Probability

Schwartz

September 20, 2017

Statistics, Data Science, and other Languages

- *acrolect*: dialect considered better than others
- *basilect*: dialect considered lower status than others
- *creole*: mixture of a European language combining one or more other languages spoken as a first language
- *diglossia*: which a language exists in a formal or literary form and an informal form with each used situationally
- *interlanguage*: mixture of two languages used by people learning a new language that uses features of the first language mixed with those of the new language
- *lexicon*: the total stock of elements that carry meaning
- *lingua franca*: language used to communicate by people with different first languages
- metalanguage: language for discussing language
- *pidgin*: language of two or more languages for communication by people with different first languages
- *prose*: ordinary written language, as opposed to poetry
- *register*: language used in a particular situation or for communicating with a particular group of people
- *sublanguage*: a variety of language with its own terms and expressions used by a particular group or to talk about a particular subject, for example, the language used by doctors to talk to each other about medicine
- vernacular: language spoken in a particular group/area, when it is different from the formal written language

- Machine languages: interpreted by hardware
- Assembly languages: machine language wrapper
- System languages: low-level tasks like memory/process management (e.g. C/++, Java)
- High-level languages: machine independent*
- Scripting languages: very high-level, expressive, and powerful (Python, R, JavaScript, Perl)
- Embedded languages: executes inside text (e.g. server and client side JavaScript)
- Compiled languages: verify, tee up, then run*
- Interpreted languages: run as is (e.g., Python, R)
- Declarative languages: what, not how (e.g., SQL)
- Imperative languages: execution of serial orders*
- Functional languages: functions only (e.g., Lisp)
- Iterative languages: generators (e.g., Python)
- Object oriented languages: classes*
- Procedural languages: functions*$^{\text{plus, e.g., SQL}}$
- Visual languages: non-text based
- Esoteric languages: not intended to be used

* C/++, Java, Python, R

Objectives

- Probability – ALL of it

Objectives

- Probability – ALL of it

- Counting
- Random Variables
  - Marginal, Joint, and Conditional distributions
- Distributions
  - Representations: pmf/pdf/mgf/characteristic functions
  - Examples: Bernoulli, Binomial, Geometric, Multinomial, Poisson Uniform, Normal, χ^2, Gamma, Exponential, Beta
  - Properties: E, Var, Cov, Cor

Objectives

- Probability – ALL of it

- Counting
- Random Variables
  - Marginal, Joint, and Conditional distributions
- Distributions
  - Representations: pmf/pdf/mgf/characteristic functions
  - Examples: Bernoulli, Binomial, Geometric, Multinomial, Poisson Uniform, Normal, χ^2, Gamma, Exponential, Beta
  - Properties: E, Var, Cov, Cor

- Exposure and comfort with a wide range of sophisticated statistical distribution theory concepts and notations

Why *I* (Scott) think YOU (<you>) should care

If you're

- Pro $\Longrightarrow$ Refresher
- Intermediate $\Longrightarrow$ Solidify
- Noob $\Longrightarrow$ Exposure (and organization for starting points)

Why *I* (Scott) think YOU (<you>) should care

If you're

- Pro $\implies$ Refresher
- Intermediate $\implies$ Solidify
- Noob $\implies$ Exposure (and organization for starting points)

1. Stats is MUSHT in the *<u>data</u> science* world
   - street cred that takes you to the next level and looks impressive
   - show's you're not messing around and you mean to business
   - **like um EVERY interview process will ask this shit**

Why *I* (Scott) think YOU (<you>) should care

If you're

- Pro $\implies$ Refresher
- Intermediate $\implies$ Solidify
- Noob $\implies$ Exposure (and organization for starting points)

1. Stats is MUSHT in the *data science* world
   - street cred that takes you to the next level and looks impressive
   - show's you're not messing around and you mean to business
   - **like um EVERY interview process will ask this shit**
2. Maths allow you to truly, madly, deeply, get methodology
   - know exactly what it is and does inside out – no more, no less

Why *I* (Scott) think YOU (<you>) should care

If you're

- Pro $\implies$ Refresher
- Intermediate $\implies$ Solidify
- Noob $\implies$ Exposure (and organization for starting points)

1. Stats is MUSHT in the *data science* world
   - street cred that takes you to the next level and looks impressive
   - show's you're not messing around and you mean to business
   - **like um EVERY interview process will ask this shit**
2. Maths allow you to truly, madly, deeply, get methodology
   - know exactly what it is and does inside out – no more, no less
3. the whole inference/prediction methodology cosmology *'thing'*
   - it's part of the general theoretical framework/landscape and all

Counting

a.k.a., *combinatorics* – the discipline of mathematics dedicated to *counting*

- Permutation: How many ways can you *permute* things

- Combination: How many ways can you *combine* things

Counting

a.k.a., *combinatorics* – the discipline of mathematics dedicated to *counting*

- Permutation: How many ways can you *permute* things

  Order matters

  E.g., ABC, BAC, CBA, ACB, CAB, BCA

- Combination: How many ways can you *combine* things

Counting

a.k.a., *combinatorics* – the discipline of mathematics dedicated to *counting*

- Permutation: How many ways can you *permute* things

  Order matters
  E.g., ABC, BAC, CBA, ACB, CAB, BCA

  - There are $\frac{n!}{(n-k)!}$ k-sized permutations of n things ($k < n$)

- Combination: How many ways can you *combine* things

Counting

a.k.a., *combinatorics* – the discipline of mathematics dedicated to *counting*

- Permutation: How many ways can you *permute* things

  Order matters
  E.g., ABC, BAC, CBA, ACB, CAB, BCA

  - There are $\frac{n!}{(n-k)!}$ k-sized permutations of n things ($k < n$)

- Combination: How many ways can you *combine* things

  Order *doesn't matter*
  E.g., AB, AC, BC

Counting

a.k.a., *combinatorics* – the discipline of mathematics dedicated to *counting*

- Permutation: How many ways can you *permute* things

  Order matters
  E.g., ABC, BAC, CBA, ACB, CAB, BCA

  - There are $\frac{n!}{(n-k)!}$ k-sized permutations of n things ($k < n$)

- Combination: How many ways can you *combine* things

  Order *doesn't matter*
  E.g., AB, AC, BC

  - The number of k-sized subsets of m things ($k < m$) is

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

Counting: the cartesian product

- If there are m ways event A can happen
- And there are n ways event B can happen... then?

Counting: the cartesian product

- If there are m ways event A can happen
- And there are n ways event B can happen... then?

  Then there are $m \times n$ (A, B) can happen

Counting: the cartesian product

- If there are m ways event A can happen
- And there are n ways event B can happen... then?

  Then there are $m \times n$ (A, B) can happen

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
b_1										
b_2										
b_3										
b_4										
b_5										

Counting: the cartesian product

- If there are m ways event A can happen
- And there are n ways event B can happen... then?

  Then there are $m \times n$ (A, B) can happen

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
b_1										
b_2										
b_3										
b_4										
b_5										

- *How many ways can* $A \cdot B \cdot C$ *happen?*

Problem 1:

- I left messages for my parents and three siblings to call me.

Problem 1:

- I left messages for my parents and three siblings to call me. What's the probability the first two calls are from my parents?
  Assuming the order of callbacks is random...

Problem 1:

- I left messages for my parents and three siblings to call me. What's the probability the first two calls are from my parents? Assuming the order of callbacks is random...

v1

Problem 1:

- I left messages for my parents and three siblings to call me. What's the probability the first two calls are from my parents?
  Assuming the order of callbacks is random...

v1 $\Pr(1^{st}\text{call from parent}) \times \Pr(2^{nd}\text{call from parent}|1^{st}\text{call from parent})$

$$\frac{2}{5} \times \frac{1}{4} = \frac{2}{20} = \frac{1}{10}$$

Problem 1:

- I left messages for my parents and three siblings to call me. What's the probability the first two calls are from my parents?
  Assuming the order of callbacks is random...

v1 $\Pr(1^{st}\text{call from parent}) \times \Pr(2^{nd}\text{call from parent}|1^{st}\text{call from parent})$

$$\frac{2}{5} \times \frac{1}{4} = \frac{2}{20} = \frac{1}{10}$$

v2 Number of ways to (a) order $2P$'s and $3S$'s and (b) get $PPSSS$

Problem 1:

- I left messages for my parents and three siblings to call me. What's the probability the first two calls are from my parents? Assuming the order of callbacks is random...

v1 $\quad \Pr(1^{st}\text{call from parent}) \times \Pr(2^{nd}\text{call from parent}|1^{st}\text{call from parent})$

$$\frac{2}{5} \times \frac{1}{4} = \frac{2}{20} = \frac{1}{10}$$

v2 $\quad$ Number of ways to (a) order $2P$'s and $3S$'s and (b) get $PPSSS$

(a) 5!

Problem 1:

- I left messages for my parents and three siblings to call me. What's the probability the first two calls are from my parents? Assuming the order of callbacks is random...

v1 $\Pr(1^{st}\text{call from parent}) \times \Pr(2^{nd}\text{call from parent}|1^{st}\text{call from parent})$

$$\frac{2}{5} \times \frac{1}{4} = \frac{2}{20} = \frac{1}{10}$$

v2 Number of ways to (a) order $2P$'s and $3S$'s and (b) get $PPSSS$

(a) $5!$ and (b) $2 \times 1 \times 3 \times 2 \times 1 = 2!3!$

Problem 1:

- I left messages for my parents and three siblings to call me. What's the probability the first two calls are from my parents? Assuming the order of callbacks is random...

v1 $\Pr(1^{st}\text{call from parent}) \times \Pr(2^{nd}\text{call from parent}|1^{st}\text{call from parent})$

$$\frac{2}{5} \times \frac{1}{4} = \frac{2}{20} = \frac{1}{10}$$

v2 Number of ways to (a) order $2P$'s and $3S$'s and (b) get $PPSSS$

(a) $5!$ and (b) $2 \times 1 \times 3 \times 2 \times 1 = 2!3! \implies \Pr(PPSSS) = \frac{2!3!}{5!} = \frac{1}{10}$

Problem 1:

- I left messages for my parents and three siblings to call me. What's the probability the first two calls are from my parents? Assuming the order of callbacks is random...

v1 $\Pr(1^{st}\text{call from parent}) \times \Pr(2^{nd}\text{call from parent}|1^{st}\text{call from parent})$

$$\frac{2}{5} \times \frac{1}{4} = \frac{2}{20} = \frac{1}{10}$$

v2 Number of ways to (a) order $2P$'s and $3S$'s and (b) get $PPSSS$

(a) $5!$ and (b) $2 \times 1 \times 3 \times 2 \times 1 = 2!3! \Longrightarrow \Pr(PPSSS) = \frac{2!3!}{5!} = \frac{1}{10}$

v3 One way to have parents call first: $PPSSS$

Problem 1:

- I left messages for my parents and three siblings to call me. What's the probability the first two calls are from my parents? Assuming the order of callbacks is random...

v1 $\Pr(1^{st}\text{call from parent}) \times \Pr(2^{nd}\text{call from parent} | 1^{st}\text{call from parent})$

$$\frac{2}{5} \times \frac{1}{4} = \frac{2}{20} = \frac{1}{10}$$

v2 Number of ways to (a) order $2P$'s and $3S$'s and (b) get $PPSSS$

(a) $5!$ and (b) $2 \times 1 \times 3 \times 2 \times 1 = 2!3! \Longrightarrow \Pr(PPSSS) = \frac{2!3!}{5!} = \frac{1}{10}$

v3 One way to have parents call first: $PPSSS$

How many ways to choose first two $XY\_\,\_\,\_$?

Problem 1:

- I left messages for my parents and three siblings to call me. What's the probability the first two calls are from my parents? Assuming the order of callbacks is random...

v1 $\Pr(1^{st}\text{call from parent}) \times \Pr(2^{nd}\text{call from parent}|1^{st}\text{call from parent})$

$$\frac{2}{5} \times \frac{1}{4} = \frac{2}{20} = \frac{1}{10}$$

v2 Number of ways to (a) order $2P$'s and $3S$'s and (b) get $PPSSS$

(a) $5!$ and (b) $2 \times 1 \times 3 \times 2 \times 1 = 2!3! \Longrightarrow \Pr(PPSSS) = \frac{2!3!}{5!} = \frac{1}{10}$

v3 One way to have parents call first: $PPSSS$

How many ways to choose first two $XY\_\,\_\,\_$?

$$\binom{5}{2} \Longrightarrow \Pr(PPSSS) = 1 \Big/ \binom{5}{2} = \frac{2!3!}{5!} = \frac{1}{10}$$

Problem 2:

There are k=3 instructors and n=5 students

Problem 2:

There are k=3 instructors and n=5 students

The instructors partition student support responsibilities

Problem 2:

There are k=3 instructors and n=5 students

The instructors partition student support responsibilities

[i.e., each instructor only works with “their students”]

Problem 2:

There are k=3 instructors and n=5 students

The instructors partition student support responsibilities

[i.e., each instructor only works with "their students"]

- How many ways are there to divvy up the workloads between the instructors if each instructor has at least one student?

Problem 2:

There are k=3 instructors and n=5 students

The instructors partition student support responsibilities

[i.e., each instructor only works with “their students”]

- How many ways are there to divvy up the workloads between the instructors if each instructor has at least one student?

$$\star \star \star \star \star$$

Problem 2:

There are k=3 instructors and n=5 students

The instructors partition student support responsibilities

[i.e., each instructor only works with “their students”]

- How many ways are there to divvy up the workloads between the instructors if each instructor has at least one student?

$$\star \star \star \star \star$$

$$\star \star \mid \star \mid \star \star$$

Problem 2:

There are k=3 instructors and n=5 students

The instructors partition student support responsibilities

[i.e., each instructor only works with “their students”]

- How many ways are there to divvy up the workloads between the instructors if each instructor has at least one student?

$$\star\ \star\ \star\ \star\ \star$$

$$\star\ \star\ \Big|\ \star\ \Big|\ \star\ \star$$

$$\Big|\ \star\ \star\ \Big|\ \star\ \star\ \star$$

Problem 2:

There are k=3 instructors and n=5 students

The instructors partition student support responsibilities

[i.e., each instructor only works with "their students"]

- How many ways are there to divvy up the workloads between the instructors if each instructor has at least one student?

$$\star \star \star \star \star$$

$$\star \star \mid \star \mid \star \star$$

$$\mid \star \star \mid \star \star \star$$

$$\binom{n+k-1}{k-1} = \binom{7}{2}$$

Problem 2:

There are k=3 instructors and n=5 students

The instructors partition student support responsibilities

[i.e., each instructor only works with “their students”]

- How many ways are there to divvy up the workloads between the instructors if each instructor has at least one student?

$$\star \star \star \star \star$$

$$\star \star \mid \star \mid \star \star$$

$$\mid \star \star \mid \star \star \star$$

$$\binom{n+k-1}{k-1} = \binom{7}{2}$$

The world-famous “stars and bars” solution

The way you get good at solving these types of problems is by knowing what the solutions are

Problem 3:

```
'''
The scenario: Donald Trump becomes president and make the python built-in
              'set' class illegal. What do we do?
'''
```

Problem 3:

$$\Pr(Hillary_{Awesome}) = 0.20$$
$$Pr(Hillary_{Wins} | Hillary_{Awesome}) = 0.98$$
$$Pr(Hillary_{Wins} | Hillary_{\cancel{Awesome}}) = 0.90$$

$$Pr(Hillary_{Awesome} | Hillary_{Wins}) = ?$$

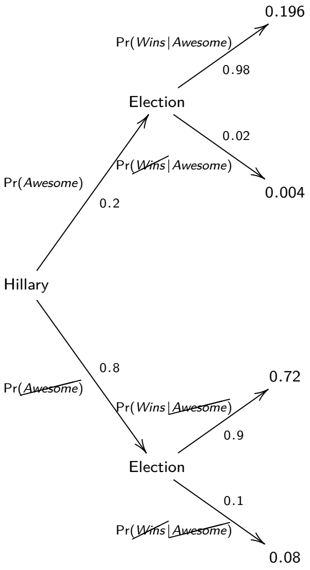
Problem 3:

$$\Pr(Hillary_{Awesome}) = 0.20$$

$$Pr(Hillary_{Wins} | Hillary_{Awesome}) = 0.98$$

$$Pr(Hillary_{Wins} | Hillary_{\cancel{Awesome}}) = 0.90$$

$$Pr(Hillary_{Awesome} | Hillary_{Wins}) = ?$$

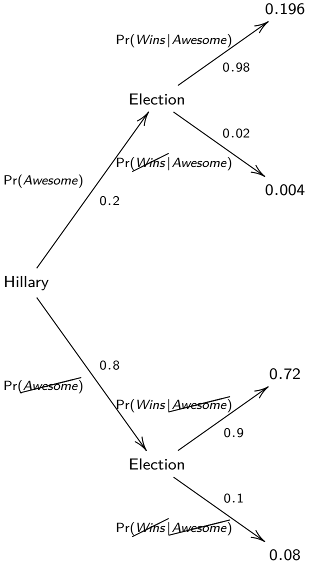


Problem 3:

$$\Pr(Hillary_{Awesome}) = 0.20$$
$$Pr(Hillary_{Wins} | Hillary_{Awesome}) = 0.98$$
$$Pr(Hillary_{Wins} | Hillary_{\cancel{Awesome}}) = 0.90$$

$$Pr(Hillary_{Awesome} | Hillary_{Wins}) = ?$$

$$\Pr(Hillary_{Wins} \& Hillary_{Awesome}) = 0.20 \cdot 0.98$$
$$\Pr(Hillary_{\cancel{Wins}} \& Hillary_{Awesome}) = 0.20 \cdot 0.02$$
$$\Pr(Hillary_{Wins} \& Hillary_{\cancel{Awesome}}) = 0.80 \cdot 0.90$$
$$\Pr(Hillary_{\cancel{Wins}} \& Hillary_{\cancel{Awesome}}) = 0.80 \cdot 0.10$$

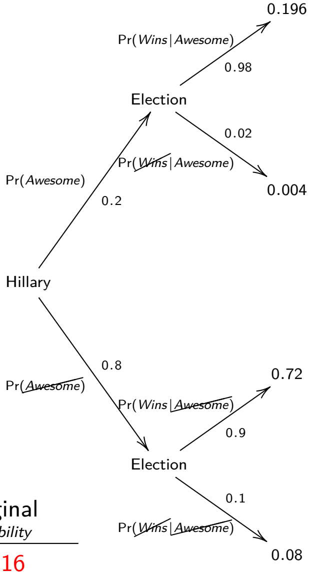


Problem 3:

$$\Pr(Hillary_{Awesome}) = 0.20$$
$$Pr(Hillary_{Wins} | Hillary_{Awesome}) = 0.98$$
$$Pr(Hillary_{Wins} | Hillary_{\cancel{Awesome}}) = 0.90$$

$$Pr(Hillary_{Awesome} | Hillary_{Wins}) = ?$$

$$\Pr(Hillary_{Wins} \& Hillary_{Awesome}) = 0.20 \cdot 0.98$$
$$\Pr(Hillary_{\cancel{Wins}} \& Hillary_{Awesome}) = 0.20 \cdot 0.02$$
$$\Pr(Hillary_{Wins} \& Hillary_{\cancel{Awesome}}) = 0.80 \cdot 0.90$$
$$\Pr(Hillary_{\cancel{Wins}} \& Hillary_{\cancel{Awesome}}) = 0.80 \cdot 0.10$$



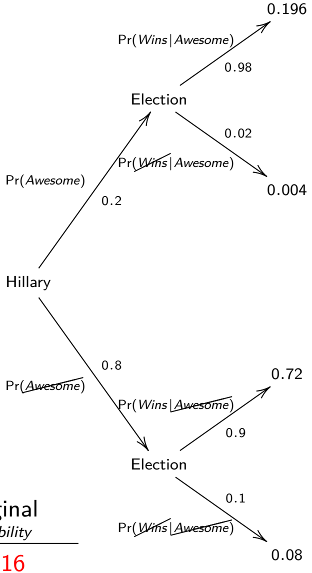
	Awesome Hillary	~~*Awesome*~~ Hillary	Marginal *Probability*
Wins Hillary	0.196	0.720	0.916
~~*Wins*~~ Hillary	0.004	0.080	0.084
Marginal *Probability*	0.200	0.800	1.000

Problem 3:

$$Pr(Hillary_{Awesome}) = 0.20$$
$$Pr(Hillary_{Wins} | Hillary_{Awesome}) = 0.98$$
$$Pr(Hillary_{Wins} | Hillary_{\cancel{Awesome}}) = 0.90$$

$$Pr(Hillary_{Awesome} | Hillary_{Wins}) = ?$$

$$Pr(Hillary_{Wins} \& Hillary_{Awesome}) = 0.20 \cdot 0.98$$
$$Pr(Hillary_{\cancel{Wins}} \& Hillary_{Awesome}) = 0.20 \cdot 0.02$$
$$Pr(Hillary_{Wins} \& Hillary_{\cancel{Awesome}}) = 0.80 \cdot 0.90$$
$$Pr(Hillary_{\cancel{Wins}} \& Hillary_{\cancel{Awesome}}) = 0.80 \cdot 0.10$$



	Awesome Hillary	*~~Awesome~~* Hillary	Marginal *Probability*
Wins Hillary	0.196	0.720	0.916
~~Wins~~ Hillary	0.004	0.080	0.084
Marginal *Probability*	0.200	0.800	1.000

$$\frac{0.196}{0.916}$$

Problem 3:

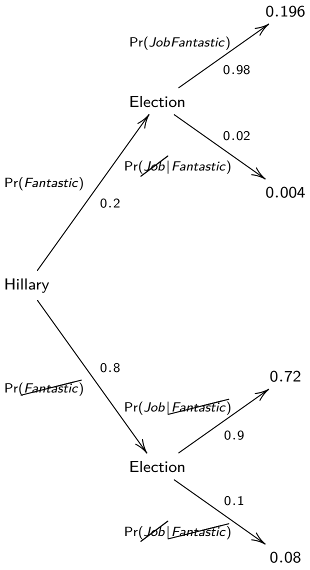
$$
\begin{aligned}
\mathrm{Pr}(Trump_{Fantastic}) &= 0.20 \\
Pr(DS_{Job} | Trump_{Fantastic}) &= 0.98 \\
Pr(DS_{Job} | Trump_{\cancel{Fantastic}}) &= 0.90
\end{aligned}
$$

$$
Pr(Trump_{Fantastic} | DS_{Job}) = ?
$$

Problem 3:

$$Pr(Trump_{Fantastic}) = 0.20$$
$$Pr(DS_{Job}|Trump_{Fantastic}) = 0.98$$
$$Pr(DS_{Job}|Trump_{\cancel{Fantastic}}) = 0.90$$

$$Pr(Trump_{Fantastic}|DS_{Job}) = ?$$

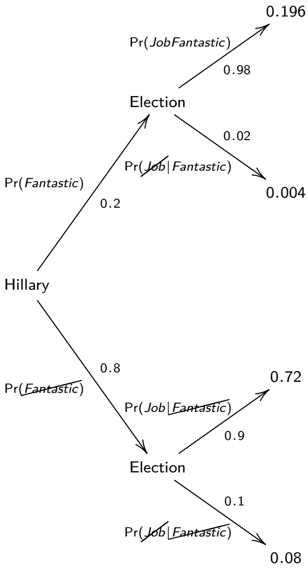


Problem 3:

$$\Pr(Trump_{Fantastic}) = 0.20$$
$$Pr(DS_{Job}|Trump_{Fantastic}) = 0.98$$
$$Pr(DS_{Job}|Trump_{\cancel{Fantastic}}) = 0.90$$

$$Pr(Trump_{Fantastic}|DS_{Job}) = ?$$

$$\Pr(DS_{Job} \& Trump_{Fantastic}) = 0.20 \cdot 0.98$$
$$\Pr(DS_{\cancel{Job}} \& Trump_{Fantastic}) = 0.20 \cdot 0.02$$
$$\Pr(DS_{Job} \& Trump_{\cancel{Fantastic}}) = 0.80 \cdot 0.90$$
$$\Pr(DS_{\cancel{Job}} \& Trump_{\cancel{Fantastic}}) = 0.80 \cdot 0.10$$

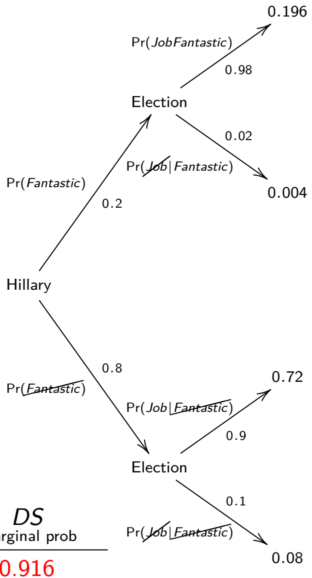


Problem 3:

$$\Pr(Trump_{Fantastic}) = 0.20$$
$$Pr(DS_{Job}|Trump_{Fantastic}) = 0.98$$
$$Pr(DS_{Job}|Trump_{\cancel{Fantastic}}) = 0.90$$

$$Pr(Trump_{Fantastic}|DS_{Job}) = ?$$

$$\Pr(DS_{Job}\& Trump_{Fantastic}) = 0.20 \cdot 0.98$$
$$\Pr(DS_{\cancel{Job}}\& Trump_{Fantastic}) = 0.20 \cdot 0.02$$
$$\Pr(DS_{Job}\& Trump_{\cancel{Fantastic}}) = 0.80 \cdot 0.90$$
$$\Pr(DS_{\cancel{Job}}\& Trump_{\cancel{Fantastic}}) = 0.80 \cdot 0.10$$



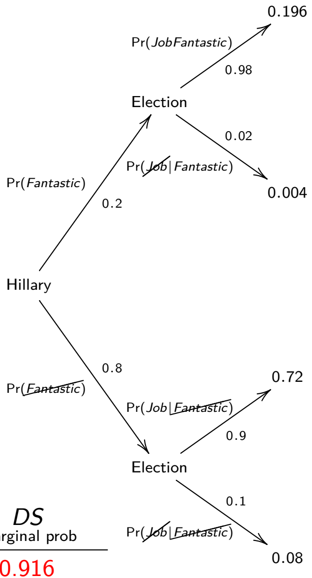
	Trump Fantastic	*Trump* $\cancel{Fantastic}$	*DS* Marginal prob
DS Job	0.196	0.720	0.916
DS $\cancel{Job}$	0.004	0.080	0.084
Trump Marginal prob	0.200	0.800	1.000

Problem 3:

$$\Pr(Trump_{Fantastic}) = 0.20$$
$$Pr(DS_{Job}|Trump_{Fantastic}) = 0.98$$
$$Pr(DS_{Job}|Trump_{\cancel{Fantastic}}) = 0.90$$

$$Pr(Trump_{Fantastic}|DS_{Job}) = ?$$

$$\Pr(DS_{Job} \& Trump_{Fantastic}) = 0.20 \cdot 0.98$$
$$\Pr(DS_{\cancel{Job}} \& Trump_{Fantastic}) = 0.20 \cdot 0.02$$
$$\Pr(DS_{Job} \& Trump_{\cancel{Fantastic}}) = 0.80 \cdot 0.90$$
$$\Pr(DS_{\cancel{Job}} \& Trump_{\cancel{Fantastic}}) = 0.80 \cdot 0.10$$



	$Trump$ Fantastic	$Trump$ ~~Fantastic~~	DS Marginal prob
DS Job	0.196	0.720	0.916
DS ~~Job~~	0.004	0.080	0.084
$Trump$ Marginal prob	0.200	0.800	1.000

$$\frac{0.196}{0.916}$$

Problem 4:

Three types of fair coins are in an urn:

HH, HT, and TT

You pull a coin, flip it, and it comes up H

- What is the probability it comes up H if you flip it a second time?

Random Variables

- Random Variable X can take on values in the *sample space* $\mathcal{S}$

Random Variables

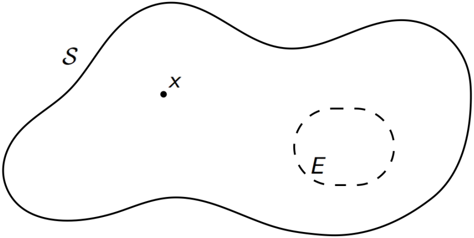
- Random Variable X can take on values in the *sample space* $\mathcal{S}$
- The actualized values x of X are called *outcomes*, and $x \in \mathcal{S}$

Random Variables

- Random Variable X can take on values in the *sample space* $\mathcal{S}$
- The actualized values x of X are called *outcomes*, and $x \in \mathcal{S}$
- An event E is a subset of the sample space, i.e., $E \subset \mathcal{S}$

Random Variables

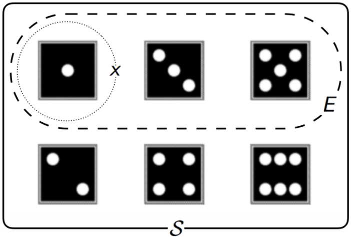
- Random Variable X can take on values in the *sample space* $\mathcal{S}$
- The actualized values x of X are called *outcomes*, and $x \in \mathcal{S}$
- An event E is a subset of the sample space, i.e., $E \subset \mathcal{S}$



Support space $\mathcal{S}$, event E, and outcome x for random variable X

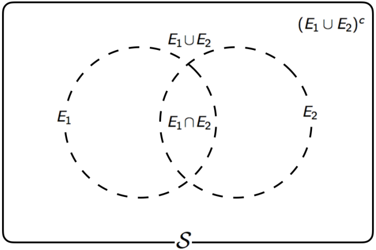
Random Variables

- Random Variable X can take on values in the *sample space* $\mathcal{S}$
- The actualized values x of X are called *outcomes*, and $x \in \mathcal{S}$
- An event E is a subset of the sample space, i.e., $E \subset \mathcal{S}$



Support space $\mathcal{S}$, event E, and outcome x for random variable X

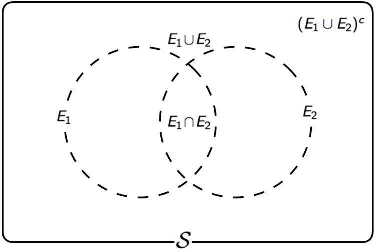
Random Variables: Obvious Rules



Venn Diagram

- $\Pr(E^c) = 1 - \Pr(E)$

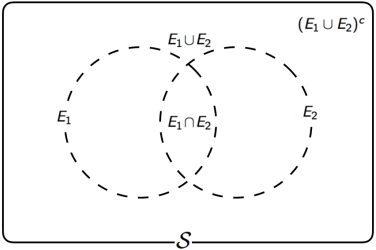
Random Variables: Obvious Rules



Venn Diagram

- $\Pr(E^c) = 1 - \Pr(E)$
- $\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2)$

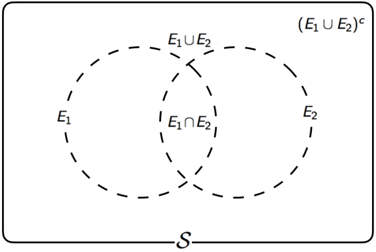
Random Variables: Obvious Rules



Venn Diagram

- $\Pr(E^c) = 1 - \Pr(E)$
- $\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2)$
- $\Pr(E_1 \cap E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cup E_2)$

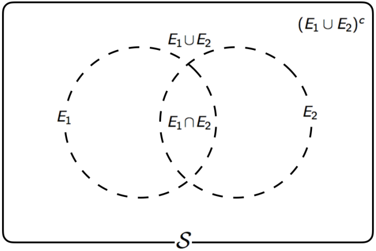
Random Variables: Obvious Rules



Venn Diagram

- $\Pr(E^c) = 1 - \Pr(E)$
- $\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2)$
- $\Pr(E_1 \cap E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cup E_2)$
- $\Pr(E \cap E^c) = 0$

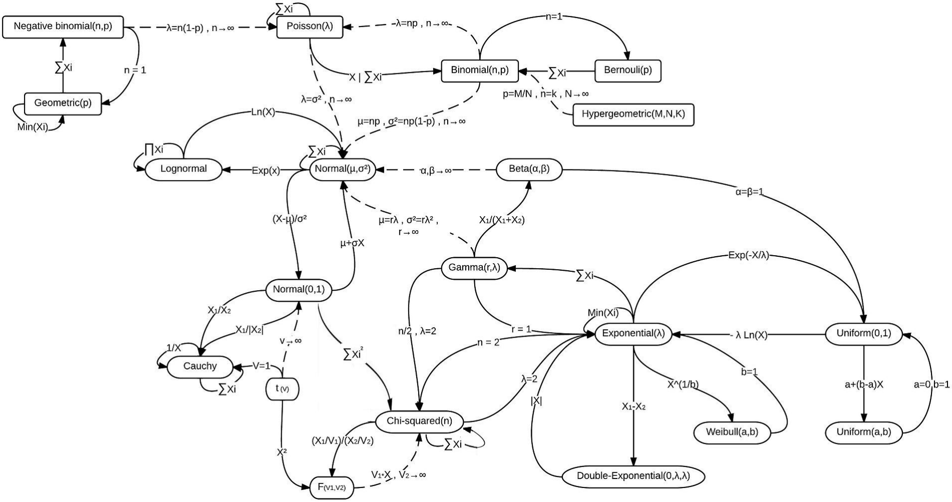
Random Variables: Obvious Rules



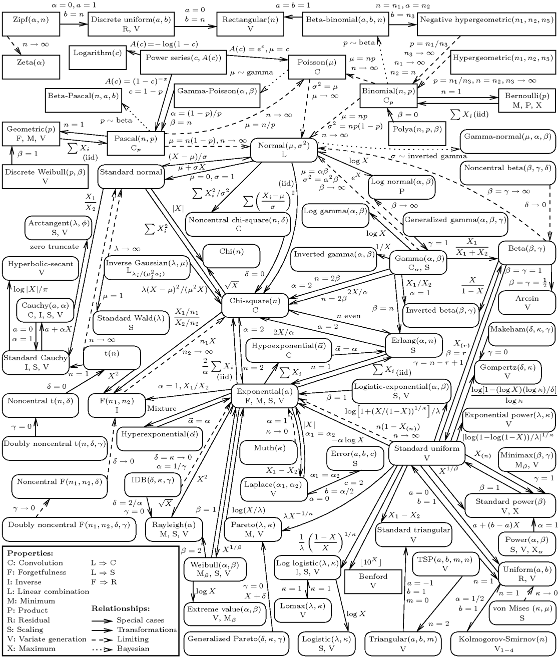
Venn Diagram

- $\Pr(E^c) = 1 - \Pr(E)$
- $\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2)$
- $\Pr(E_1 \cap E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cup E_2)$
- $\Pr(E \cap E^c) = 0$
- DeMorgan's Laws
  - $\Pr\left((E_1 \cup E_2)^c\right) = Pr\left(E_1^c \cap E_2^c\right)$
  - $\Pr\left((E_1 \cap E_2)^c\right) = Pr\left(E_1^c \cup E_2^c\right)$

Distributions



More Distributions



Discrete Distributions: *Bernoulli*

The "coin flip"

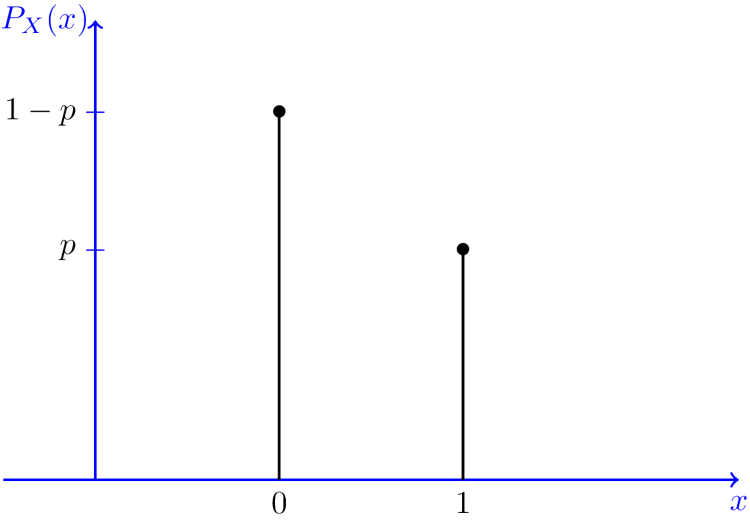
$$x \in \{0, 1\}$$

$$\Pr(X = x|p) = p^x(1-p)^{1-x}$$

$$p \in [0, 1]$$

Discrete Distributions: *Bernoulli*

$$X \sim Bernoulli(p)$$



Discrete Distributions: *Geometric*

The "how many times until"

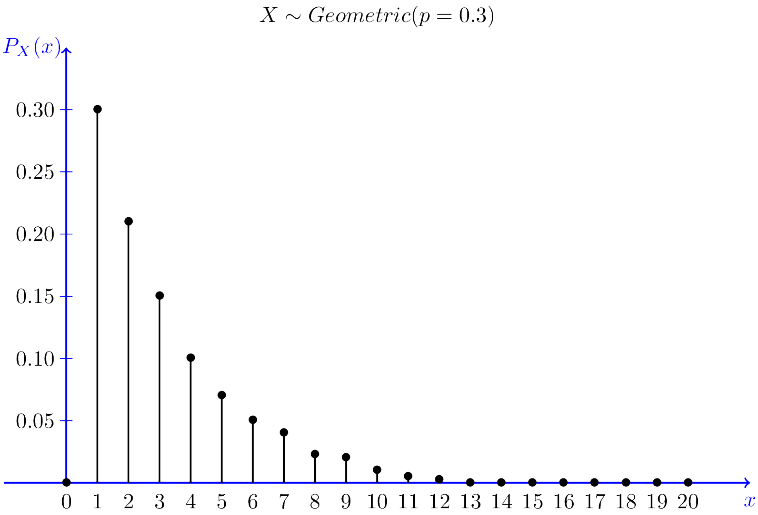
$$x \in \{0, 1, \cdots \infty\}$$

$$\Pr(X = x|p) = (1-p)^{x-1}p$$

$$p \in [0, 1]$$

"If at first you don't succeed, Try, try, try again" – William Edward Hickson

Discrete Distributions: *Geometric*



Discrete Distributions: *Binomial*

The "number of success in *n* trials"

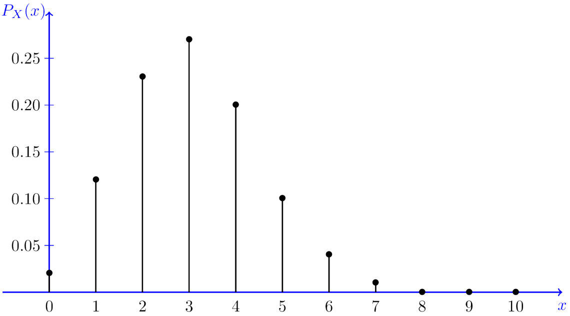
$$x \in \{1, 2, \cdots n\}$$

$$\Pr(X = x|p, n) = \binom{n}{x} p^x (1-p)^{n-x}$$

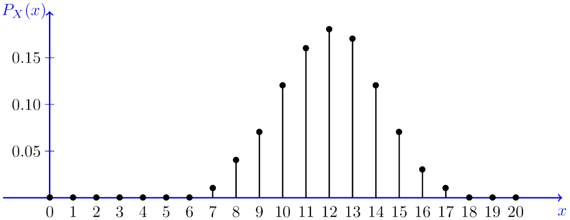
$$p \in [0, 1]$$

Discrete Distributions: *Binomial*

$X \sim Binomial(n = 10, p = 0.3)$



$X \sim Binomial(n = 20, p = 0.6)$



Discrete Distributions: *Multinomial*

The "fancy binomial"

$$\mathbf{x} = (x_1, x_2, \cdots, x_k)$$
$$x_j \in \{0, 1, \cdots m\} \text{ such that } \sum x_j = m$$

$$\Pr(\mathbf{X} = \mathbf{x} | \theta_1, \theta_2, \cdots \theta_k, m) = \frac{m!}{x_1! x_2! \cdots x_k!} \prod_{j=1}^{k} \theta_j^{x_j}$$

$$\theta_j \in [0, 1] \text{ such that } \sum \theta_j = 1$$

Discrete Distributions: *Poisson*

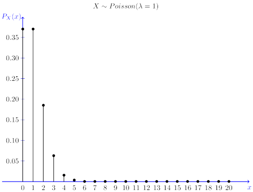
The "number of arrivals"

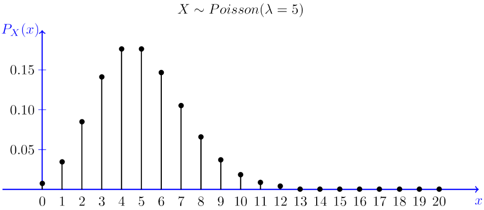
$$k \in \{0, 1, \cdots \infty\}$$

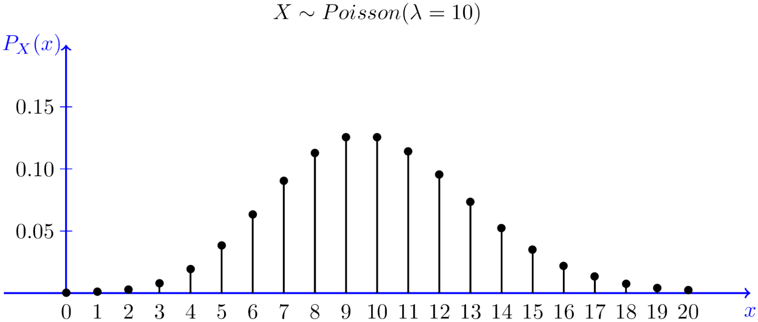
$$\Pr(X = k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$\lambda \in \mathbb{R}^+$$

Discrete Distributions: *Poisson*







Discrete Distributions: *Poisson* $\approx$ *Bionomial*

- If $n\theta = \lambda$ then $\theta = \frac{\lambda}{n}$ so that

$$
\begin{aligned}
&= \binom{n}{k} \theta^k (1-\theta)^{n-k} \\
&= \frac{n \cdot (n-1) \cdots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n \cdot (n-1) \cdots (n-k+1)}{n^k k!} \lambda^k \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&\approx \frac{1}{k!} \lambda^k e^{-\lambda} 1 \ (\text{as } n \to \infty) \\
&= \frac{\lambda^k e^{-\lambda}}{k!}
\end{aligned}
$$

Continuous Distributions: *Uniform*

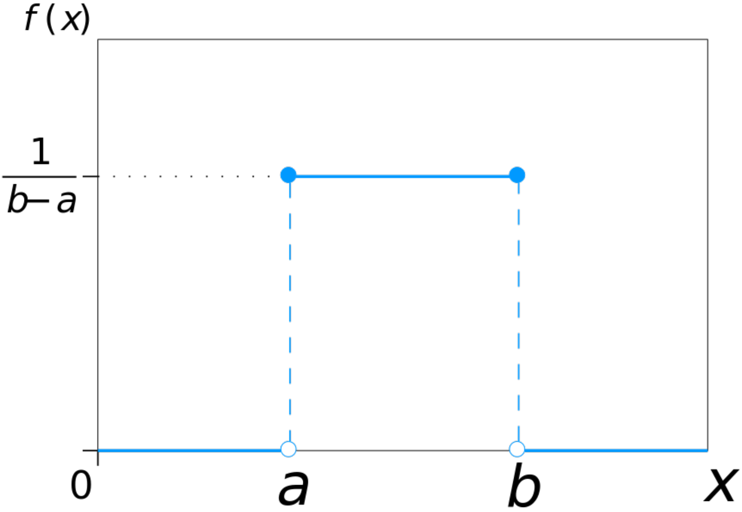
The "random continuous number"

$$u \in \mathbb{R}$$

$$f(X = u|a, b) = \frac{1}{b - a} 1_{[a,b]}(u)$$

$$a, b \in \mathbb{R}, a < b$$

Discrete Distributions: *Uniform*



Continuous Distributions: *Normal*

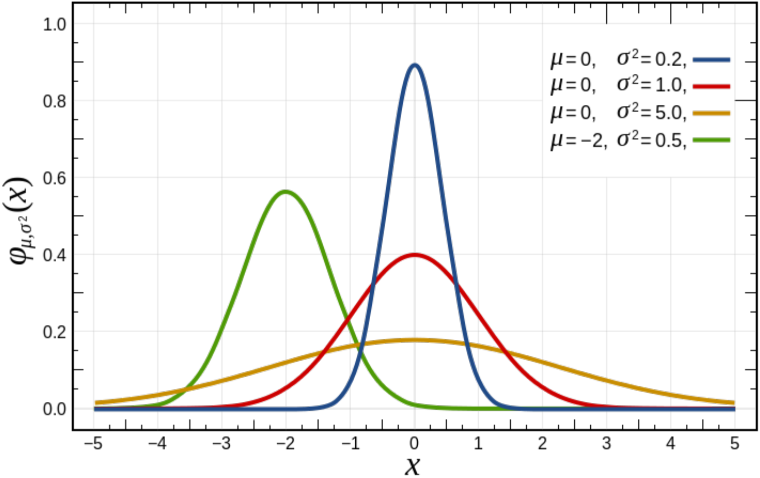
The "bell curve"

$$x \in \mathbb{R}$$

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

$$\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+$$

Continuous Distributions: *Normal*



Continuous Distributions: *Normal*^{2}

- If $X_j \sim$ *Normal* (μ, σ^2) are normal random variables

Continuous Distributions: *Normal*[2]

- If $X_j \sim \mathit{Normal}\,(\mu, \sigma^2)$ are normal random variables

$$Z_j = \frac{X_i - \mu}{\sigma}$$

Continuous Distributions: *Normal*[2]

- If $X_j \sim \textit{Normal}\left(\mu, \sigma^2\right)$ are normal random variables

$$Z_j = \frac{X_i - \mu}{\sigma}$$

are standard normal random variables

Continuous Distributions: *Normal*[2]

- If $X_j \sim Normal\left(\mu, \sigma^2\right)$ are normal random variables

$$Z_j = \frac{X_i - \mu}{\sigma}$$

are standard normal random variables

$$\text{i.e., } Z_j \sim Normal\left(0, 1\right), \text{ and}$$

Continuous Distributions: *Normal*[2]

- If $X_j \sim Normal\left(\mu, \sigma^2\right)$ are normal random variables

  $$Z_j = \frac{X_i - \mu}{\sigma}$$

  are standard normal random variables

  $$\text{i.e., } Z_j \sim Normal\,(0, 1)\,, \text{ and}$$

  $$\sum_{j=1}^{k} Z_i^2 \sim \chi_k^2$$

Continuous Distributions: *Normal*[2]

- If $X_j \sim Normal\left(\mu, \sigma^2\right)$ are normal random variables

  $$Z_j = \frac{X_i - \mu}{\sigma}$$

  are standard normal random variables

  $$\text{i.e., } Z_j \sim Normal\left(0, 1\right), \text{ and}$$

  $$\sum_{j=1}^{k} Z_i^2 \sim \chi_k^2$$

  is *chi-squared* random variable with k “degrees of freedom”

Continuous Distributions: *Normal*[2]

- If $X_j \sim Normal\left(\mu, \sigma^2\right)$ are normal random variables

$$Z_j = \frac{X_i - \mu}{\sigma}$$

are standard normal random variables

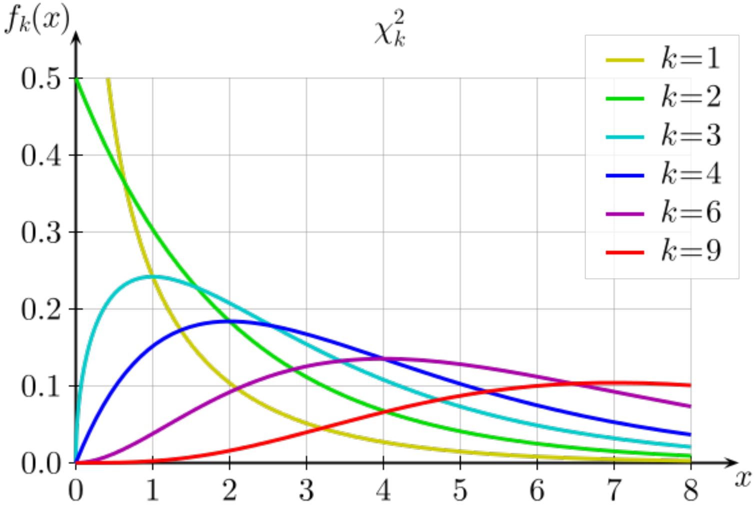
i.e., $Z_j \sim Normal\left(0, 1\right)$, and

$$\sum_{j=1}^{k} Z_i^2 \sim \chi_k^2$$

is *chi-squared* random variable with k "degrees of freedom"

The χ_{df}^2 distribution is a key distribution in hypothesis testing

Continuous Distributions: *Normal*^{2} : χ^2_{df}



Continuous Distributions: *Gamma*

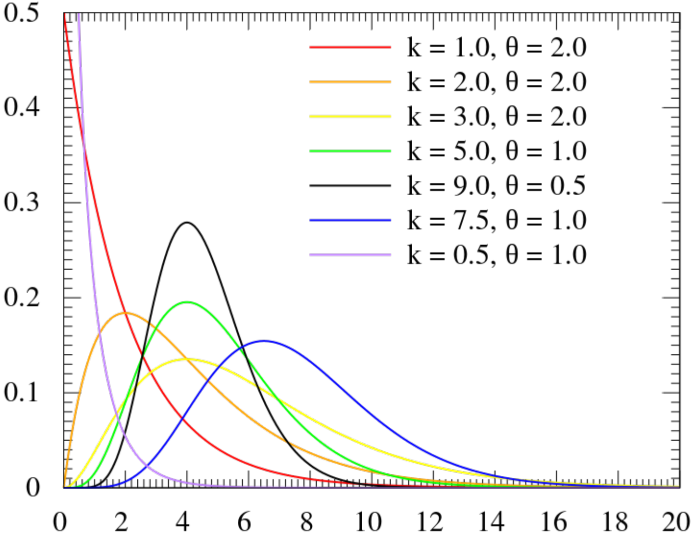
The "Bayesian model for variance"

$$x \in \mathbb{R}^+$$

$$f(X = x|\theta, k) = \frac{\theta^k}{\Gamma(k)} x^{k-1} e^{-x\theta}$$

$$\theta \in \mathbb{R}^+$$

Continuous Distributions: *Gamma*



Continuous Distributions: *Gamma ($\theta = 1/2$, Chi-squared)*

- We previously derived the χ^2_{df} distribution as the "sum of squared standard normal distributions"
- and noted its importance in hypothesis testing

The χ^2_{df} is also a special case of the gamma distribution

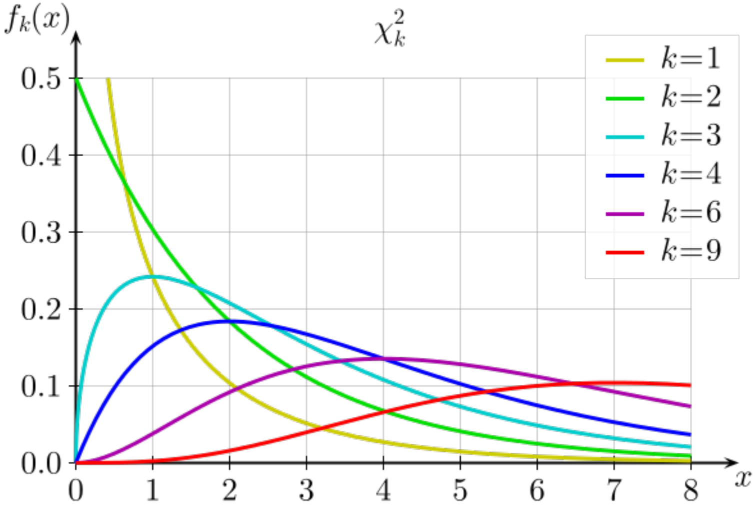
$$x \in \mathbb{R}^+$$

$$f(X = x|k) = \frac{\frac{1}{2}^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$$

$$k \in \mathbb{R}^+$$

Bonus: if $X \sim \chi^2_v$ and $X \sim \chi^2_w$, then $\frac{\frac{1}{v}\chi^2_v}{\frac{1}{w}\chi^2_w} \sim F_{v,w}$

Continuous Distributions: *Gamma* ($\theta = 1/2, \chi^2_{df}$)



Continuous Distributions: *Gamma (k=1, Exponential)*

The *Exponential* is another special case of the gamma

$$x \in \mathbb{R}^{+}$$

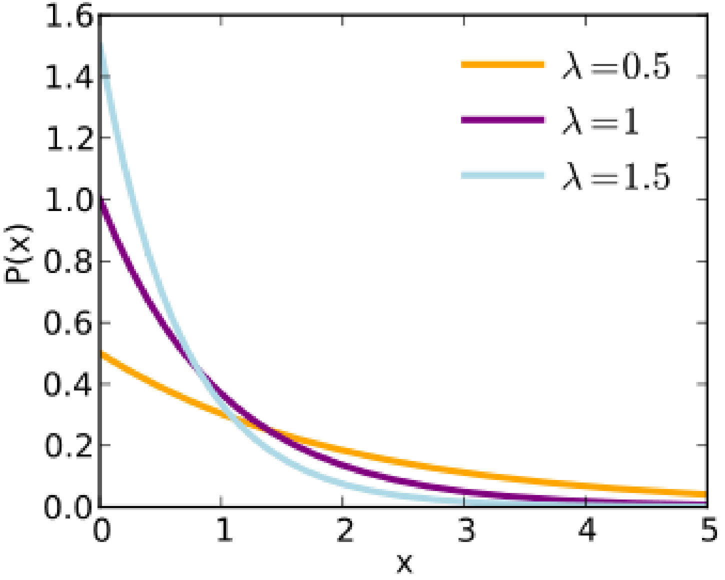
$$f\left(X = x|\theta\right) = \theta e^{-x\theta}$$

$$\theta \in \mathbb{R}^{+}$$

- The Exponential is often used to model time to failure
- It has an interesting "ageless" property, however, in that

  $\Pr(X = x + c|x = 0) = \Pr(X = x + c|x)$ for any value of x

Continuous Distributions: *Gamma (k=1, Exponential)*



Continuous Distributions: *Beta*

The "distribution for modeling random probabilities"

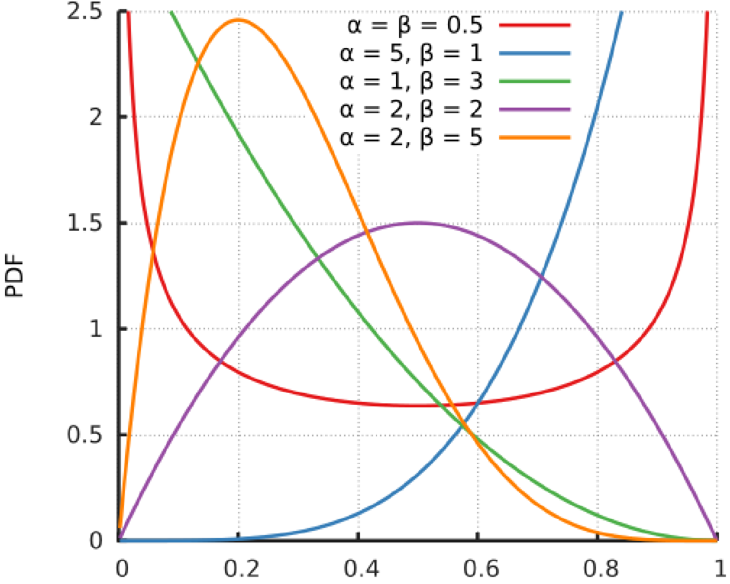
$$p \in [0, 1]$$

$$f(X = p|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1 - p)^{\beta-1}$$

$$\alpha, \beta \in \mathbb{R}^+$$

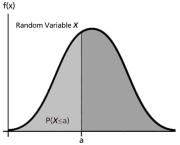
$\alpha = \beta = 1$ results in a *uniform distribution* over the unit interval

Continuous Distributions: *Beta*

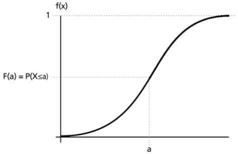


PDF/*PMF*, *CDF*, and characteristic functions

1. We have thus far defined the distribution of a random variable in terms of the *PDF and PMF*

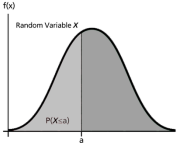


2. We can equivalently alternatively define the distribution of X by it's *cumulative distribution function*

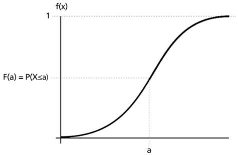


PDF / *PMF*, *CDF*, and characteristic functions

1. We have thus far defined the distribution of a random variable in terms of the *PDF and PMF*



2. We can equivalently alternatively define the distribution of X by it's *cumulative distribution function*

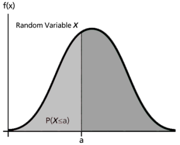


3. Yet another way to define the distribution of X is by its *moment generating function* or its *characteristic function*:

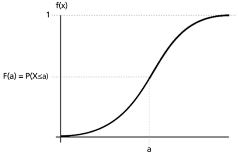
$$\mathrm{E}\left[tX\right], \text{ or } \mathrm{E}\left[itX\right], \text{ respectively}$$

PDF / PMF, *CDF*, and characteristic functions

1. We have thus far defined the distribution of a random variable in terms of the *PDF and PMF*



2. We can equivalently alternatively define the distribution of X by it's *cumulative distribution function*



3. Yet another way to define the distribution of X is by its *moment generating function* or its *characteristic function*:

$$\mathrm{E}\left[tX\right], \text{ or } \mathrm{E}\left[itX\right], \text{ respectively}$$

Interestingly, the characteristic function of $X + Y$ for *independent* random variables X and Y is the product of the characteristic functions of X and Y

Discrete Distributions: *Poisson* + *Poisson*

- The characteristic function of a *Poisson* random variable is

$$e^{\lambda\left(e^{it}-1\right)}$$

Discrete Distributions: *Poisson* + *Poisson*

- The characteristic function of a *Poisson* random variable is

$$e^{\lambda\left(e^{it}-1\right)}$$

- For X and Y Poisson distributed random variables with parameters λ_X and λ_Y, the characteristic function of $X+Y$ is

$$e^{\lambda_X\left(e^{it}-1\right)}e^{\lambda_Y\left(e^{it}-1\right)} = e^{(\lambda_X+\lambda_Y)\left(e^{it}-1\right)}$$

Discrete Distributions: *Poisson* + *Poisson*

- The characteristic function of a *Poisson* random variable is

$$e^{\lambda\left(e^{it}-1\right)}$$

- For X and Y Poisson distributed random variables with parameters λ_X and λ_Y, the characteristic function of $X+Y$ is

$$e^{\lambda_X\left(e^{it}-1\right)}e^{\lambda_Y\left(e^{it}-1\right)} = e^{(\lambda_X+\lambda_Y)\left(e^{it}-1\right)}$$

  So $X+Y$ is Poisson distributed with parameter $\lambda_X+\lambda_Y$

Discrete Distributions: *Poisson* + *Poisson*

- The characteristic function of a *Poisson* random variable is

$$e^{\lambda\left(e^{it}-1\right)}$$

- For X and Y Poisson distributed random variables with parameters λ_X and λ_Y, the characteristic function of $X+Y$ is

$$e^{\lambda_X\left(e^{it}-1\right)}e^{\lambda_Y\left(e^{it}-1\right)} = e^{(\lambda_X+\lambda_Y)\left(e^{it}-1\right)}$$

  So $X+Y$ is Poisson distributed with parameter $\lambda_X+\lambda_Y$

- Quiz: name the distributions of $X+X$ and $Y+Y$ if

$$X \sim \mathit{Bernoulli}(\theta) \text{ and } Y \sim \mathit{Binomial}(\theta, n)$$

  with respective characteristic functions

$$1-\theta+\theta e^{it} \text{ and } \left(1-\theta+\theta e^{it}\right)^n$$

Continuous Distributions: *Normal* + *Normal*

- The characteristic function of a normal random variable is

$$e^{it\mu - \frac{1}{2}t^2\sigma^2}$$

Continuous Distributions: *Normal* + *Normal*

- The characteristic function of a normal random variable is

$$e^{it\mu - \frac{1}{2}t^2\sigma^2}$$

What is the distribution of $X + Y$ if

$$X \sim \mathit{Normal}\left(\mu_X, \sigma_X^2\right) \text{ and } Y \sim \mathit{Normal}\left(\mu_Y, \sigma_Y^2\right)?$$

Continuous Distributions: $X_1 + X_2 + \cdots + X_n \sim Normal$

- The moment generating function (MGF) of a normal random variable is

$$e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

Continuous Distributions: $X_1 + X_2 + \cdots + X_n \sim Normal$

- The moment generating function (MGF) of a normal random variable is

$$e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

The MGF of the sum of n *arbitrary* $i.i.d.$ random variables $\sum_{i=1}^{n} X_i$ is

$$\left(1 + t\mathrm{E}[X] + \frac{t^2\mathrm{E}[X^2]}{2!} + \frac{t^3\mathrm{E}[X^3]}{3!} + \cdots\right)^n$$

Continuous Distributions: $X_1 + X_2 + \cdots + X_n \sim Normal$

- The moment generating function (MGF) of a normal random variable is

$$e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

The MGF of the sum of n *arbitrary* $i.i.d.$ random variables $\sum_{i=1}^{n} X_i$ is

$$\left(1 + t\mathrm{E}[X] + \frac{t^2\mathrm{E}[X^2]}{2!} + \frac{t^3\mathrm{E}[X^3]}{3!} + \cdots\right)^n$$

Using $\log\left((1+u)^n\right) = n\left(u - \frac{u^2}{2!} + \frac{u^3}{3!} - \cdots\right)$, we have

Continuous Distributions: $X_1 + X_2 + \cdots + X_n \sim \mathit{Normal}$

- The moment generating function (MGF) of a normal random variable is

$$e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

The MGF of the sum of n *arbitrary i.i.d.* random variables $\sum_{i=1}^{n} X_i$ is

$$\left(1 + t\mathrm{E}[X] + \frac{t^2\mathrm{E}[X^2]}{2!} + \frac{t^3\mathrm{E}[X^3]}{3!} + \cdots\right)^n$$

Using $\log\left((1+u)^n\right) = n\left(u - \frac{u^2}{2!} + \frac{u^3}{3!} - \cdots\right)$, we have

$$\exp\left(n\left(t\mathrm{E}[X] + \frac{t^2\mathrm{E}[X^2]}{2} + \cdots - \frac{t^2\mathrm{E}[X]^2}{2} - \cdots + \cdots\right)\right)$$

$$= e^{tn\mathrm{E}[X] + \frac{1}{2}t^2 n(\mathrm{E}[X^2] - \mathrm{E}[X]^2) + n(\cdots)}$$

$$\approx e^{tn\mathrm{E}[X] + \frac{1}{2}t^2 n(\mathrm{E}[X^2] - \mathrm{E}[X]^2)} \text{ as } n \to \infty$$

Continuous Distributions: $X_1 + X_2 + \cdots + X_n \sim Normal$

- The moment generating function (MGF) of a normal random variable is

$$e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

The MGF of the sum of n *arbitrary* $i.i.d.$ random variables $\sum_{i=1}^{n} X_i$ is

$$\left(1 + t\mathrm{E}[X] + \frac{t^2\mathrm{E}[X^2]}{2!} + \frac{t^3\mathrm{E}[X^3]}{3!} + \cdots\right)^n$$

Using $\log\left((1+u)^n\right) = n\left(u - \frac{u^2}{2!} + \frac{u^3}{3!} - \cdots\right)$, we have

$$\exp\left(n\left(t\mathrm{E}[X] + \frac{t^2\mathrm{E}[X^2]}{2} + \cdots - \frac{t^2\mathrm{E}[X]^2}{2} - \cdots + \cdots\right)\right)$$

$$= e^{tn\mathrm{E}[X] + \frac{1}{2}t^2 n(\mathrm{E}[X^2] - \mathrm{E}[X]^2) + n(\cdots)}$$

$$\approx e^{tn\mathrm{E}[X] + \frac{1}{2}t^2 n(\mathrm{E}[X^2] - \mathrm{E}[X]^2)} \text{ as } n \to \infty \quad \text{so } \sum_{i=1}^{n} X_i \sim N(n\mathrm{E}[X], n(\mathrm{E}[X^2] - \mathrm{E}[X]^2))$$

Continuous Distributions: $X_1 + X_2 + \cdots + X_n \sim Normal$

- The moment generating function (MGF) of a normal random variable is

$$e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

The MGF of the sum of n *arbitrary i.i.d.* random variables $\sum_{i=1}^{n} X_i$ is

$$\left(1 + t\mathrm{E}[X] + \frac{t^2\mathrm{E}[X^2]}{2!} + \frac{t^3\mathrm{E}[X^3]}{3!} + \cdots\right)^n$$

Using $\log\left((1+u)^n\right) = n\left(u - \frac{u^2}{2!} + \frac{u^3}{3!} - \cdots\right)$, we have

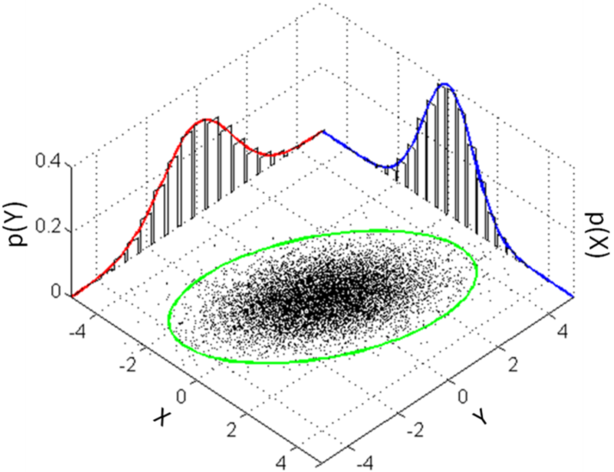
$$\exp\left(n\left(t\mathrm{E}[X] + \frac{t^2\mathrm{E}[X^2]}{2} + \cdots - \frac{t^2\mathrm{E}[X]^2}{2} - \cdots + \cdots\right)\right)$$

$$= e^{tn\mathrm{E}[X] + \frac{1}{2}t^2 n(\mathrm{E}[X^2] - \mathrm{E}[X]^2) + n(\cdots)}$$

$$\approx e^{tn\mathrm{E}[X] + \frac{1}{2}t^2 n(\mathrm{E}[X^2] - \mathrm{E}[X]^2)} \text{ as } n \to \infty \quad \text{so } \sum_{i=1}^{n} X_i \sim N(n\mathrm{E}[X], n(\mathrm{E}[X^2] - \mathrm{E}[X]^2))$$

The binomial distribution with large n is approximately normal: why?

Joint Distributions



This multivariate normal joint distribution models the strength of relationship between X and Y with a correlation parameter ρ and it's marginal distributions that are themselves normally distributed

Joint Distributions

- *Errata:* joint distribution **is allowed/encouraged** on campus

Joint Distributions

- *Errata:* joint distribution **is allowed/encouraged** on campus

Joint distributions are just a collection of random variables that may or may not have some dependencies on each other

$$\Pr(X_1, X_2, \cdots, X_k)$$

Joint Distributions

- *Errata:* joint distribution **is allowed/encouraged** on campus

Joint distributions are just a collection of random variables that may or may not have some dependencies on each other

$$\Pr(X_1, X_2, \cdots, X_k)$$

- We actually already saw an example of a *joint distribution*

$$\Pr(\mathbf{X} = \mathbf{x} | \theta_1, \theta_2, \cdots \theta_k, m) = \frac{m!}{x_1! x_2! \dots x_k!} \prod_{j=1}^{k} \theta_j^{x_j}$$

Do these X_j's have some *dependence*?

Joint Distributions: *discrete*

- *Joint distributions* factor as *conditional* $\times$ *marginal* dist.'s

$$\Pr(\mathbf{X}_1, \mathbf{X}_2) = \Pr(\mathbf{X}_1 | \mathbf{X}_2)\,\Pr(\mathbf{X}_2)$$

Joint Distributions: *discrete*

- *Joint distributions* factor as *conditional* $\times$ *marginal* dist.'s

$$\Pr(\mathbf{X}_1, \mathbf{X}_2) = \Pr(\mathbf{X}_1|\mathbf{X}_2)\Pr(\mathbf{X}_2)$$

- *Bayes theorem* is derived from joint distribution factoring

$$\Pr(\mathbf{X}_2|\mathbf{X}_1) = \frac{\Pr(\mathbf{X}_1|\mathbf{X}_2)\Pr(\mathbf{X}_2)}{\Pr(\mathbf{X}_1)}$$

Joint Distributions: *discrete*

- *Joint distributions* factor as *conditional* $\times$ *marginal* dist.'s

$$\Pr(\mathbf{X}_1, \mathbf{X}_2) = \Pr(\mathbf{X}_1 | \mathbf{X}_2) \Pr(\mathbf{X}_2)$$

- *Bayes theorem* is derived from joint distribution factoring

$$\Pr(\mathbf{X}_2 | \mathbf{X}_1) = \frac{\Pr(\mathbf{X}_1 | \mathbf{X}_2) \Pr(\mathbf{X}_2)}{\Pr(\mathbf{X}_1)}$$

- *Independence* of $\mathbf{X}_1$ and $\mathbf{X}_2$ is when

$$\Pr(\mathbf{X}_1 | \mathbf{X}_2) = \Pr(\mathbf{X}_1), \text{ or}$$

$$\Pr(\mathbf{X}_1, \mathbf{X}_2) = \Pr(\mathbf{X}_1) \Pr(\mathbf{X}_2)$$

Joint Distributions: *discrete*

- *Joint distributions* factor as *conditional* $\times$ *marginal* dist.'s

$$\Pr(\mathbf{X}_1, \mathbf{X}_2) = \Pr(\mathbf{X}_1 | \mathbf{X}_2) \Pr(\mathbf{X}_2)$$

- *Bayes theorem* is derived from joint distribution factoring

$$\Pr(\mathbf{X}_2 | \mathbf{X}_1) = \frac{\Pr(\mathbf{X}_1 | \mathbf{X}_2) \Pr(\mathbf{X}_2)}{\Pr(\mathbf{X}_1)}$$

- *Independence* of $\mathbf{X}_1$ and $\mathbf{X}_2$ is when

$$\Pr(\mathbf{X}_1 | \mathbf{X}_2) = \Pr(\mathbf{X}_1), \text{ or}$$

$$\Pr(\mathbf{X}_1, \mathbf{X}_2) = \Pr(\mathbf{X}_1) \Pr(\mathbf{X}_2)$$

- *Marginal distributions* are derived from *joint distributions*

$$\Pr(\mathbf{X}_1) = \sum_{x_2 \in \mathcal{S}_{X_2}} \Pr(\mathbf{X}_1, X_2 = x_2)$$

Joint Distributions: *continuous*

- *Joint distributions* factor as *conditional* $\times$ *marginal* dist.'s

$$f(\mathbf{X}_1, \mathbf{X}_2) = f(\mathbf{X}_1 | \mathbf{X}_2) f(\mathbf{X}_2)$$

- *Bayes theorem* is derived from joint distribution factoring

$$f(\mathbf{X}_2 | \mathbf{X}_1) = \frac{f(\mathbf{X}_1 | \mathbf{X}_2) f(\mathbf{X}_2)}{f(\mathbf{X}_1)}$$

- *Independence* of $\mathbf{X}_1$ and $\mathbf{X}_2$ is when

$$f(\mathbf{X}_1 | \mathbf{X}_2) = f(\mathbf{X}_1), \text{ or}$$

$$f(\mathbf{X}_1, \mathbf{X}_2) = f(\mathbf{X}_1) f(\mathbf{X}_2)$$

- *Marginal distributions* are derived from *joint distributions*

$$f(\mathbf{X}_1) = \int_{\mathcal{S}_{X_2}} f(\mathbf{X}_1, X_2 = x_2) dx_2$$

For Discrete DISTRIBUTIONS

- The *Expected Value* of X $\mathrm{E}_X[X] = \sum_{x \in \mathcal{S}_X} x \Pr(X = x)$

For Discrete DISTRIBUTIONS

- The *Expected Value* of X $\mathrm{E}_X[X] = \sum_{x \in \mathcal{S}_X} x \Pr(X = x)$

- The *Variance* of X $$\mathrm{Var}[X] = \sum_{x \in \mathcal{S}_X} (x - \mathrm{E}_X[X])^2 \Pr(X = x) = \mathrm{E}_X[X^2] - \mathrm{E}_X[X]^2$$

For Discrete DISTRIBUTIONS

- The *Expected Value* of X $\mathrm{E}_X[X] = \sum_{x \in \mathcal{S}_X} x \Pr(X = x)$
- The *Variance* of X $\mathrm{Var}[X] = \sum_{x \in \mathcal{S}_X} (x - \mathrm{E}_X[X])^2 \Pr(X = x)$
  $= \mathrm{E}_X[X^2] - \mathrm{E}_X[X]^2$
- The *Covariance* of X & Y

  $$\mathrm{Cov}(X, Y) = \sum_{(x,y) \in \mathcal{S}_{XY}} (x - \mathrm{E}_X[X])(y - \mathrm{E}_Y[Y]) \Pr(X = x, Y = y)$$

For Discrete DISTRIBUTIONS

- The *Expected Value* of X $\mathrm{E}_X[X] = \sum_{x \in \mathcal{S}_X} x \Pr(X = x)$
- The *Variance* of X
$$\mathrm{Var}[X] = \sum_{x \in \mathcal{S}_X} (x - \mathrm{E}_X[X])^2 \Pr(X = x) = \mathrm{E}_X[X^2] - \mathrm{E}_X[X]^2$$
- The *Covariance* of X & Y
$$\mathrm{Cov}(X, Y) = \sum_{(x,y) \in \mathcal{S}_{XY}} (x - \mathrm{E}_X[X])(y - \mathrm{E}_Y[Y]) \Pr(X = x, Y = y)$$
- The *Correlation* of X & Y $\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}} \in [-1, 1]$

For Discrete DISTRIBUTIONS

- The *Expected Value* of X $\mathrm{E}_X[X] = \sum_{x \in \mathcal{S}_X} x \Pr(X = x)$
- The *Variance* of X $$\mathrm{Var}[X] = \sum_{x \in \mathcal{S}_X} (x - \mathrm{E}_X[X])^2 \Pr(X = x)$$ $$= \mathrm{E}_X[X^2] - \mathrm{E}_X[X]^2$$
- The *Covariance* of X & Y $$\mathrm{Cov}(X, Y) = \sum_{(x,y) \in \mathcal{S}_{XY}} (x - \mathrm{E}_X[X])(y - \mathrm{E}_Y[Y]) \Pr(X = x, Y = y)$$
- The *Correlation* of X & Y $\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}} \in [-1, 1]$
- For $a, b, c \in \mathbb{R}$, $\mathrm{E}[aX + bY + c] = ?$

For Discrete DISTRIBUTIONS

- The *Expected Value* of X $\mathrm{E}_X[X] = \sum_{x \in \mathcal{S}_X} x \Pr(X = x)$
- The *Variance* of X
$$\begin{aligned} \mathrm{Var}[X] &= \sum_{x \in \mathcal{S}_X} (x - \mathrm{E}_X[X])^2 \Pr(X = x) \\ &= \mathrm{E}_X[X^2] - \mathrm{E}_X[X]^2 \end{aligned}$$
- The *Covariance* of X & Y
$$\mathrm{Cov}(X, Y) = \sum_{(x,y) \in \mathcal{S}_{XY}} (x - \mathrm{E}_X[X])(y - \mathrm{E}_Y[Y]) \Pr(X = x, Y = y)$$
- The *Correlation* of X & Y $\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}} \in [-1, 1]$
- For $a, b, c \in \mathbb{R}$, $\mathrm{E}[aX + bY + c] = ?$
- $\mathrm{Var}[aX + bY + c] \stackrel{?}{=} a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y] + 2ab\mathrm{Cov}(X, Y)$

For Discrete DISTRIBUTIONS

- The *Expected Value* of X $\mathrm{E}_X[X] = \sum_{x \in \mathcal{S}_X} x \Pr(X = x)$
- The *Variance* of X $$\begin{aligned}\mathrm{Var}[X] &= \sum_{x \in \mathcal{S}_X} (x - \mathrm{E}_X[X])^2 \Pr(X = x) \\ &= \mathrm{E}_X[X^2] - \mathrm{E}_X[X]^2\end{aligned}$$
- The *Covariance* of X & Y
  $$\mathrm{Cov}(X, Y) = \sum_{(x,y) \in \mathcal{S}_{XY}} (x - \mathrm{E}_X[X])(y - \mathrm{E}_Y[Y]) \Pr(X = x, Y = y)$$
- The *Correlation* of X & Y $\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}} \in [-1, 1]$
- For $a, b, c \in \mathbb{R}$, $\mathrm{E}[aX + bY + c] = ?$
- $\mathrm{Var}[aX + bY + c] \stackrel{?}{=} a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y] + 2ab\mathrm{Cov}(X, Y)$
- If X and Y are *independent*, $\mathrm{E}_{XY}[XY] \stackrel{?}{=} \mathrm{E}_X[X]\mathrm{E}_Y[Y]$

For Continuous DISTRIBUTIONS

- The *Expected Value* of X $\mathsf{E}_X[X] = \int_{x \in \mathcal{S}_X} x\, f(X = x) dx$
- The *Variance* of X $\mathsf{Var}[X] = \int_{x \in \mathcal{S}_X} (x - \mathsf{E}_X[X])^2\, f(X = x) dx$
  $= \mathsf{E}_X[X^2] - \mathsf{E}_X[X]^2$
- *The Covariance* of X and Y
  $$\mathsf{Cov}(X, Y) = \int_{\substack{(x,y) \\ \in \mathcal{S}_{XY}}} (x - \mathsf{E}_X[X])(y - \mathsf{E}_Y[Y]) f(X = x, Y = y) d_{xy}$$
- The *Correlation* of X & Y $\mathsf{Cor}(X, Y) = \frac{\mathsf{Cov}(X,Y)}{\sqrt{\mathsf{Var}[X]\mathsf{Var}[Y]}} \in [-1, 1]$
- For $a, b, c \in \mathbb{R}$, $\mathsf{E}[aX + bY + c] = ?$
- $\mathsf{Var}[aX + bY + c] \stackrel{?}{=} a^2\mathsf{Var}[X] + b^2\mathsf{Var}[Y] + 2ab\mathsf{Cov}(X, Y)$
- If X and Y are *independent*, $\mathsf{E}_{XY}[XY] \stackrel{?}{=} \mathsf{E}_X[X]\mathsf{E}_Y[Y]$

For SAMPLES we have *STATISTICS*

Statistics are functions of the original random variables *and hence are themselves random variables*

For SAMPLES we have *STATISTICS*

Statistics are functions of the original random variables *and hence are themselves random variables*

- The *Sample Mean* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

For SAMPLES we have *STATISTICS*

Statistics are functions of the original random variables *and hence are themselves random variables*

- The *Sample Mean* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- The *Sample Variance* $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

For SAMPLES we have *STATISTICS*

Statistics are functions of the original random variables *and hence are themselves random variables*

- The *Sample Mean* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- The *Sample Variance* $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
- The *Sample Covariance* $S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$

For SAMPLES we have *STATISTICS*

Statistics are functions of the original random variables *and hence are themselves random variables*

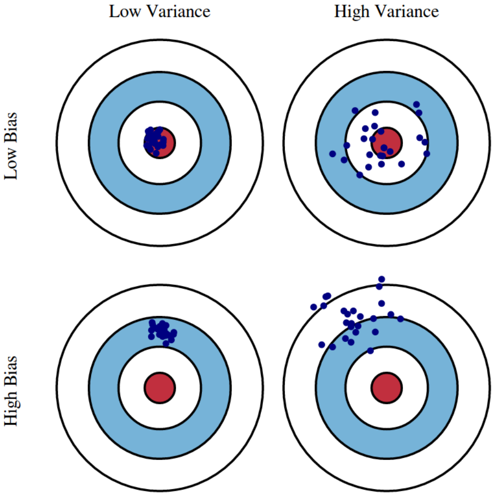
- The *Sample Mean* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- The *Sample Variance* $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
- The *Sample Covariance* $S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$
- The *Sample Correlation** $R_{XY} = \frac{S_{XY}}{\sqrt{S_X^2 S_Y^2}} \in [-1, 1]$

*not robust... correlation of ranks?

Why $n-1$?

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}\left(x_i^2-\frac{1}{n}\sum_{j=1}^{n}x_j\right)^2\right] &= E\left[\sum_{i=1}^{n}\left(x_i^2-\frac{2x_i}{n}\sum_{j=1}^{n}x_j+\left(\frac{1}{n}\sum_{j=1}^{n}x_j\right)^2\right)\right]\\
&= E\left[\sum_{i=1}^{n}x_i^2-\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_j+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_j\right]\\
&= E\left[\sum_{i=1}^{n}x_i^2-\frac{2}{n}\sum_{i=1}^{n}x_i^2-\frac{2}{n}\sum_{j\neq i}x_ix_j+\frac{1}{n}\sum_{i=1}^{n}x_i^2+\frac{1}{n}\sum_{j\neq i}x_ix_j\right]\\
&= E\left[\sum_{i=1}^{n}x_i^2-\frac{2}{n}\sum_{i=1}^{n}x_i^2+\frac{1}{n}\sum_{i=1}^{n}x_i^2-\frac{2}{n}\sum_{j\neq i}x_ix_j+\frac{1}{n}\sum_{j\neq i}x_ix_j\right]\\
&= E\left[\frac{n-1}{n}\sum_{i=1}^{n}x_i^2-\frac{1}{n}\sum_{j\neq i}x_ix_j\right]=\frac{n-1}{n}\sum_{i=1}^{n}E\left[x_i^2\right]-\frac{1}{n}\sum_{j\neq i}E\left[x_ix_j\right]\\
&= \frac{n-1}{n}\sum_{i=1}^{n}(\sigma^2+\mu^2)-\frac{1}{n}\sum_{j\neq i}\mu^2 \quad (why?)\\
&= (n-1)(\sigma^2+\mu^2)-\frac{n^2-n}{n}\mu^2=(n-1)\sigma^2
\end{aligned}
$$

Bias versus Variance of Estimators



Covariance is not Correlation is not Causation

$$\text{Cov}[X, Y] \neq \text{Cor}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}}$$

and "correlation *is not* causation"

or more generally, "association *is not* causation"

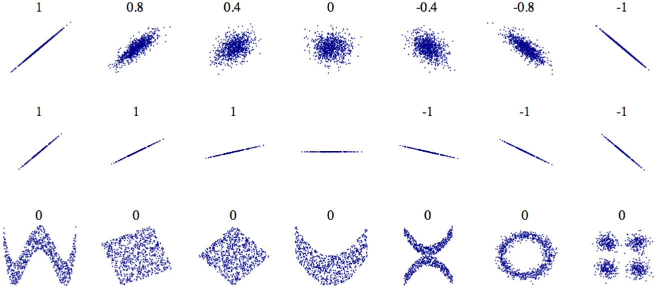
Covariance is not Correlation is not Causation

$$\mathrm{Cov}[X,Y] \neq \mathrm{Cor}[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}}$$

and "correlation *is not* causation"

or more generally, "association *is not* causation"

Conversely, *uncorrelated variables* **are not** necessarily *independent*



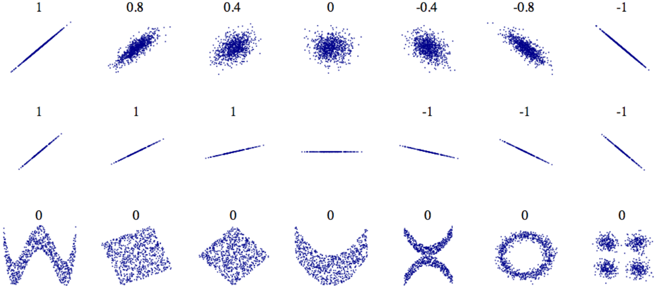
Covariance is not Correlation is not Causation

$$\mathrm{Cov}[X,Y] \neq \mathrm{Cor}[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}}$$

and "correlation *is not* causation"

or more generally, "association *is not* causation"

Conversely, *uncorrelated variables* **are not** necessarily *independent*



*Mutually exclusive events E_1 and E_2 quite dependent (as opposed to independent)