local}}!}$$

$$\mathbf{P}(R_t \mid I_0, I_1, I_2, \ \ldots, I_{t-\tau-1}, I_{t-\tau}^{\text{local}}, I_{t-\tau+1}^{\text{local}}, \ \ldots, I_t^{\text{local}}, \ w_s)$$

$$\propto \mathbf{P}(I_{t-\tau}^{\text{local}}, I_{t-\tau+1}^{\text{local}}, \ldots, I_t^{\text{local}} \mid I_0, \ \ldots, I_{t-\tau-1}, w_s, R_t)\mathbf{P}(R_t)$$



Middle

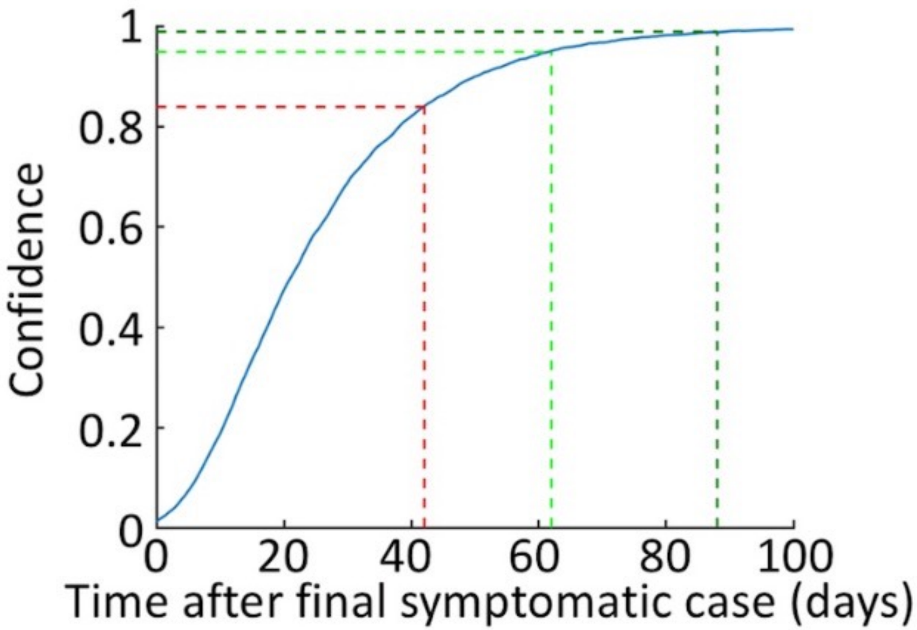
- How effective are current interventions?
- Which interventions will minimise numbers of cases?

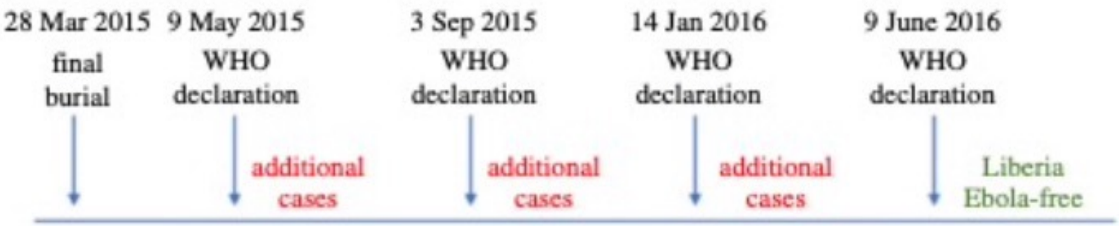
Thompson *et al.*, Epidemics, 2019

Course Overview

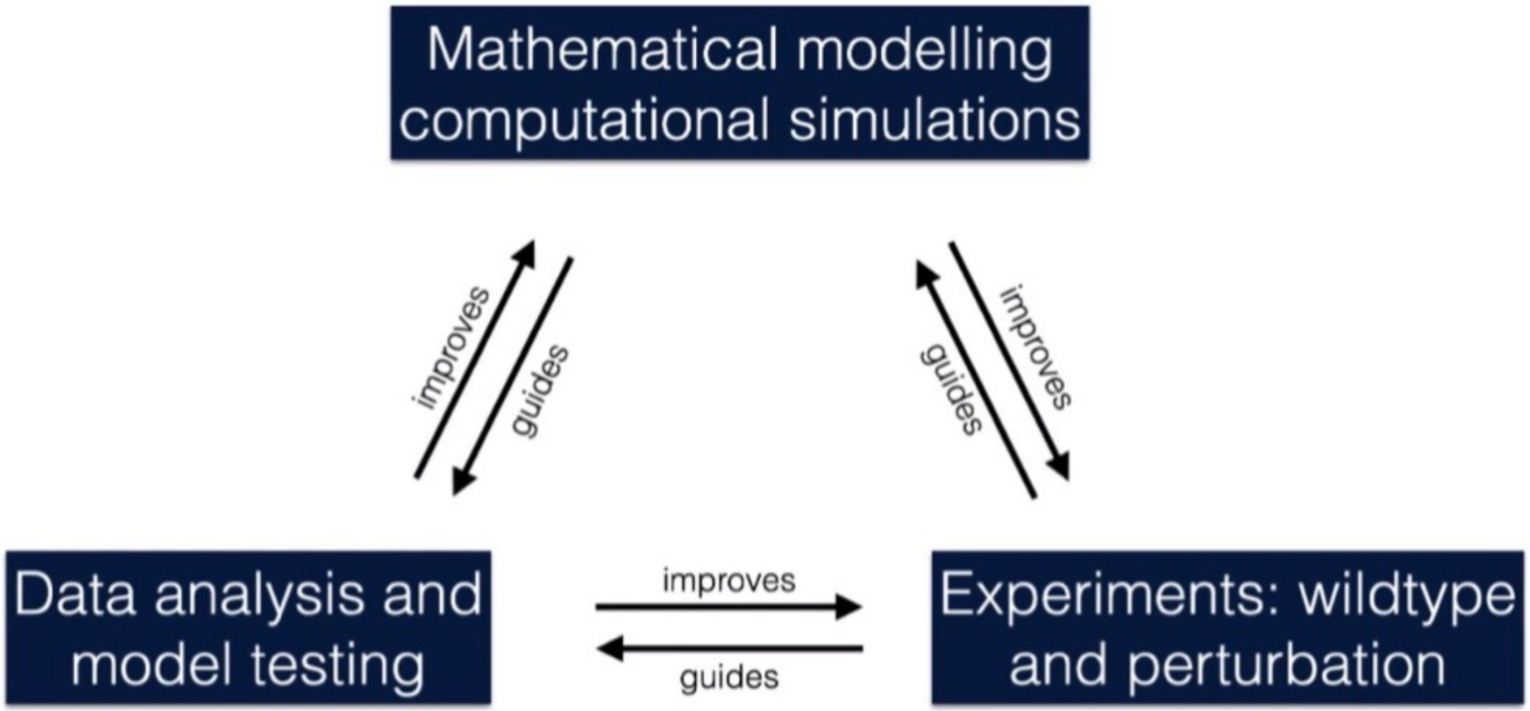
End

- How should interventions be lifted?
- Is the epidemic over?





Course Overview



Data analysis is integral in the cycle of predict – test – refine – predict

Outline

- Course Overview
- Rules of Probability
- Discrete and Continuous Distributions
- Cumulative Distributions
- Common Distributions
- Probability Generating Functions
- Characteristic Functions

Rules of Probability

- Consider an experiment with possible *outcomes* $\omega \subseteq \Omega$
- Ω is referred to as the *sample space*

Rules of Probability

- Consider an experiment with possible *outcomes* $\omega \subseteq \Omega$
- Ω is referred to as the *sample space*

Example: When Throwing two dice,

$$\Omega = \{(x, y): 1 \leq x, y \leq 6\}$$

- An event $A \subseteq \Omega$ occurs if the outcome is a member of A

Getting a total of 3,

$$A = \{(1,2), (2,1)\}$$

Rules of Probability

- **Rule 1. Fundamental convention**
  For a complete set of non-overlapping events $A \subseteq \Omega$, then
  $$\sum_{A \subseteq \Omega} \boldsymbol{P}(A) = 1.$$

Rules of Probability

- **Rule 1. Fundamental convention**

  For a complete set of non-overlapping events $A \subseteq \Omega$, then

$$\sum_{A \subseteq \Omega} \boldsymbol{P}(A) = 1.$$

- **Rule 2. Positivity**

  Probabilities are non-negative, $0 \leq \boldsymbol{P}(A) \leq 1$.

Rules of Probability

- **Rule 1. Fundamental convention**

  For a complete set of non-overlapping events $A \subseteq \Omega$, then

  $$\sum_{A \subseteq \Omega} \boldsymbol{P}(A) = 1.$$

- **Rule 2. Positivity**

  Probabilities are non-negative, $0 \leq \boldsymbol{P}(A) \leq 1$.

- **Rule 3. Union**

  $$\boldsymbol{P}(A \cup B) = \boldsymbol{P}(A) + \boldsymbol{P}(B) - \boldsymbol{P}(A \cap B)$$

  If $A \cap B = \emptyset$, then A and B are said to be *disjoint* or *mutually exclusive.*

- **Rule 4. Conditionality product rule**

$$\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A|B)\boldsymbol{P}(B)$$

For independent events, $\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A)\boldsymbol{P}(B)$.

Example. Suppose that there are two baskets. basket 1 has 2 blue balls and 1 red ball, and basket 2 has 3 blue balls and 1 red ball. We select a basket uniformly at random, and pick out a ball. What is the probability of selecting a blue ball from basket 1?

Rules of Probability

- **Rule 4. Conditionality product rule**

$$\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A|B)\boldsymbol{P}(B)$$

For independent events, $\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A)\boldsymbol{P}(B)$.

Example. Suppose that there are two baskets. basket 1 has 2 blue balls and 1 red ball, and basket 2 has 3 blue balls and 1 red ball. We select a basket uniformly at random, and pick out a ball. What is the probability of selecting a blue ball from basket 1?

$$\begin{aligned}\boldsymbol{P}(\text{blue} \cap \text{basket1}) &= \boldsymbol{P}(\text{blue}|\text{basket1})\boldsymbol{P}(\text{basket1}) \\ &= \frac{2}{3} \times \frac{1}{2} = \frac{1}{3}\end{aligned}$$

- **Rule 4. Conditionality product rule**

$$\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A|B)\boldsymbol{P}(B)$$

For independent events, $\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A)\boldsymbol{P}(B)$.

Exercise 1.1. Show that the conditionality product rule extends to three variables, so that

$$\boldsymbol{P}(A \cap B \cap C) = \boldsymbol{P}(A|B \cap C)\boldsymbol{P}(B|C)\boldsymbol{P}(C)$$

What is the conditionality product rule for n variables? Prove it.

Rules of Probability

- **Rule 5. Conditioning**

  For a complete set of non-overlapping events A, then

$$\boldsymbol{P}(B) = \sum_{A \subseteq \Omega} \boldsymbol{P}(B|A)\boldsymbol{P}(A)$$

Example 1.1 (board). Probability of an infectious disease epidemic.

Suppose that an infected person enters a new host population, and that the number of infected hosts acts as a simple process in which each sequential event is either an infection or a recovery, with $\boldsymbol{P}(\text{infection}) = 0.8$ and $\boldsymbol{P}(\text{recovery}) = 0.2$.

What is the probability that the pathogen fades out in the population, without causing an epidemic with large numbers of infections?

Rules of Probability

- **Rule 5. Conditioning**

  For a complete set of non-overlapping events A, then

$$\boldsymbol{P}(B) = \sum_{A \subseteq \Omega} \boldsymbol{P}(B|A)\boldsymbol{P}(A)$$

Exercise 1.2. Probability of an infectious disease epidemic.

Verify the result of Example 1.1 by writing computing code to simulate lots of outbreaks. Record the proportion of outbreaks that fade out before hitting 20 simultaneously infected individuals.

- **Rule 6. Bayes' Rule**

$$\boldsymbol{P}(A|B) = \frac{\boldsymbol{P}(A \cap B)}{\boldsymbol{P}(B)} = \frac{\boldsymbol{P}(B|A)\boldsymbol{P}(A)}{\boldsymbol{P}(B)}$$

Example. Positive/negative predictive value.

The positive predictive value (PPV) is the probability that a patient is infected if they test positive.
The negative predictive value (NPV) is the probability that a patient is uninfected if they test negative.

- **Rule 6. Bayes' Rule**

$$\boldsymbol{P}(A|B) = \frac{\boldsymbol{P}(A \cap B)}{\boldsymbol{P}(B)} = \frac{\boldsymbol{P}(B|A)\boldsymbol{P}(A)}{\boldsymbol{P}(B)}$$

Example 1.2(board). i) What is the PPV and NPV given the probabilities below, obtained from a clinical trial?

		True infection status	
		Infected	Uninfected
Test result	Positive	8/1000	0/1000
	Negative	2/1000	990/1000

Rules of Probability

- **Rule 6. Bayes' Rule**

$$\boldsymbol{P}(A|B) = \frac{\boldsymbol{P}(A \cap B)}{\boldsymbol{P}(B)} = \frac{\boldsymbol{P}(B|A)\boldsymbol{P}(A)}{\boldsymbol{P}(B)}$$

Example 1.2(board). ii) What about in this example, where there is lots of disease in population?

		True infection status	
		Infected	Uninfected
Test result	Positive	940/1000	0/1000
	Negative	50/1000	10/1000

- **Rule 6. Bayes' Rule**

$$\boldsymbol{P}(A|B) = \frac{\boldsymbol{P}(A \cap B)}{\boldsymbol{P}(B)} = \frac{\boldsymbol{P}(B|A)\boldsymbol{P}(A)}{\boldsymbol{P}(B)}$$

Example 1.2(board). ii) What about in this example, where there is lots of disease in population?

		True infection status	
		Infected	Uninfected
Test result	Positive	940/1000	0/1000
	Negative	50/1000	10/1000

Since there is lots of disease, a negative test does not guarantee you are not infected!

Rules of Probability

- **Rule 6. Bayes' Rule**

$$\boldsymbol{P}(A|B) = \frac{\boldsymbol{P}(A \cap B)}{\boldsymbol{P}(B)} = \frac{\boldsymbol{P}(B|A)\boldsymbol{P}(A)}{\boldsymbol{P}(B)}$$

Exercise 1.3. A political party runs a social media campaign in half of all districts in which they have a candidate up for election. The results are:

Ad or no ad?

Win or lose?		Advert	No advert
	Win	4/10	2/10
	Lose	1/10	3/10

What are $\boldsymbol{P}$(advert), $\boldsymbol{P}$(no advert), $\boldsymbol{P}$(win), $\boldsymbol{P}$(lose)? What is $\boldsymbol{P}$(win|advert)? What is $\boldsymbol{P}$(no advert|lose)? Did placing an advert appear to affect the chance of winning?

Outline

- Course Overview
- Rules of Probability
- Discrete and Continuous Distributions
- Cumulative Distributions
- Common Distributions
- Probability Generating Functions
- Characteristic Functions

Discrete and Continuous Distributions

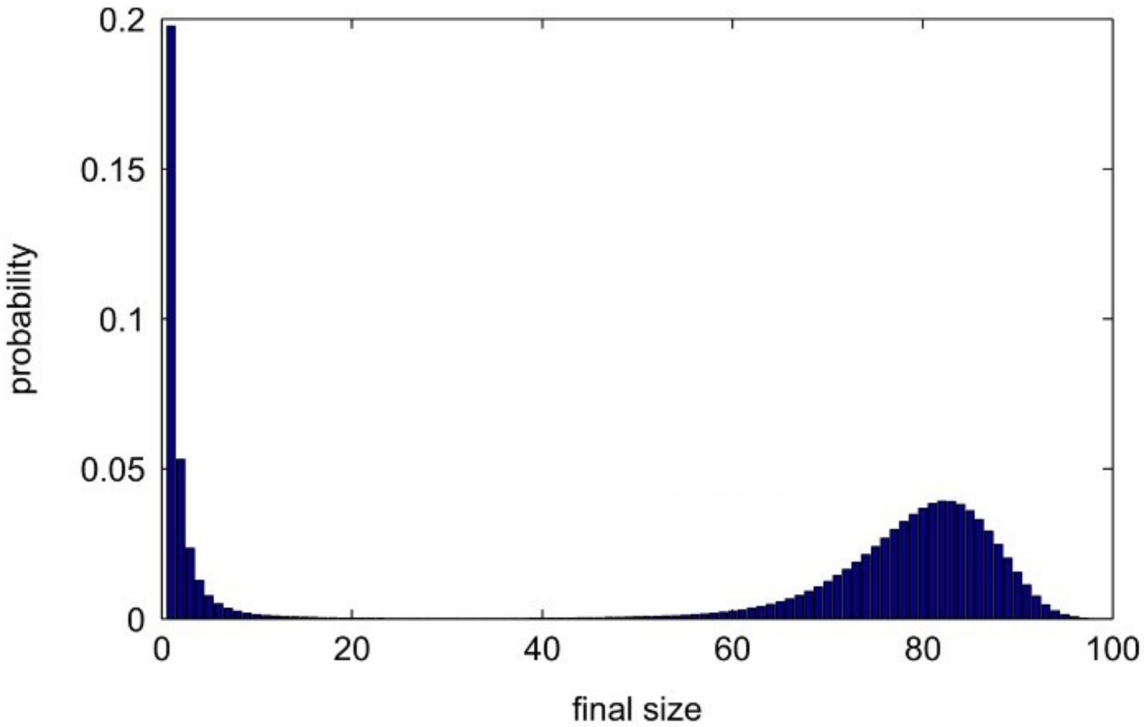
Single-variable (X) events can be labelled with either

- A **discrete** label k for a countable number of states
- A **continuous** label x for a density of states

Discrete and Continuous Distribution

Discrete states have probability $P(k)$[which denotes $\boldsymbol{P}(X = k)$]

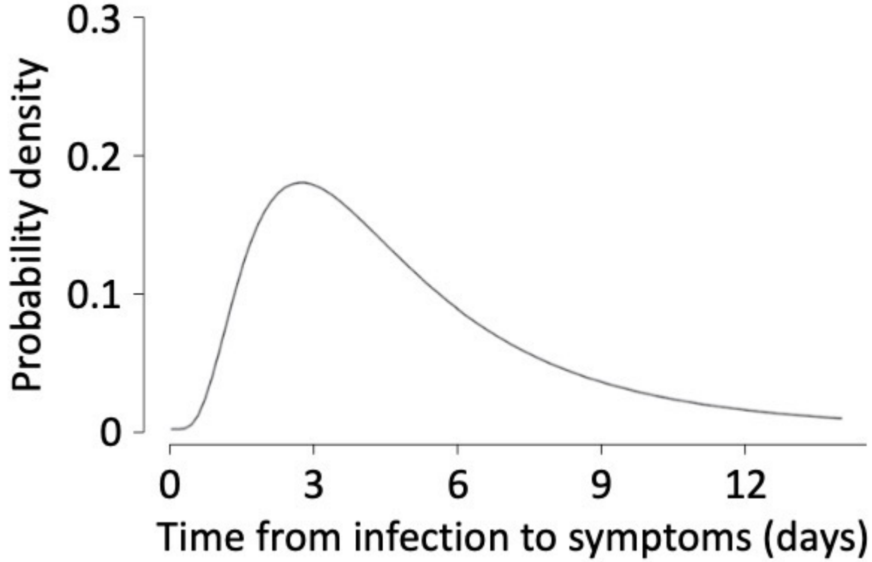
- Summation rule: $\sum_k P(k) = 1$
- Expectation of function: $\mathbf{E}(h(X))= \sum_k h(k)P(k)$



Discrete and Continuous Distribution

Continuous states have probability density $f(x)$, where $\int_a^b f(x)\,\mathrm{d}x$ is the probability that the variable lies between a and b

- Summation (integration) rule: $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$
- Expectation of function: $\mathbf{E}\big(h(X)\big) = \int_{-\infty}^{\infty} h(x) f(x) \mathrm{d}x$



Outline

- Course Overview
- Rules of Probability
- Discrete and Continuous Distributions
- Cumulative Distributions
- Common Distributions
- Probability Generating Functions
- Characteristic Functions

Cumulative Distributions

$F(x)$ – the probability that the variable is less than or equal to x

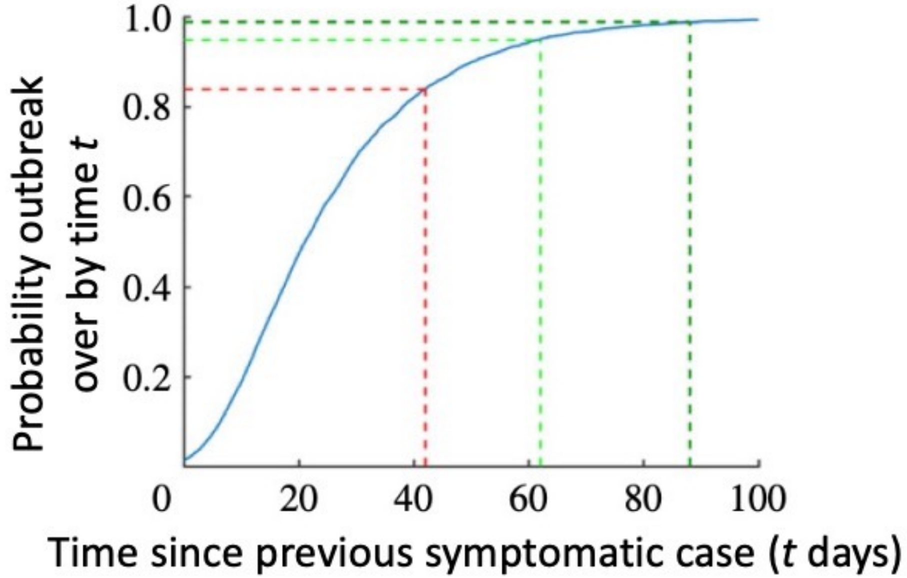
- $F(x) \;= \boldsymbol{P}(X \leq x)$
- For discrete variables, $F(x) = \sum_{k \leq x} P(k)$
- For continuous variables, $F(x) = \int_{-\infty}^{x} f(s)\mathrm{d}s$

It maps probabilities and densities onto the range [0,1]. Useful for generating random numbers from any distribution (will return to this later)

Cumulative Distributions

$F(x)$ – the probability that the variable is less than or equal to x

- $F(x) = \boldsymbol{P}(X \leq x)$
- For discrete variables, $F(x) = \sum_{k \leq x} P(k)$
- For continuous variables, $F(x) = \int_{-\infty}^{x} f(s)\mathrm{d}s$



Cumulative Distributions

Example 1.3 (board). Exponential distribution

Exponentially distributed random variables obey $f(x) = \lambda \exp(-\lambda x)$ for $x \in [0, \infty)$

- What is the mean?
- What is the variance, $\mathbf{E}(X^2) - \mathbf{E}(X)^2$?
- What is the cumulative distribution?

Cumulative Distributions

Example 1.4 (board). Sampling exponential distributed random numbers

Important example (e.g. for simulation models). Given the cumulative distribution of an exponentially distributed random variable, demonstrate that simulating exponential random numbers simply requires sampling a uniform random number on $[0, 1]$ (easy!), followed by transforming it to $x = -\frac{1}{\lambda}\ln(u)$.

Exercise 1.4. Using a computer, plot the density and cumulative distributions of an exponential distribution based on the calculations on the previous slide. Verify them by plotting the analogous distributions resulting from sampling exponentially distributed random variables a large number of times using the above result.

Outline

- Course Overview
- Rules of Probability
- Discrete and Continuous Distributions
- Cumulative Distributions
- Common Distributions
- Probability Generating Functions
- Characteristic Functions

Bernoulli distribution

- Discrete distribution, where x is a binary random number
- $x = 1$ (success) with probability p
- $x = 0$ (failure) with probability $q = 1 - p$

- $\mathbf{E}(X) = p$
- $\mathrm{Var}(X) = p(1 - p)$

Exercise 1.5. Simulate Bernoulli random numbers and check this variance.

Common Distributions

Binomial distribution

- Sum of n Bernoulli random variables, $X = \sum_{j=1}^{n} x_j$
- Number of successes out of n trials
- Discrete distribution with $n + 1$ states
- Order unimportant, so combinatorial factor is included

- $\boldsymbol{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$
- Mean np, Variance $np(1 - p)$

Exercise 1.6. Plot binomial probability distribution functions for:

i) $n = 5$, $p = 1$
ii) $n = 5$, $p = 0.5$
iii) $n = 10$, $p = 0.5$

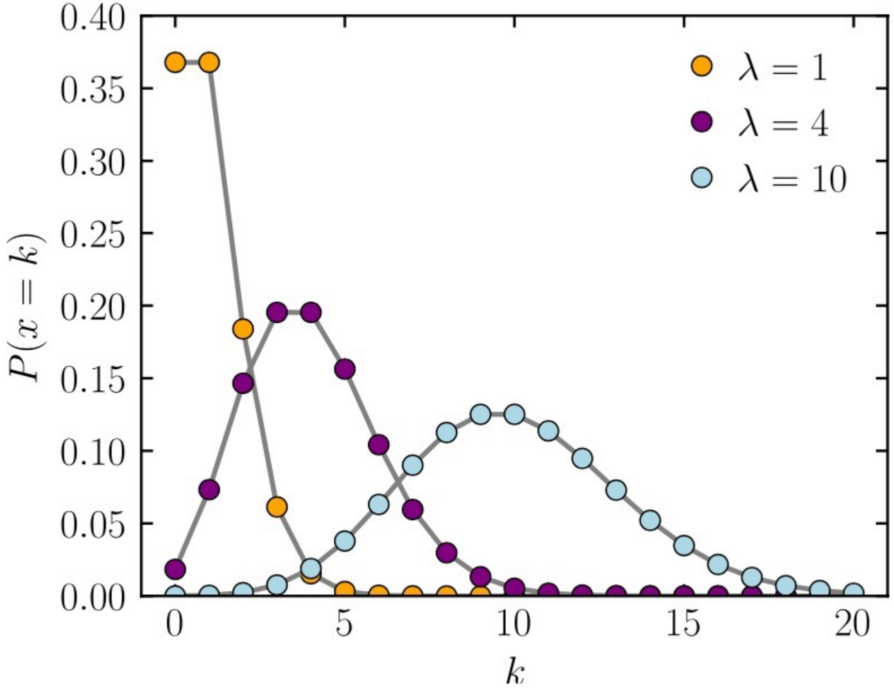
Verify these distributions using simulations of Bernoulli random variables.

Common Distributions

Poisson distribution

- Discrete distribution with countably infinite numbers of states
- k = 0, 1, 2, 3, …
- Determined by rate parameter λ

$$\boldsymbol{P}(X = k) = \frac{\exp(-\lambda)\lambda^k}{k!}$$



- Mean λ, Variance λ

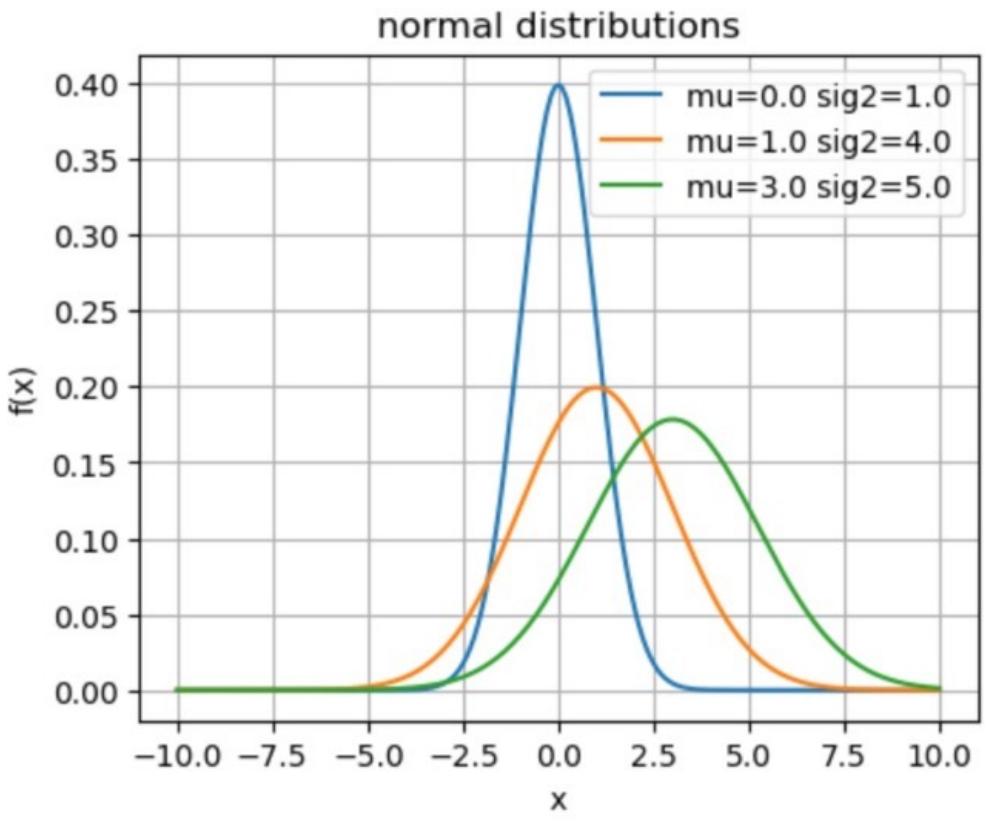
Exercise 1.7. Poisson approximation to the Binomial distribution

If X is binomially distributed (parameters n, p), where n is large and p is small, then X is approximately Poisson distributed with rate np.

Check this by considering the following example. If X is the number of patients who develop severe disease out of a population of n = 1000 patients, where p = 1/1000, then what is the probability that at least 4 patients develop severe disease?

Common Distributions

Normal distribution



Common Distributions

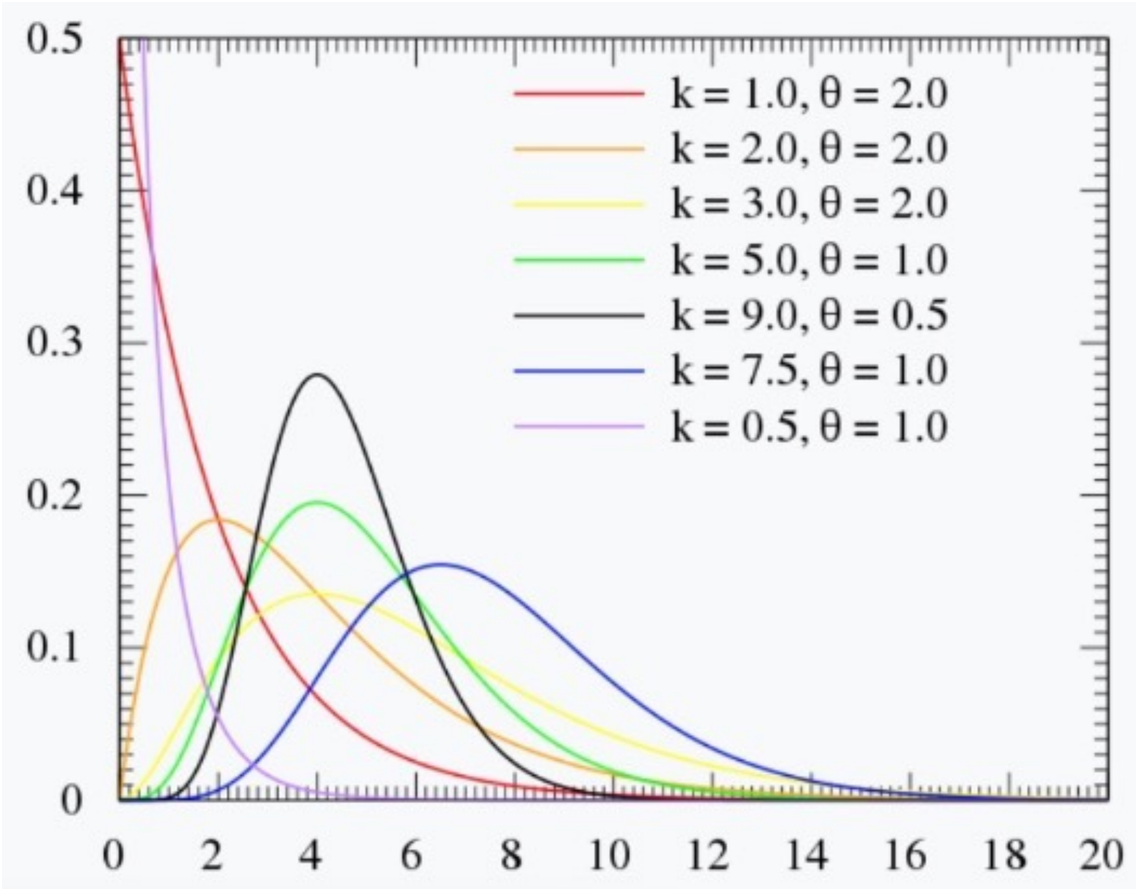
Normal distribution

- Continuous distribution, ubiquitous due to *Central Limit Theorem* (distribution of means of random samples from a population tend to a normal distribution)
- Specified by mean μ and variance σ^2

- Standard normal: $\mu = 0, \sigma^2 = 1$
- Sum of Gaussian numbers is Gaussian (with summed means and variances)

Common Distributions

Gamma distribution



Common Distributions

Gamma distribution

- Continuous distribution; shape parameter k and scale parameter θ

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} \exp(-\frac{x}{\theta})$$

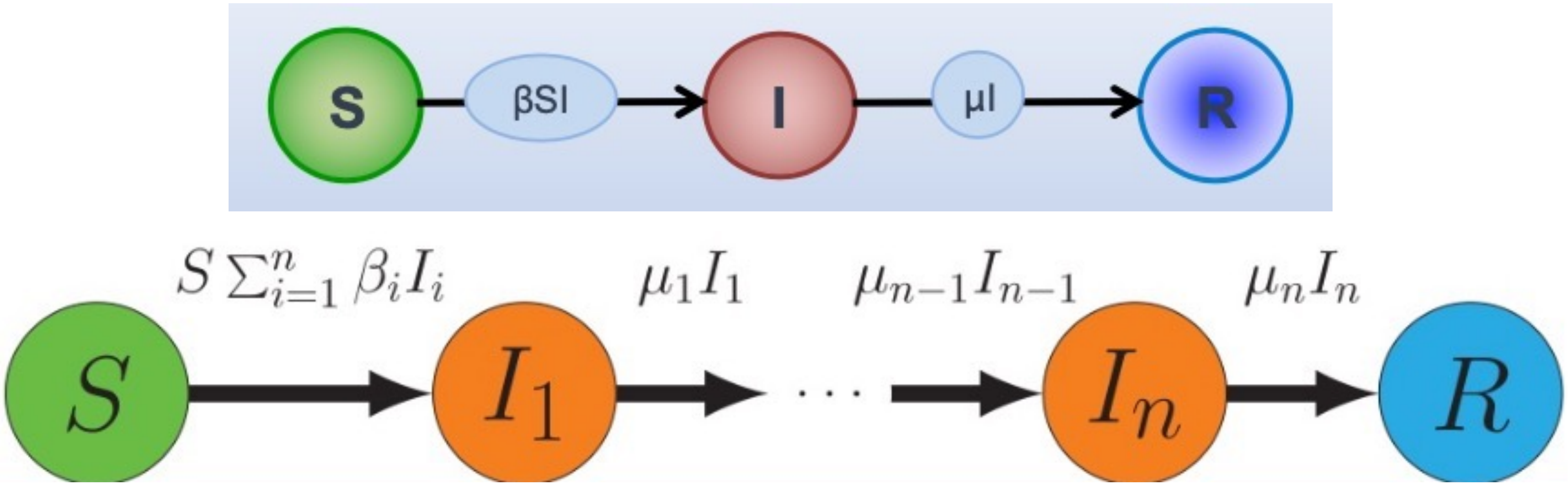
- Mean $k\theta$
- Variance $k\theta^2$

- Special case: Exponential distribution ($k = 1$)
- Special case: Erlang distribution (k is an integer)

Common Distributions

Sum of *k* independent and identically distributed exponential distributions is Erlang/gamma distributed, with shape parameter *k*

Example. Linear chain trick / Method of stages



Common Distributions

Multidimensional distributions

- Discrete case: $P(x,y)$, where $\sum_x \sum_y P(x,y) = 1$
- Continuous case: $f(x,y)$, where $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x\,\mathrm{d}y = 1$
- Marginal distribution of X: $P(\mathrm{x}) = \sum_y P(x \cap y) = \sum_y P(x|y)P(y)$ with the analogous definition in the continuous case.

Example. Marginal distributions

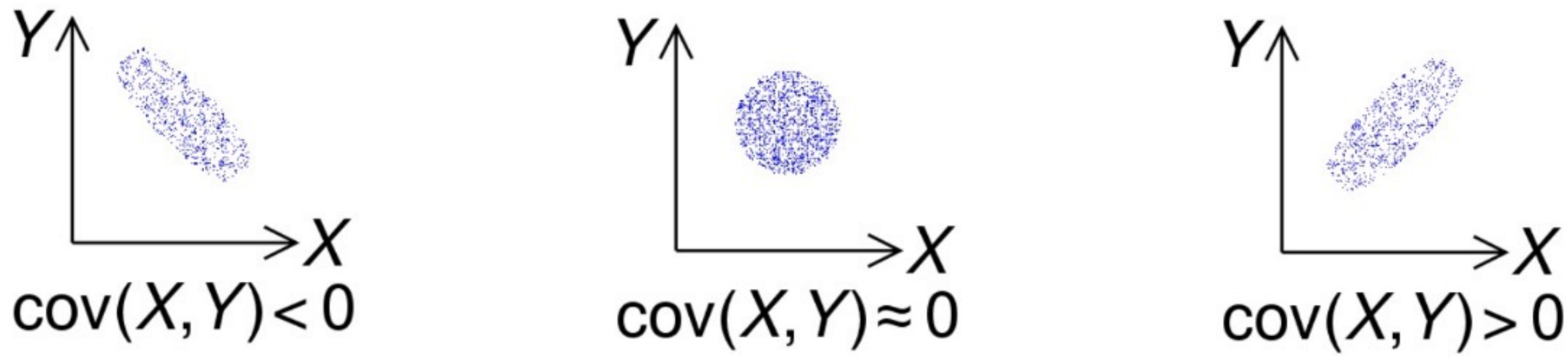
	x_1	x_2	x_3	P_y
y_1	1/16	3/16	5/16	9/16
y_2	2/16	3/16	2/16	7/16
P_x	3/16	6/16	7/16	

Common Distributions

Covariance: $\mathbf{E}((X-\mu_X)(Y-\mu_Y))$

- Measure of whether or not larger values of X correspond to larger values of Y

Example. Covariance and data shape.



Common Distributions

Covariance: $\mathbf{E}((X-\mu_X)(Y-\mu_Y))$

Example. Calculate Covariance.

$f(x,y)$		**x** 5	6	7	$f_Y(y)$
y	**8**	0	0.4	0.1	0.5
	9	0.3	0	0.2	0.5
$f_X(x)$		0.3	0.4	0.3	1

Covariance: $\mathbf{E}((X-\mu_X)(Y-\mu_Y))$

Example. Calculate Covariance.

$f(x,y)$		x = 5	x = 6	x = 7	$f_Y(y)$
y	8	0	0.4	0.1	0.5
	9	0.3	0	0.2	0.5
$f_X(x)$		0.3	0.4	0.3	1

$$\mu_X = 0.3*5+0.4*6+0.3*7=6$$
$$\mu_Y = 0.5*8+0.5*9=8.5$$

$$\begin{aligned}
&\mathbf{E}\big((X-\mu_X)(Y-\mu_Y)\big)\\
&=\sum_{(x,y)}(x-\mu_X)(y-\mu_Y)f(x,y)\\
&=(5-6)(8-8.5)*0+(6-6)(8-8.5)*0.4\\
&+(7-6)(8-8.5)*0.1+(5-6)(9-8.5)*0.3\\
&+(6-6)(9-8.5)*0+(7-6)(9-8.5)*0.2\\
&=-0.1
\end{aligned}$$

Outline

- Course Overview
- Rules of Probability
- Discrete and Continuous Distributions
- Cumulative Distributions
- Common Distributions
- Probability Generating Functions
- Characteristic Functions

Probability Generating Functions

If X is a discrete random variable taking integer values, then the probability generating function (PGF) is

$$G(z) = \mathbf{E}(z^X) = \sum_{k=0}^{\infty} P(k)\, z^k$$

The PGF can be used to calculate all the probabilities of a distribution

$$G(z) = P(0) + P(1)z + P(2)z^2 + \cdots$$

So, to find $P(n)$, simply differentiate n times, divide by $n!$, and set $z = 0$

PGFs uniquely specify a probability distribution

Exercise 1.8. If X is a discrete random variable with PGF $G(z) = \frac{z}{5}\,(2\ +\ 3z^2)$, then what is the distribution of X?

Note that: $G'(1) = \mathbf{E}(\mathrm{X})$:

$$G(z) = P(0) + P(1)z + P(2)z^2 + \cdots$$

$$G'(z) = P(1) + 2P(2)z\ + \cdots$$

$$G'(1) = P(1) + 2P(2)z + \cdots = \sum_{k=0}^{\infty} kP(k)$$

Example. PGF of a geometric random variable.

If X is a geometric RV, then $P(k) = (1-p)^k p$

$$
\begin{aligned}
G(z) = \mathbf{E}(z^X) &= \sum_{k=0}^{\infty} P(k)\, z^k \\
&= \sum_{k=0}^{\infty} (1-p)^k p\, z^k \\
&= p \sum_{k=0}^{\infty} (z(1-p))^k \\
&= \frac{p}{1 - z(1-p)}
\end{aligned}
$$

whenever $|z(1-p)| < 1$

Exercise 1.9. Suppose that England win the football world cup each time it is held with probability 0.1. Use PGFs to find the expected number of years between world cup wins.

Probability Generating Functions

Example 1.6. (board) Probability of an infectious disease epidemic.

If infections happen at rate β and removals happen at rate μ, and the probability of fadeout (no epidemic) starting from a single infectious host is denoted by z, then

$$z = \sum_k P(\text{fadeout}|\text{cause } k \text{ infections})\, P(\text{cause } k \text{ infections})$$

$$= \sum_{k=0}^{\infty} z^k \left(\frac{\beta}{\beta+\mu}\right)^k \left(\frac{\mu}{\beta+\mu}\right)$$

$= G(z)$ for a Geometric RV with probability of failure $1 - p = \frac{\beta}{\beta+\mu}$

$$= \frac{p}{1 - z(1-p)}$$

This can then be solved to find z, the probability of no epidemic.

Probability Generating Functions

PGFs are particularly useful for finding the distribution of sums of independent random variables.

This is because $G_{X+Y}(z) = \mathbf{E}(z^{X+Y})=\mathbf{E}(z^X z^Y)$

$=\mathbf{E}(z^X)\mathbf{E}(z^Y)$ by independence (see board)

$=G_X(z)G_Y(z)$

Outline

- Course Overview
- Rules of Probability
- Discrete and Continuous Distributions
- Cumulative Distributions
- Common Distributions
- Probability Generating Functions
- Characteristic Functions

Characteristic Functions

The moment generating function (MGF) of a random variable X is given by $M(t) = \mathbf{E}(\exp(tX))$.

Similarly to PGFs, the MGF of a sum of independent random variables is simply equal to the product of the individual MGFs.

A related concept is the characteristic function, $\phi(t) = \mathbf{E}(\exp(itX))$. Again, the characteristic function of a sum of independent RVs is equal to the product of the characteristic functions

Characteristic Functions

For a Bernoulli distribution with probability p,

$$\phi(t) = p\exp(it) + (1 - p)$$

Exercise 1.10. What is the characteristic function of a binomial RV?

Characteristic Functions

For a Bernoulli distribution with probability p,

$$\phi(t) = p\exp(it) + (1 - p)$$

Other distributions:

Distribution	characteristic function $\phi(t)$
Bernoulli	$1 - p + pe^{it}$
Binomial	$(1 - p + pe^{it})^n$
Poisson	$e^{\lambda(e^{it}-1)}$
Normal	$e^{it\mu-\sigma^2t^2/2}$
Gamma	$(1 - it\theta)^{-k}$

Characteristic Functions

For a Bernoulli distribution with probability p,

$$\phi(t) = p\exp(it) + (1 - p)$$

Exercise 1.11. What is distribution of a sum of independent:

i) Normally distributed random numbers?
ii) Gamma distributed random numbers with same scale parameter?

Additional Questions

Exercise 1.12. Derive the characteristic function for the Poisson distribution.

Exercise 1.13. Application to neuroscience.
Neuron A makes n synaptic contacts onto neuron B. When neuron A fires, a vesicle containing neurotransmitter is released at each contact with probability p per contact. Each vesicle release contributes to a voltage increase in neuron B that is normally distributed with mean a and variance σ^2. The total increase in voltage is the sum of the effects from each contact plus some Gaussian background noise with zero mean and variance s^2. Hence, the voltage change following one such event is $V = \phi(0, s^2) + \sum_{k=1}^{n} \delta_k \ \phi(a, \sigma^2)$, where δ is a Bernoulli random number with probability p and $\phi(\mu, \sigma^2)$ is a normal random number with mean μ and variance σ^2.

- Show that the distribution for V is given by $\boldsymbol{P}(V) = \sum_{k=1}^{n} \boldsymbol{P}_1 \ (V|k)\boldsymbol{P}_2(k)$, where $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are known distributions to be found.
- Using the values n=10, p=0.3, a=0.2, σ=0.01, s=0.05, write code to plot an empirical distribution of random simulated voltages. Compare them to the theoretical distribution.

Additional Questions

Optional. If time allows...

(Ask for help)

Consider the probability of a major epidemic for the stochastic SIR model, calculated via model simulations. How does this correspond to the probability of a major epidemic derived in the lecture?

Generate appropriate plots.