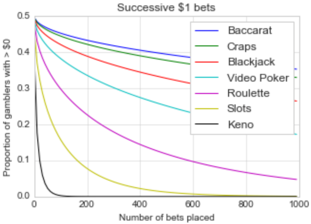
Probability

Schwartz

August 29, 2016

Beating the House

Game	House Advantage
Baccarat (no tie bets)	1.2%
Craps (pass/come)	1.4%
Blackjack (average player)	2.0%
Video Poker (average player)	0.5% - 3%
Roulette (double-zero)	5.3%
Slots	5.0%-10.0%
Keno (average)	27.0%



Blackjack can be legally beaten by keeping track of the probability of getting a high card (10,J,Q,K,A) compared to a low card (2,3,4,5,6). This is called *card counting*. In early 1979, four MIT students taught themselves card counting and along with a professional gambler and an investor who put up most of their capital ($5,000) went to Atlantic City for spring break. They went again in December and then recruited a few more MIT students as "students" for a "blackjack class". The "class" continued to visit Atlantic City intermittently until May 1980 (when the students graduated), during which time they increased their capital four-fold. At about the same time, Bill Kaplan returned to Cambridge after successfully running a blackjack team in Las Vegas. Kaplan earned his BA at Harvard in 1977 and was accepted into Harvard Business School but delayed admission while he ran the blackjack team. Kaplan ran his operation using funds he received upon graduation as Harvard's "outstanding scholar-athlete" and generated more than a 35 fold rate of return in less than nine months of play. Kaplan continued to run his Las Vegas blackjack team as a sideline while attending Harvard Business School but by the time of his graduation the players were so "burnt out" the team disbanded.

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* think you should care

If you're

- Pro $\implies$ Refresher
- Intermediate $\implies$ Solidify
- Noob $\implies$ Exposure

Why *I* think you should care

If you're

- Pro $\implies$ Refresher
- Intermediate $\implies$ Solidify
- Noob $\implies$ Exposure

1. Stats is another feather in your cap 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

Why *I* think you should care

If you're

- Pro $\implies$ Refresher
- Intermediate $\implies$ Solidify
- Noob $\implies$ Exposure

1. Stats is another feather in your cap 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
2. Allows you to completely and fully understand a methodology
   - know exactly what it is and does inside out – no more, no less

Why *I* think you should care

If you're

- Pro $\implies$ Refresher
- Intermediate $\implies$ Solidify
- Noob $\implies$ Exposure

1. Stats is another feather in your cap 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
2. Allows you to completely and fully understand a methodology
   - know exactly what it is and does inside out – no more, no less
3. Can orient prediction/inference machine learning cosmology
   - gives a general theoretical framework to place methodologies

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
- 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
  - <u>Order matters</u>
    E.g., ABC, BAC, CBA, ACB, CAB, BCA
  - There are $\frac{n!}{(n-k)!}$ k-sized permutations of m 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 m things ($k < n$)
- Combination: How many ways can you *combine* things
  - Order *does not matters*

Counting

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

- Permutation: How many ways can you *permute* things
  - <u>Order matters</u>
    E.g., ABC, BAC, CBA, ACB, CAB, BCA
  - There are $\frac{n!}{(n-k)!}$ k-sized permutations of m things ($k < n$)
- Combination: How many ways can you *combine* things
  - <u>Order *does not matters*</u>
    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 m things ($k < n$)
- Combination: How many ways can you *combine* things
  - Order *does not matters*
    E.g., AB, AC, BC
  - The number of k-sized subsets of m things ($k < m$) is

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{(n-k)!k!}$$

Counting: the cartesian product

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

Counting: the cartesian product

- If there are m ways event A can happen
- And there are n ways event B can happen
- Then there are $m \times n$ $\ A \cdot 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 there are $m \times n$ $A \cdot B$ 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 there are $m \times n$ $A \cdot B$ 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?

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

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

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

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

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

(a) 5!

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

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 $XY\_\,\_\,\_$?

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

Problem 2:

- There are k=3 instructors and n=9 students

Problem 2:

- There are k=3 instructors and n=9 students

  The instructors partition student support responsibilities

Problem 2:

- There are k=3 instructors and n=9 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=9 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=9 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?

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

Problem 2:

- There are k=3 instructors and n=9 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?

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

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

Problem 2:

- There are k=3 instructors and n=9 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?

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

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

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

Problem 2:

- There are k=3 instructors and n=9 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?

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

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

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

$$\binom{n-1}{k-1} \quad \binom{8}{2}$$

Problem 3:

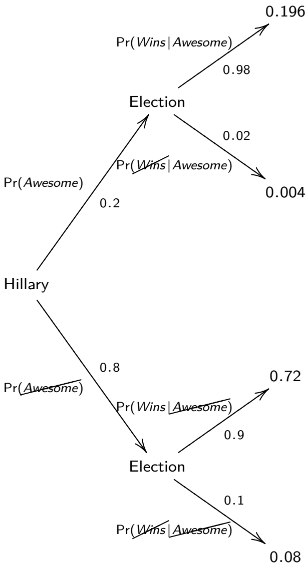
$$\begin{aligned}
\Pr(Hillary_{Awesome}) &= 0.20 \\
Pr(Hillary_{Wins} | Hillary_{Awesome}) &= 0.98 \\
Pr(Hillary_{Wins} | Hillary_{\cancel{Awesome}}) &= 0.90
\end{aligned}$$

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

Problem 3:

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

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

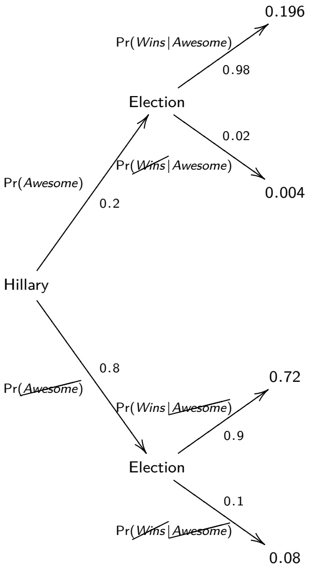


Problem 3:

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

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

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

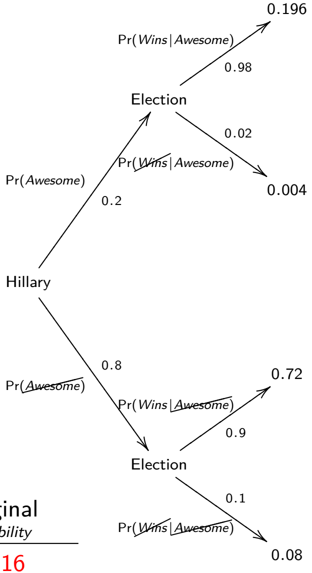


Problem 3:

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

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

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



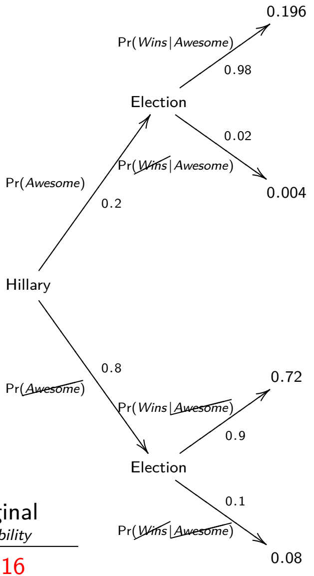
	Awesome Hillary	*~~Awesome~~* Hillary	Marginal *Probability*
Wins Hillary	0.196	0.720	0.916
~~Wins~~ Hillary	0.004	0.080	0.084
Marginal *Probability*	0.200	0.800	1.000

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

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

Random Variables

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

Random Variables

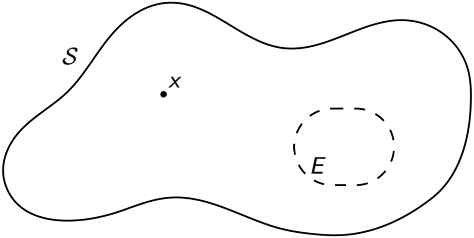
- Random Variable X can take on values in the *sample space* $\mathcal{S}$
- The actualized values x of X are called *outcomes*, and $x \in \mathcal{S}$

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

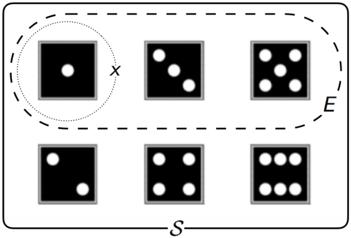
- Random Variable X can take on values in the *sample space* $\mathcal{S}$
- The actualized values x of X are called *outcomes*, and $x \in \mathcal{S}$
- An event E is a subset of the sample space, i.e., $E \subset \mathcal{S}$



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

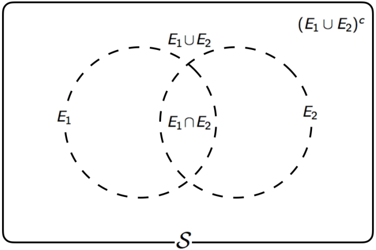
Random Variables

- 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

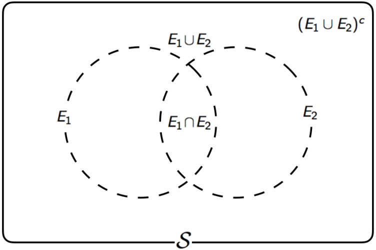
Random Variables: Obvious Rules



Venn Diagram

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

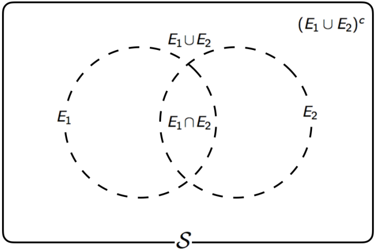
Random Variables: Obvious Rules



Venn Diagram

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

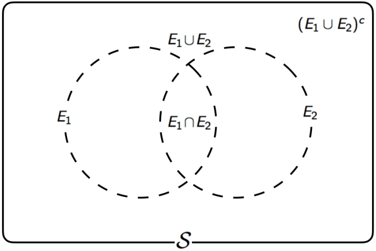
Random Variables: Obvious Rules



Venn Diagram

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

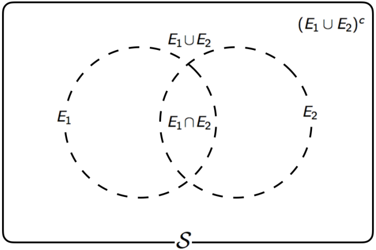
Random Variables: Obvious Rules



Venn Diagram

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

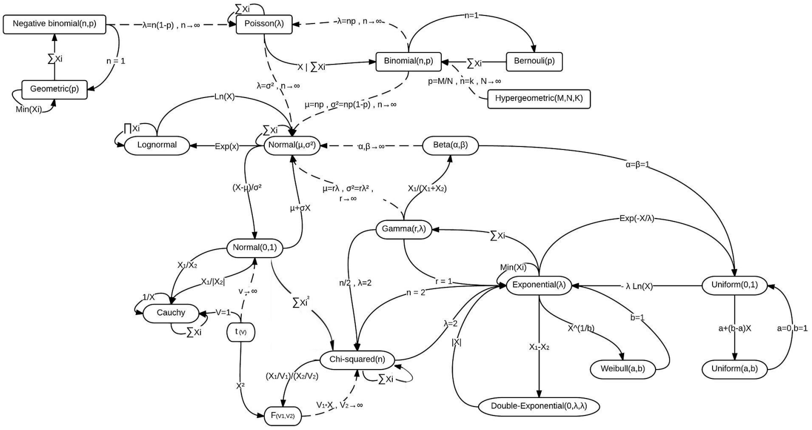
Random Variables: Obvious Rules



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



Discrete Distributions: *Bernoulli*

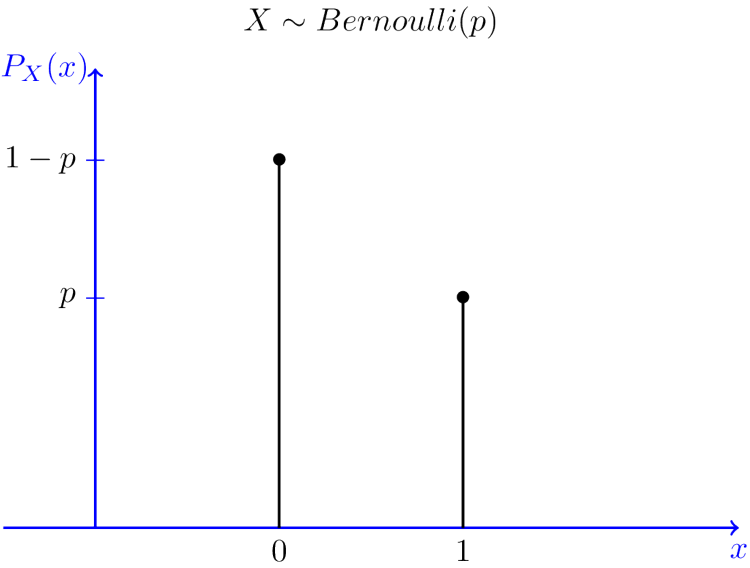
The “coin flip”

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

$$\Pr(Y = y|\theta) = \theta^y (1 - \theta)^{1-y}$$

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

Discrete Distributions: *Bernoulli*



Discrete Distributions: *Geometric*

The "how many times until"

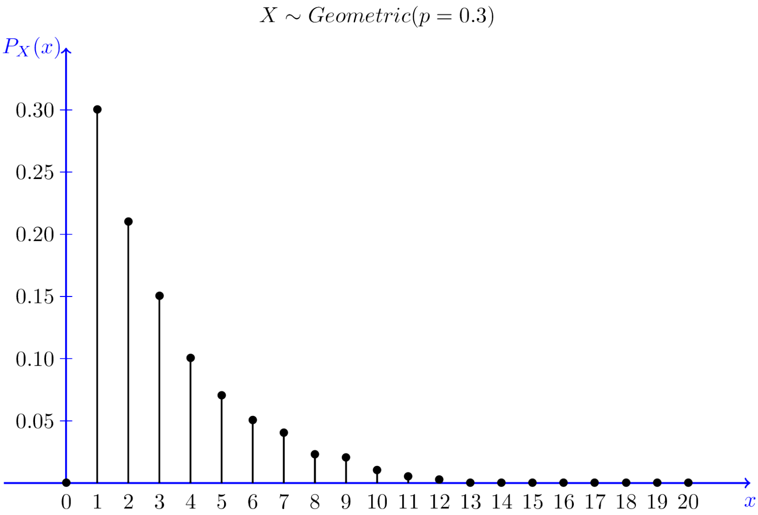
$$k \in \{0, 1, \cdots \infty\}$$

$$\Pr(X = k|\theta) = (1-\theta)^{k-1}\theta$$

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

"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"

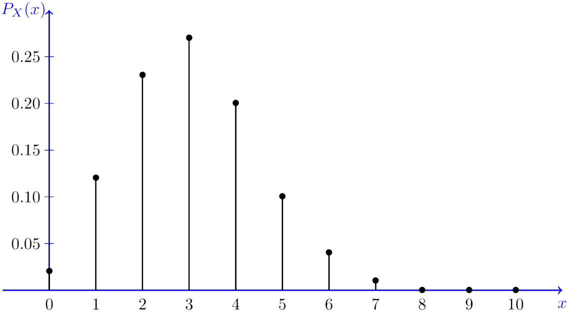
$$k \in \{1, 2, \cdots n\}$$

$$\Pr(X = k|\theta, n) = \begin{pmatrix} n \\ k \end{pmatrix} \theta^k (1-\theta)^{n-k}$$

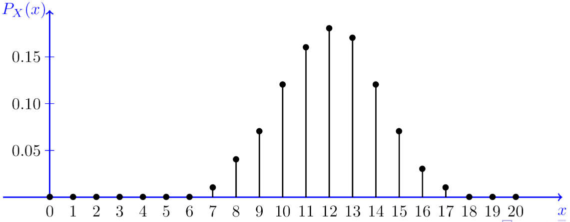
$$\theta \in [0, 1]$$

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"

$$\mathbf{x} = (k_1, x_2, \cdots, x_k)$$
$$x_j \in \{0, 1, \cdots m\} \text{ such that } \sum x_j = m$$

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

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

Discrete Distributions: *Poisson*

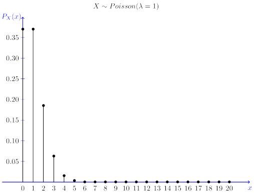
The "number of arrivals"

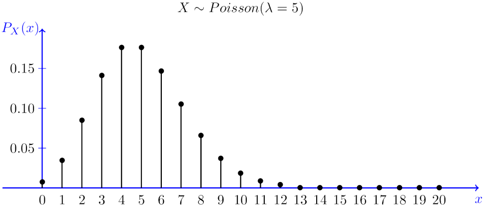
$$k \in \{0, 1, \cdots \infty\}$$

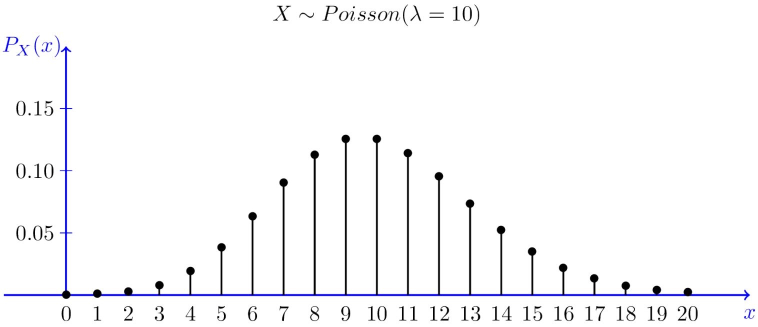
$$\Pr(X = k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

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

Discrete Distributions: *Poisson*

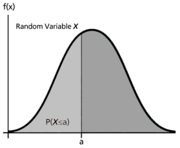




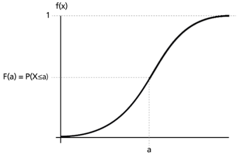


pmf's, *cdf*'s, and characteristic functions

1. We have thus far defined the distribution of a random variable by it's *probability mass function*

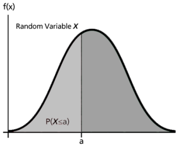


2. We can equivalently alternatively define the distribution of X by it's *cumulative distribution function*

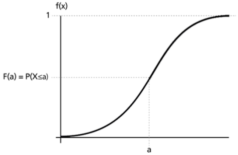


pmf's, *cdf*'s, and characteristic functions

1. We have thus far defined the distribution of a random variable by it's <u>*probability mass function*</u>



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

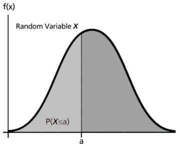


3. Yet another way to define the distribution of X is by its *moment generating function* or its *characteristic function*:

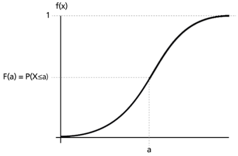
$$\mathrm{E}\left[tX\right], \text{ or } \mathrm{E}\left[itX\right], \text{ respectively}$$

pmf's, *cdf*'s, and characteristic functions

1. We have thus far defined the distribution of a random variable by it's *probability mass function*



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

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

  with respective characteristic functions

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

Discrete Distributions: *Poisson* ≈ *Bionomial*

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

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

Continuos Distributions: *Uniform*

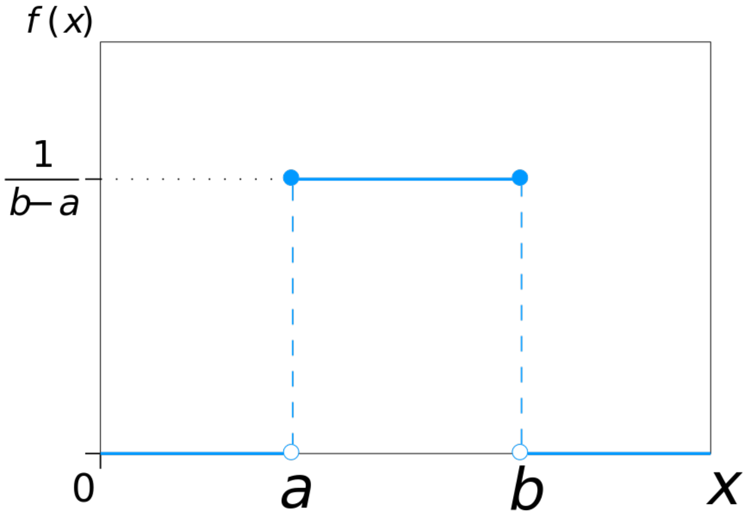
The "random continuous number"

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

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

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

Discrete Distributions: *Uniform*



Continuos Distributions: *Normal*

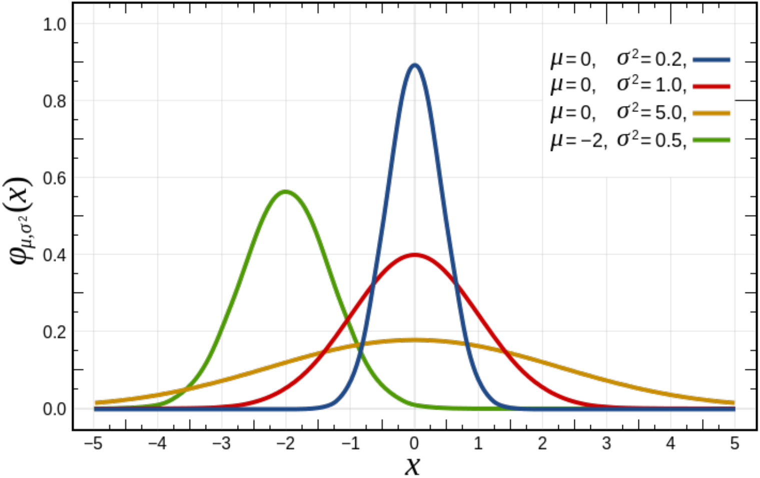
The "bell curve"

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

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

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

Continuos Distributions: *Normal*



Continuos Distributions: *Normal*²

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

Continuos Distributions: *Normal*[2]

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

$$Z_j = \frac{X_i - \mu}{\sigma^2}$$

Continuos Distributions: *Normal*^{2}

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

  $$Z_j = \frac{X_i - \mu}{\sigma^2}$$

  are standard normal random variables

Continuos Distributions: *Normal*[2]

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

$$Z_j = \frac{X_i - \mu}{\sigma^2}$$

are standard normal random variables

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

Continuos Distributions: *Normal*[2]

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

$$Z_j = \frac{X_i - \mu}{\sigma^2}$$

are standard normal random variables

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

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

Continuos Distributions: *Normal*[2]

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

  $$Z_j = \frac{X_i - \mu}{\sigma^2}$$

  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"

Continuos Distributions: *Normal*[2]

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

$$Z_j = \frac{X_i - \mu}{\sigma^2}$$

are standard normal random variables

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

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

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

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

Continuos 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

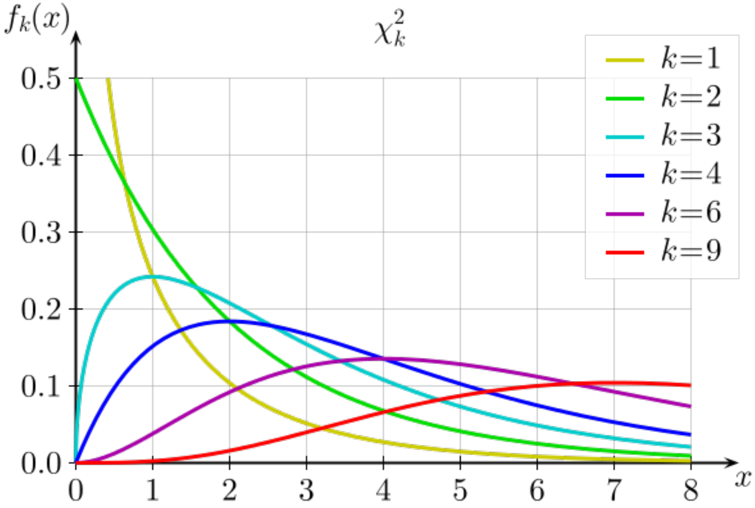
i.e., $Z_j \sim Normal\left(0, 1\right)$, and

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

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

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

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



Continuos Distributions: *Normal* + *Normal*

- The characteristic function of a normal random variable is

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

Continuos 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

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

Continuos Distributions: *Normal* $\sim X_1 + X_2 + \cdots + X_n$

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

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

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

- 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 is

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

Continuos Distributions: *Normal* $\sim X_1 + X_2 + \cdots + X_n$

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

Continuos Distributions: *Normal* $\sim X_1 + X_2 + \cdots + X_n$

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

$$\begin{aligned}
& n\log\left(t\mathrm{E}[X] + \frac{t^2\mathrm{E}[X^2]}{2} + \cdots - \frac{t^2\mathrm{E}[X]^2}{2} - \cdots + \cdots\right) \\
&= n\log\left(t\mathrm{E}[X] + \frac{t^2(\mathrm{E}[X^2] - \mathrm{E}[X]^2)}{2!} + \cdots\right) \\
&= e^{tn\mathrm{E}[X] + \frac{1}{2}t^2 n(\mathrm{E}[X^2] - \mathrm{E}[X]^2) + n(\cdots)} \\
&\approx e^{tn\mathrm{E}[X] + \frac{1}{2}t^2 n(\mathrm{E}[X^2] - \mathrm{E}[X]^2)} \text{ as } n \to \infty
\end{aligned}$$

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

Continuos Distributions: *Gamma*

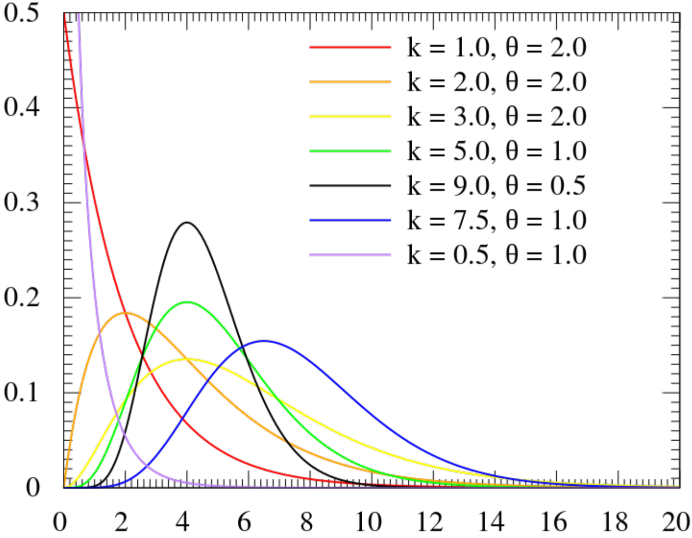
The “Bayesian model for variance”

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

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

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

Continuos Distributions: *Gamma*



Continuos 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

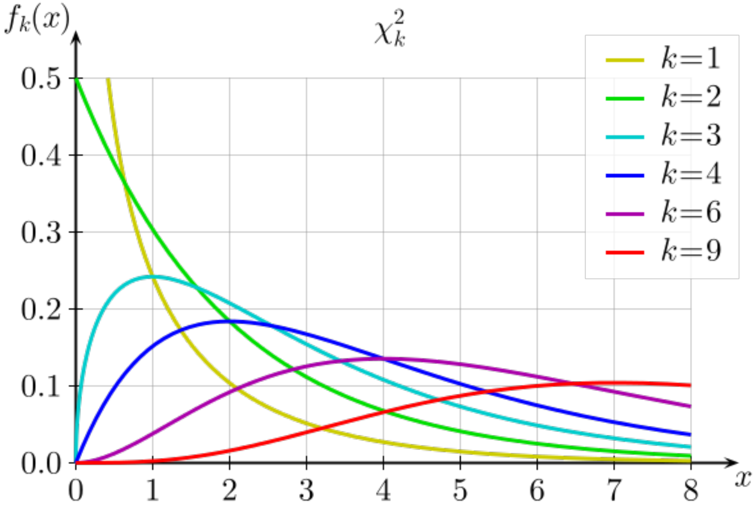
$$x \in \mathbb{R}^+$$

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

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

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

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



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

- The *Exponential* is another special case of the gamma
- The Exponential is often used to model time to failure
- It has an interesting "ageless" property, however, in that

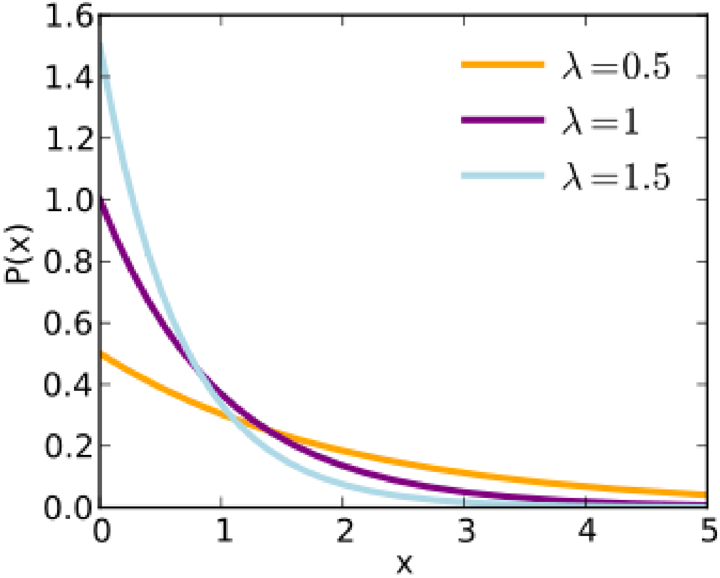
$$\Pr(X = x + c|x = 0) = \Pr(X = x + c|x) \text{ for any value of } x$$

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

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

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

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



Continuos Distributions: *Beta*

The "distribution for modeling random probabilities"

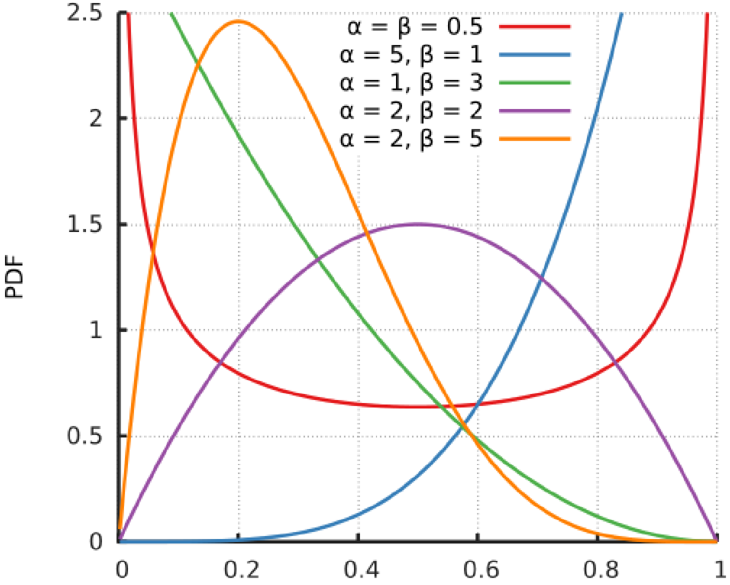
$$p \in [0, 1]$$

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

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

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

Continuos Distributions: *Beta*



Joint Distributions

- A *joint distribution* is set of more than one random variable
- Let X_j be the number of joints Snoop Dogg's sells to client j

Joint Distributions

- A *joint distribution* is set of more than one random variable
- Let X_j be the number of joints Snoop Dogg's sells to client j
- We model Snoop's joint distribution with a *joint distribution*

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

Joint Distributions

- A *joint distribution* is set of more than one random variable
- Let X_j be the number of joints Snoop Dogg's sells to client j
- We model Snoop's joint distribution with a *joint distribution*

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

- We'll use a *multinomial distribution* for the *joint distribution*

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

Joint Distributions

- A *joint distribution* is set of more than one random variable
- Let X_j be the number of joints Snoop Dogg's sells to client j
- We model Snoop's joint distribution with a *joint distribution*

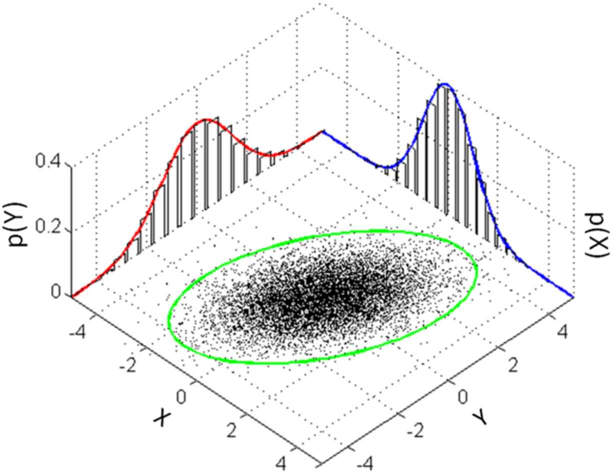
$$\Pr(X_1, X_2, \cdots, X_k)$$

- We'll use a *multinomial distribution* for the *joint distribution*

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

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

Joint Distributions



It turns out this is a multivariate normal distribution – the marginals are themselves normal and strength of relationship between X and Y is determined by a correlation parameter ρ

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 $\mathrm{E}[X] = \sum_{x \in \mathcal{S}_X} x \Pr(X = x)$

For Discrete DISTRIBUTIONS

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

For Discrete DISTRIBUTIONS

- The *Expected Value* of X $\mathsf{E}[X] = \sum_{x \in \mathcal{S}_X} x \Pr(X = x)$
- The *Variance* of X $$\begin{aligned}\mathsf{Var}[X] &= \sum_{x \in \mathcal{S}_X} (x - \mathsf{E}[X])^2 \Pr(X = x) \\ &= \mathsf{E}[X^2] - \mathsf{E}[X]^2\end{aligned}$$
- The *Covariance* of X & Y
  $$\mathsf{Cov}[X, Y] = \sum_{(x,y) \in \mathcal{S}_{XY}} (x - \mathsf{E}[X])(y - \mathsf{E}[Y]) \Pr(X = x, Y = y)$$

For Discrete DISTRIBUTIONS

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

For Discrete DISTRIBUTIONS

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

For Continuous DISTRIBUTIONS

- The *Expected Value* of X $\mathrm{E}[X] = \int_{x \in \mathcal{S}_X} x \Pr(X = x) dx$
- The *Variance* of X $\mathrm{Var}[X] = \int_{x \in \mathcal{S}_X} (x - \mathrm{E}[X])^2 \Pr(X = x) dx$
$$= \mathrm{E}[X^2] - \mathrm{E}[X]^2$$
- *The Covariance* of X and Y
$$\mathrm{Cov}[X, Y] = \int_{\substack{(x,y) \\ \in \mathcal{S}_{XY}}} (x - \mathrm{E}[X])(y - \mathrm{E}[Y]) \Pr(X = x, Y = y) d_{xy}$$
- 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*, $\mathrm{E}[XY] \stackrel{?}{=} \mathrm{E}[X]\mathrm{E}[Y]$

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*

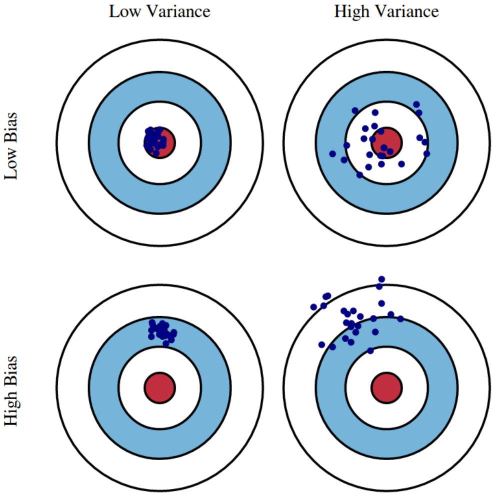
- The *Sample Mean* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- The *Sample Variance* $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
- The *Sample Covariance* $S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$
- The *Sample Correlation** $R_{XY} = \frac{S_{XY}}{\sqrt{S_X^2 S_Y^2}}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

*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



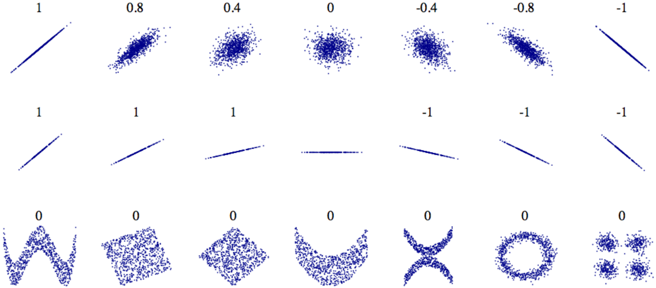
Covariance is not Correlation is not Causation

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

and "correlation *is not* causation"

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

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



*Also, *mutually exclusive events* E_1 and E_2 quite *dependent* as opposed to independent