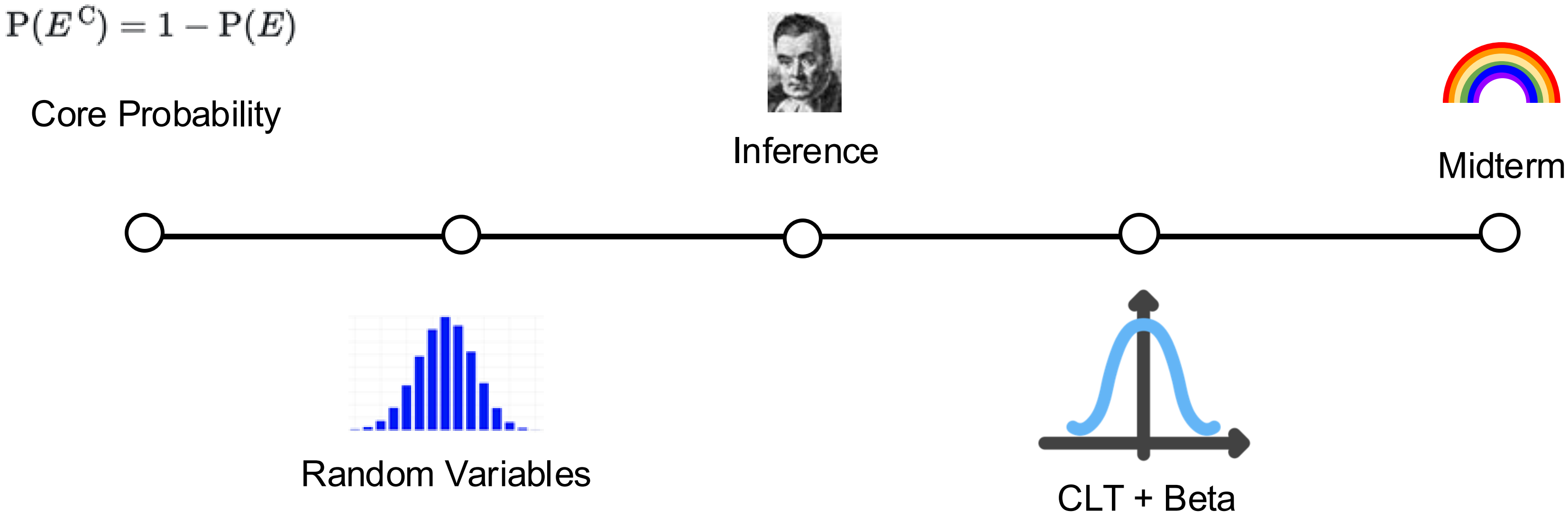
Midterm Review

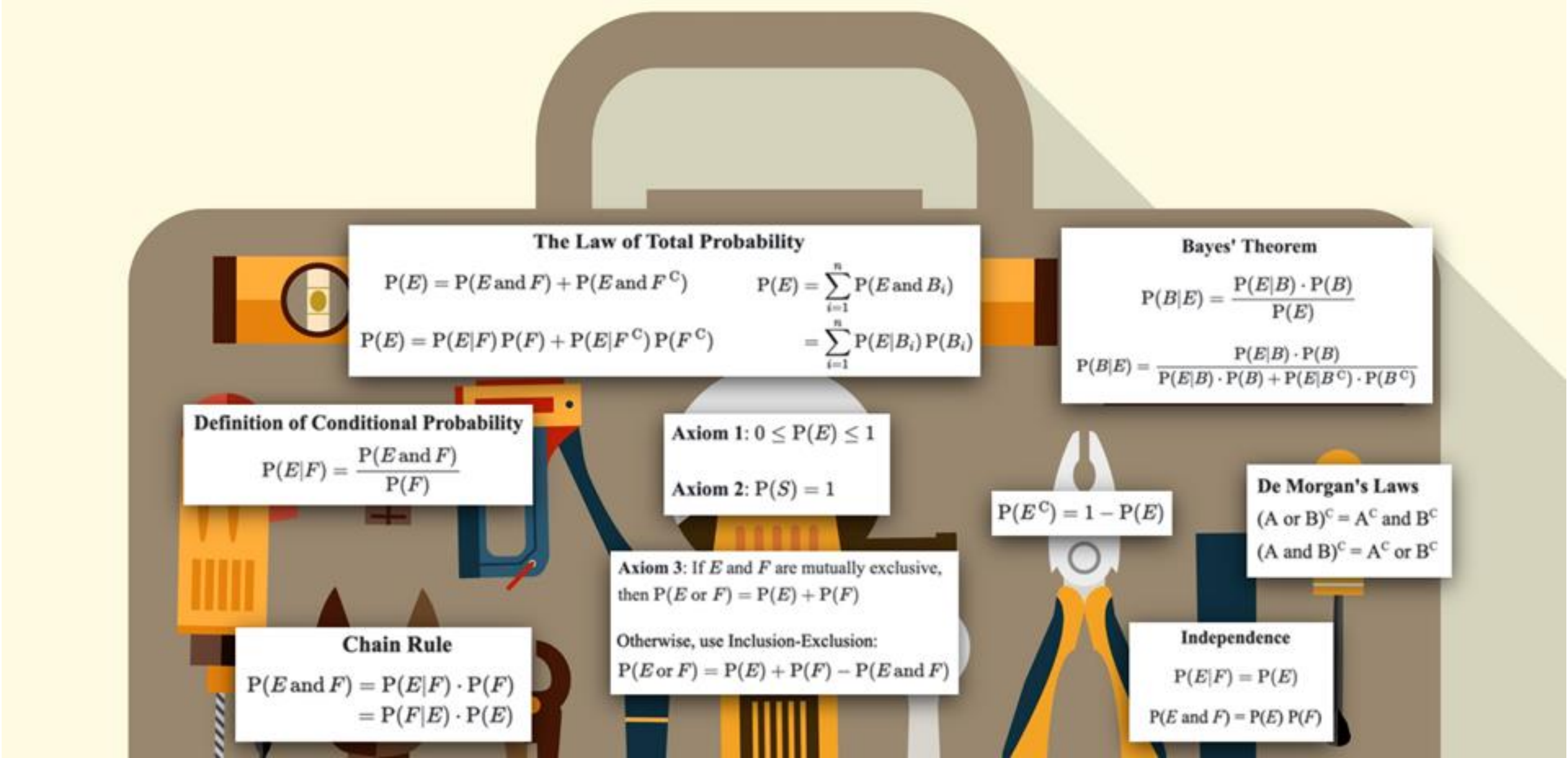
CS109, Fall 2025

Good luck on your exam

Roadmap So Far

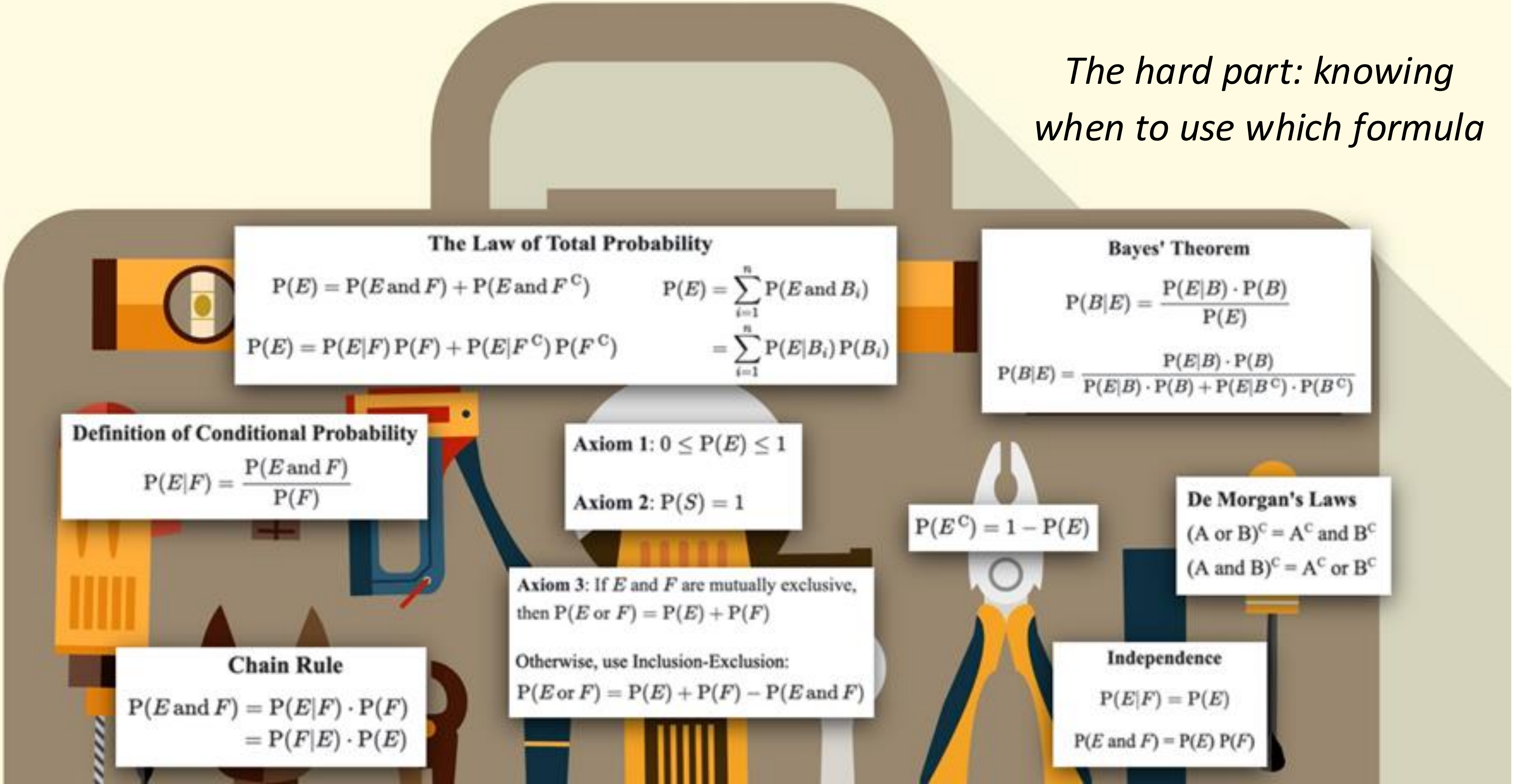


Core Probability: The Formula Toolkit



Core Probability: The Formula Toolkit

The hard part: knowing when to use which formula



Core Probability: The Formulas

Axiom 1: $0 \leq \mathrm{P}(E) \leq 1$

Axiom 2: $\mathrm{P}(S) = 1$

"The probability of anything happening…is 1"

Core Probability: The Formulas

Axiom 1: $0 \le \mathrm{P}(E) \le 1$

Axiom 2: $\mathrm{P}(S) = 1$

Taking the Complement:

$$\mathrm{P}(E^{\mathrm{C}}) = 1 - \mathrm{P}(E)$$

"Everything either happens...or doesn't"

Many times it is easier to calculate $P(E^C)$.

Core Probability: The Formulas

Axiom 1: $0 \leq \mathrm{P}(E) \leq 1$

Axiom 2: $\mathrm{P}(S) = 1$

Taking the Complement:

$$\mathrm{P}(E^C) = 1 - \mathrm{P}(E)$$

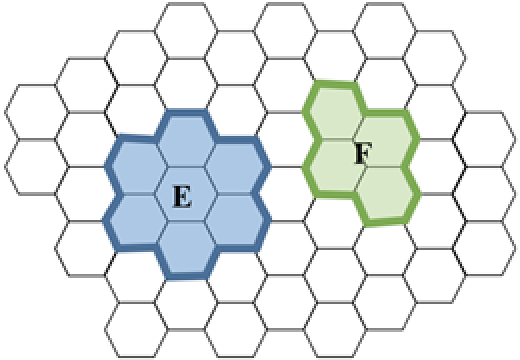
Probability of "Or":

If E and F are mutually exclusive,

$$\mathrm{P}(E \text{ or } F) = \mathrm{P}(E) + \mathrm{P}(F)$$

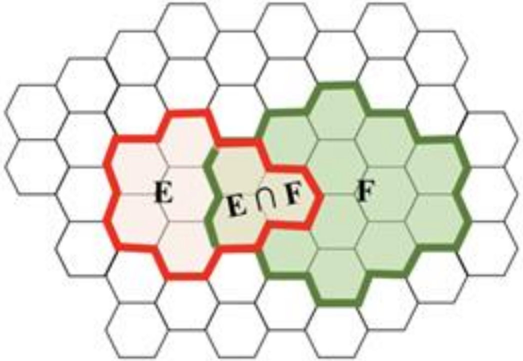
If E and F are *not* mutually exclusive, use Inclusion-Exclusion:

$$\mathrm{P}(E \text{ or } F) = \mathrm{P}(E) + \mathrm{P}(F) - \mathrm{P}(E \text{ and } F)$$



Mutually Exclusive

$\mathrm{P}(E \text{ and } F) = 0$



Not Mutually Exclusive

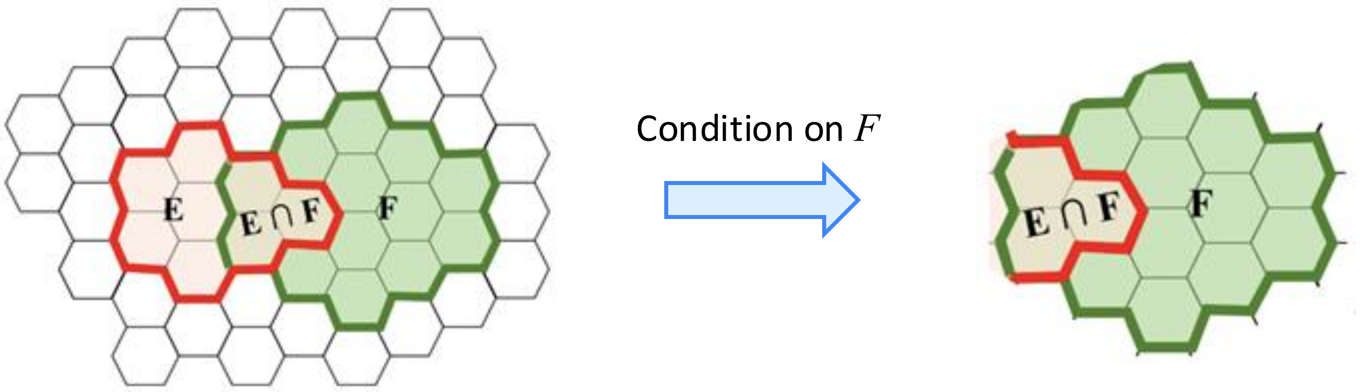
Like "overcounting", then correcting the overcount

Core Probability: Conditionals

$P(E \mid F)$ is $P(E)$, in a world where we know F happened.

$$P(E|F) = \frac{P(E \text{ and } F)}{P(F)}$$

- There is no longer any uncertainty about F: $P(F) = 1$
- When we condition on F, the **sample space shrinks** to outcomes in F



Note! $P(E \mid F) \neq P(E \text{ and } F)$, and $P(E \mid F) \neq P(F \mid E)$

Core Probability: The Formulas

Axiom 1: $0 \leq \mathrm{P}(E) \leq 1$

Axiom 2: $\mathrm{P}(S) = 1$

Taking the Complement:

$$\mathrm{P}(E^{C}) = 1 - \mathrm{P}(E)$$

Probability of "Or":

If E and F are mutually exclusive,

$$\mathrm{P}(E \text{ or } F) = \mathrm{P}(E) + \mathrm{P}(F)$$

If E and F are *not* mutually exclusive, use Inclusion-Exclusion:

$$\mathrm{P}(E \text{ or } F) = \mathrm{P}(E) + \mathrm{P}(F) - \mathrm{P}(E \text{ and } F)$$

Definition of Conditional Probability:

$$\mathrm{P}(E|F) = \frac{\mathrm{P}(E \text{ and } F)}{\mathrm{P}(F)}$$

Chain Rule:

$$\mathrm{P}(E \text{ and } F) = \mathrm{P}(E|F) \cdot \mathrm{P}(F)$$
$$= \mathrm{P}(F|E) \cdot \mathrm{P}(E)$$

Always true, regardless of independence

Core Probability: The Formulas

Axiom 1: $0 \leq P(E) \leq 1$

Axiom 2: $P(S) = 1$

Definition of Conditional Probability:

$$P(E|F) = \frac{P(E \text{ and } F)}{P(F)}$$

Law of Total Probability:

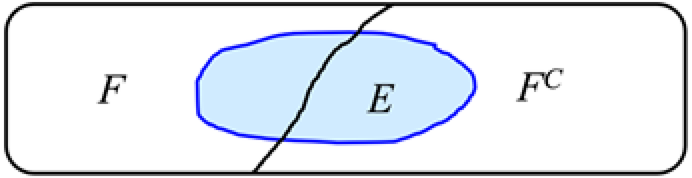
$$P(E) = P(E \text{ and } F) + P(E \text{ and } F^C)$$
$$= P(E|F)\,P(F) + P(E|F^C)\,P(F^C)$$

Taking the Complement:

$$P(E^C) = 1 - P(E)$$

Chain Rule:

$$P(E \text{ and } F) = P(E|F) \cdot P(F)$$
$$= P(F|E) \cdot P(E)$$



When we need $P(E)$, but only know $P(E|F)$

Probability of "Or":

If E and F are mutually exclusive,

$$P(E \text{ or } F) = P(E) + P(F)$$

If E and F are *not* mutually exclusive, use Inclusion-Exclusion:

$$P(E \text{ or } F) = P(E) + P(F) - P(E \text{ and } F)$$

Core Probability: The Formulas

Axiom 1: $0 \le P(E) \le 1$

Axiom 2: $P(S) = 1$

Taking the Complement:

$$P(E^C) = 1 - P(E)$$

Probability of "Or":

If E and F are mutually exclusive,

$$P(E \text{ or } F) = P(E) + P(F)$$

If E and F are *not* mutually exclusive, use Inclusion-Exclusion:

$$P(E \text{ or } F) = P(E) + P(F) - P(E \text{ and } F)$$

Definition of Conditional Probability:

$$P(E|F) = \frac{P(E \text{ and } F)}{P(F)}$$

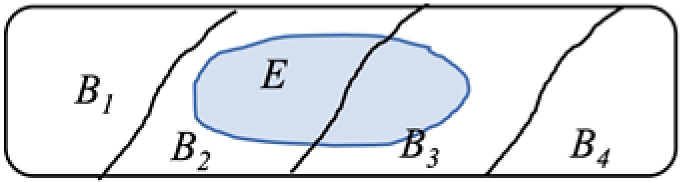
Chain Rule:

$$P(E \text{ and } F) = P(E|F) \cdot P(F) = P(F|E) \cdot P(E)$$

Law of Total Probability:

$$P(E) = P(E \text{ and } F) + P(E \text{ and } F^C) = P(E|F)\,P(F) + P(E|F^C)\,P(F^C)$$

$$P(E) = \sum_{i=1}^{n} P(E \text{ and } B_i) = \sum_{i=1}^{n} P(E|B_i)\,P(B_i)$$



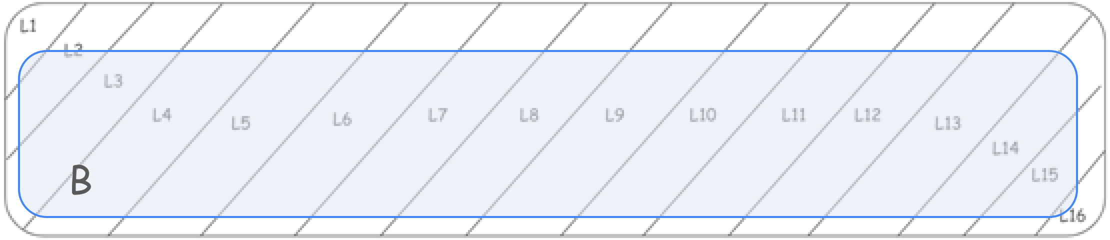
Core Probability: Law of Total Probability

Cellphone

Your cell phone is constantly trying to keep track of where you are. At any given point in time, for all nearby locations, your phone stores a probability that you are in that location.

Right now your phone believes that you are in one of 16 different locations arranged in a grid with the following probabilities (see the figure on the left, `prior`):

$$P(B) = \sum_{i=1}^{16} P(B \cap L = \text{location i})$$
$$= \sum_{i=1}^{16} P(B|L = \text{location i})P(L = \text{location i})$$



Core Probability: The Formulas

Axiom 1: $0 \leq \mathrm{P}(E) \leq 1$

Axiom 2: $\mathrm{P}(S) = 1$

Taking the Complement:

$$\mathrm{P}(E^C) = 1 - \mathrm{P}(E)$$

Probability of "Or":

If E and F are mutually exclusive,

$$\mathrm{P}(E \text{ or } F) = \mathrm{P}(E) + \mathrm{P}(F)$$

If E and F are *not* mutually exclusive, use Inclusion-Exclusion:

$$\mathrm{P}(E \text{ or } F) = \mathrm{P}(E) + \mathrm{P}(F) - \mathrm{P}(E \text{ and } F)$$

Definition of Conditional Probability:

$$\mathrm{P}(E|F) = \frac{\mathrm{P}(E \text{ and } F)}{\mathrm{P}(F)}$$

Chain Rule:

$$\mathrm{P}(E \text{ and } F) = \mathrm{P}(E|F) \cdot \mathrm{P}(F)$$
$$= \mathrm{P}(F|E) \cdot \mathrm{P}(E)$$

Law of Total Probability:

$$\mathrm{P}(E) = \mathrm{P}(E \text{ and } F) + \mathrm{P}(E \text{ and } F^C)$$
$$= \mathrm{P}(E|F)\,\mathrm{P}(F) + \mathrm{P}(E|F^C)\,\mathrm{P}(F^C)$$

$$\mathrm{P}(E) = \sum_{i=1}^{n} \mathrm{P}(E \text{ and } B_i)$$
$$= \sum_{i=1}^{n} \mathrm{P}(E|B_i)\,\mathrm{P}(B_i)$$

Bayes' Theorem:

$$\mathrm{P}(B|E) = \frac{\mathrm{P}(E|B) \cdot \mathrm{P}(B)}{\mathrm{P}(E)}$$

$$\mathrm{P}(B|E) = \frac{\mathrm{P}(E|B) \cdot \mathrm{P}(B)}{\mathrm{P}(E|B) \cdot \mathrm{P}(B) + \mathrm{P}(E|B^C) \cdot \mathrm{P}(B^C)}$$

LOTP in the denominator!

$$\mathrm{P}(B|E) = \frac{\mathrm{P}(E|B) \cdot \mathrm{P}(B)}{\sum_{i=1}^{n} \mathrm{P}(E|B_i)\,\mathrm{P}(B_i)}$$

Core Probability: Independence

Events A and B are independent if knowing A happened doesn't change our belief in B.

$$P(A|B) = P(A)$$

$$P(AB) = P(A)P(B)$$

Examples:

- Completely unrelated things
- Coin flips + dice rolls (generally, repeat trials)
- Sampling with replacement

If you can assume independence, you get this math for free

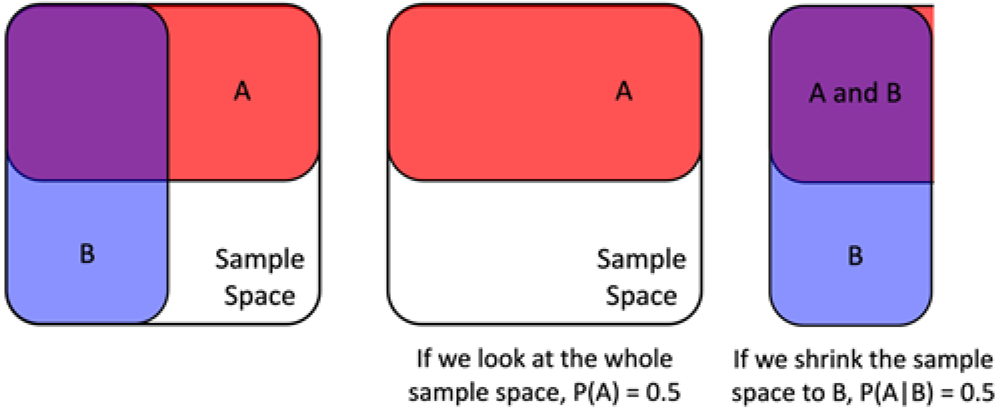
If you need to prove independence, these are the equations you show hold true

Core Probability: Independence

Events A and B are independent if knowing A happened doesn't change our belief in B.

$$P(A|B) = P(A)$$

$$P(AB) = P(A)P(B)$$



What *independence* looks like:

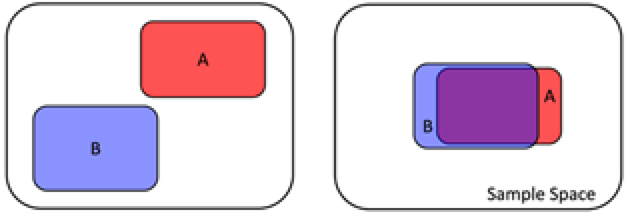
Core Probability: Independence

Events A and B are independent if knowing A happened doesn't change our belief in B.

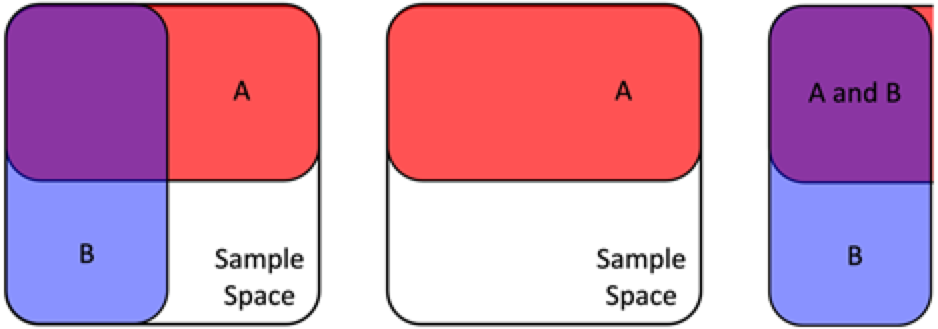
$$P(A|B) = P(A)$$

$$P(AB) = P(A)P(B)$$

What *dependence* looks like:



What *independence* looks like:



If we look at the whole sample space, P(A) = 0.5

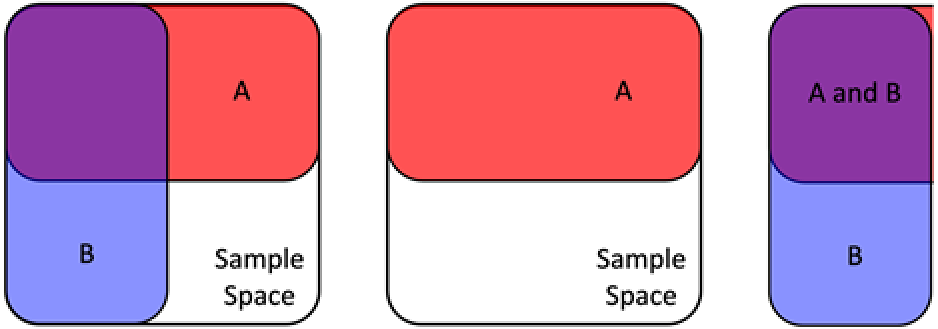
If we shrink the sample space to B, P(A|B) = 0.5

Core Probability: Independence

Events A and B are independent if knowing A happened doesn't change our belief in B.

$$P(A|B) = P(A)$$

$$P(AB) = P(A)P(B)$$



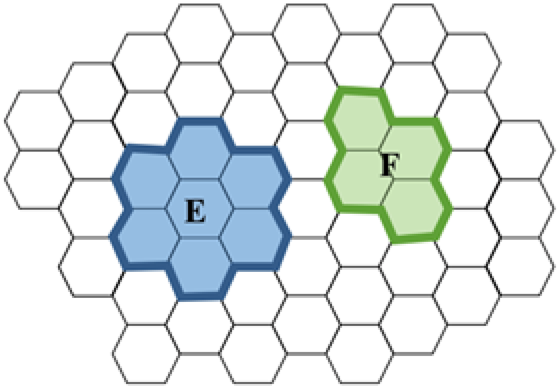
If we look at the whole sample space, P(A) = 0.5

If we shrink the sample space to B, P(A|B) = 0.5

If A and B are independent, A and B^C are independent.

Core Probability: Mutual Exclusion

When two events are mutually exclusive, we know that both events can never co-occur.

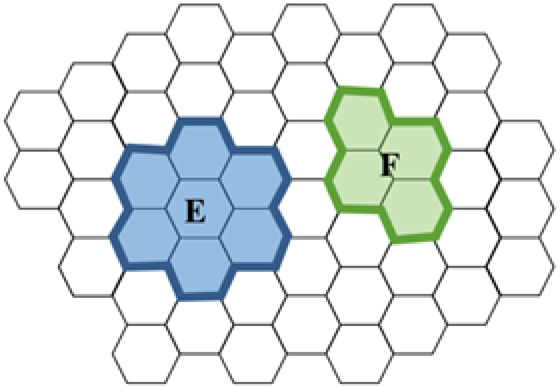


Examples:

- Winning vs. losing the lottery (complements)
- For random variable X, $X = 0$ vs. $X = 1$
- Heads-Heads-Tails vs. Tails-Heads-Tails coin flips

Core Probability: Mutual Exclusion

When two events are mutually exclusive, we know that both events can never co-occur.



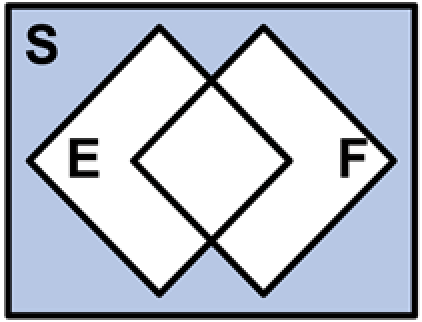
Examples:

- Winning vs. losing the lottery
- For random variable X, $X = 0$ vs. $X = 1$
- Heads-Heads-Tails vs. Tails-Heads-Tails coin flips

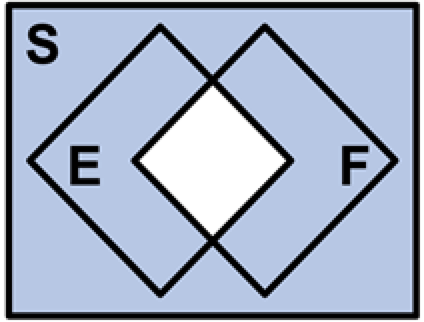
Mutually exclusive events are *never* independent.

- If A and B are mutually exclusive, then knowing A happened tells me B can't happen – that's the opposite of independence!

Core Probability: De Morgan's Laws



$(A \text{ or } B)^C = A^C \text{ and } B^C$



$(A \text{ and } B)^C = A^C \text{ or } B^C$

Neither event happens = A doesn't happen, AND B doesn't happen

"Not both events" = A doesn't happen, OR B doesn't happen

Counting

Why do we do counting? Because a lot of probability involves counting outcomes.

event

$$P(E) = \frac{|E|}{|S|}$$

outcomes in event space

outcomes in sample space

Counting with steps

Definition: Step Rule of Counting (aka Product Rule of Counting)

If an experiment has two parts, where the first part can result in one of m outcomes and the second part can result in one of n outcomes regardless of the outcome of the first part, then the total number of outcomes for the experiment is $m \cdot n$.

Scales nicely to many steps

Counting

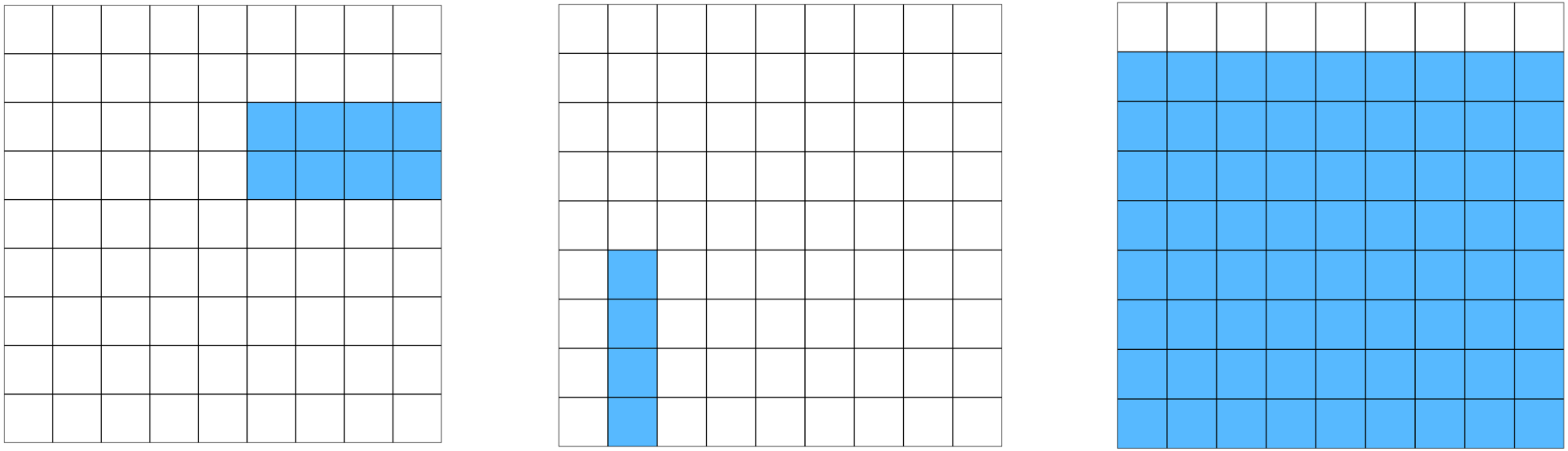
(H, H, H, H, T, T, T, T, T, T)
(H, H, H, T, H, T, T, T, T, T)
(H, H, H, T, T, H, T, T, T, T)
(H, H, H, T, T, T, H, T, T, T)
(H, H, H, T, T, T, T, H, T, T)
(H, H, H, T, T, T, T, T, H, T)
(H, H, H, T, T, T, T, T, T, H)
(H, H, T, H, H, T, T, T, T, T)
(H, H, T, H, T, H, T, T, T, T)
(H, H, T, H, T, T, H, T, T, T)
(H, H, T, H, T, T, T, H, T, T)
(H, H, T, H, T, T, T, T, H, T)
(H, H, T, H, T, T, T, T, T, H)
(H, H, T, T, H, H, T, T, T, T)
(H, H, T, T, H, T, H, T, T, T)
(H, H, T, T, H, T, T, H, T, T)
(H, H, T, T, H, T, T, T, H, T)
(H, H, T, T, H, T, T, T, T, H)
(H, H, T, T, T, H, H, T, T, T)

How many ways can you draw a subset of k items from a set of n?

$$\binom{n}{k}$$

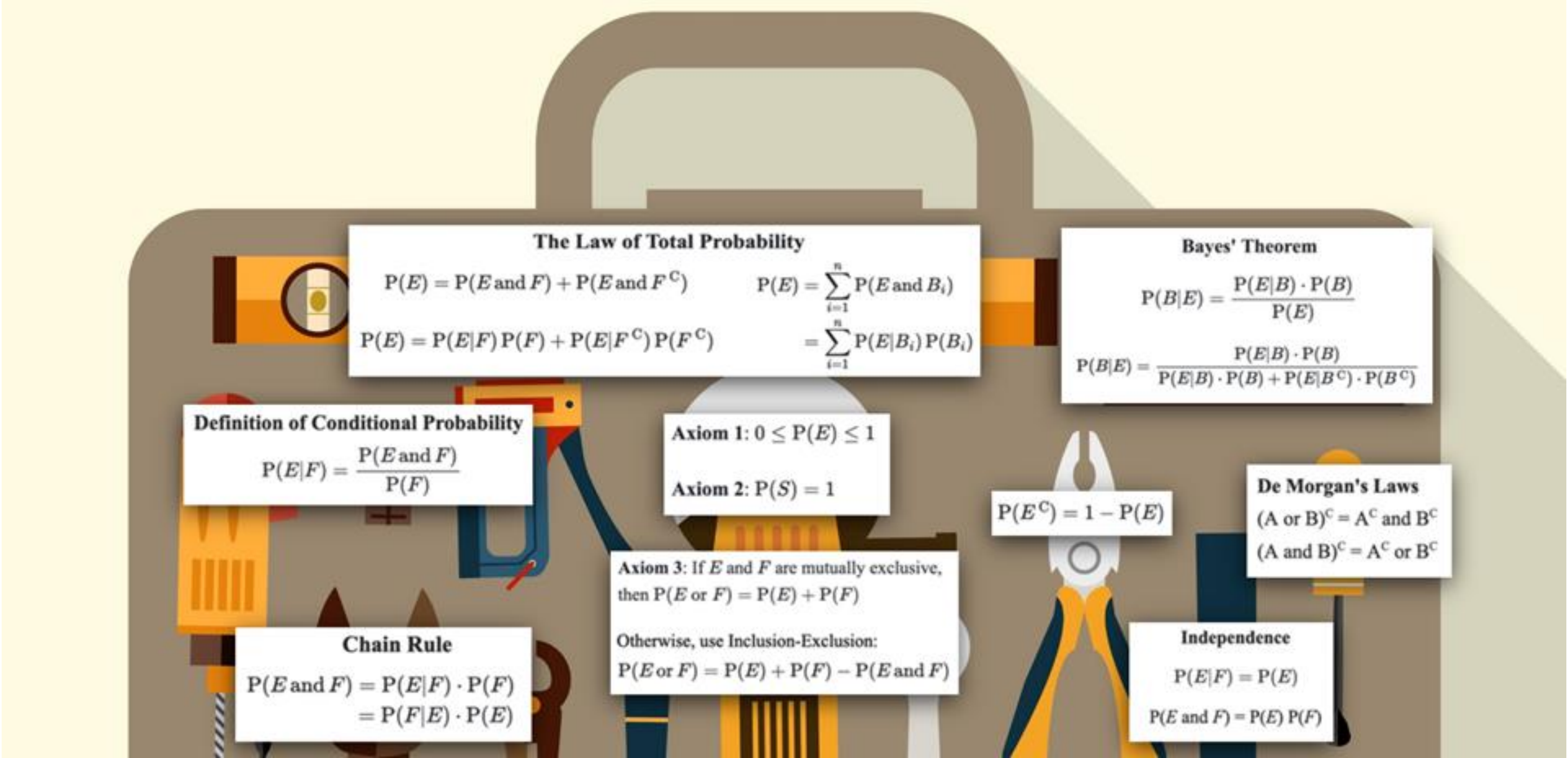
Number of Rectangles in a 9x9 grid

Why is the answer 2025? Hint: $\binom{9}{2} = 45$

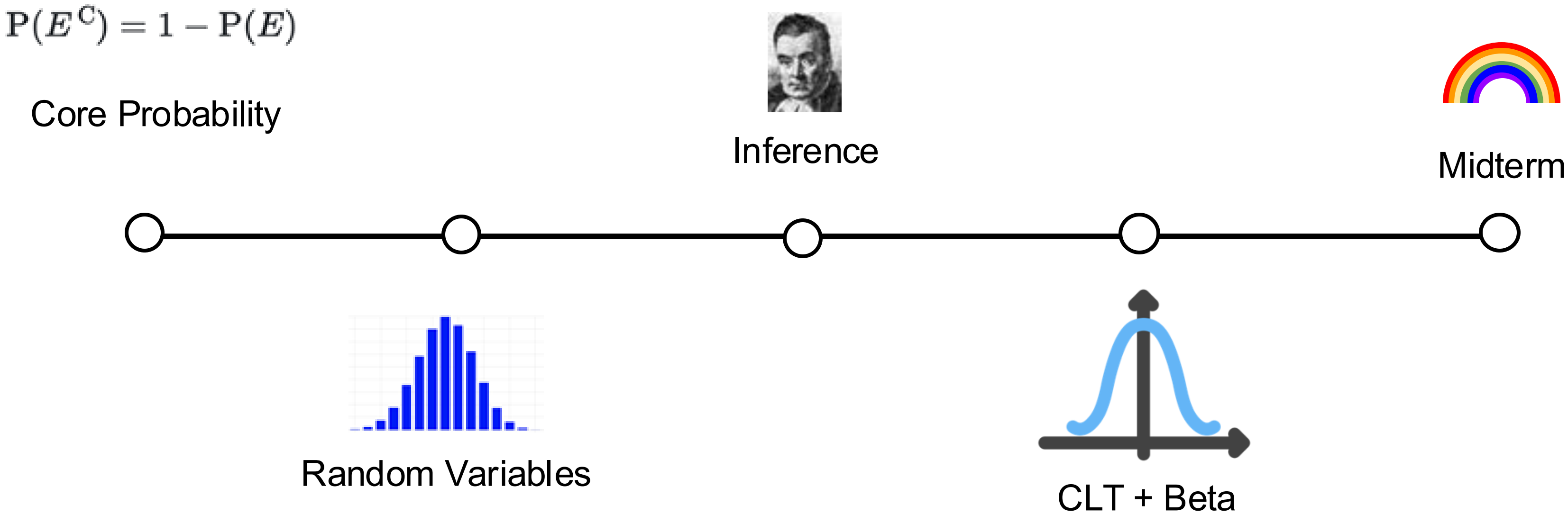


https://probabilityforcs.firebaseapp.com/book/rectangles_in_a_grid

Core Probability: The Formula Toolkit



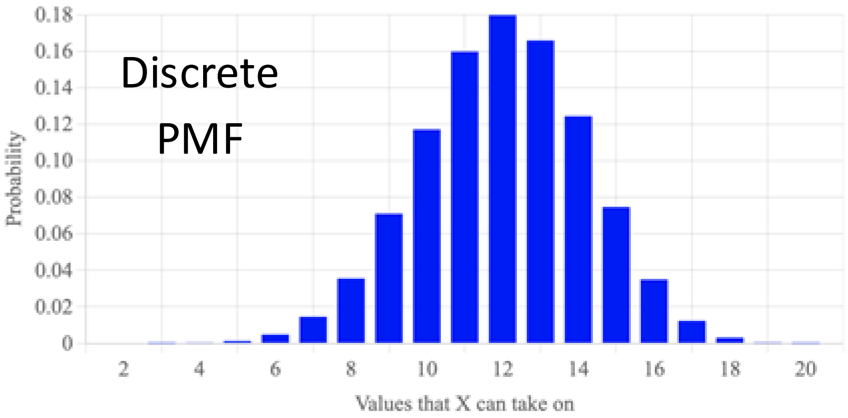
Roadmap So Far

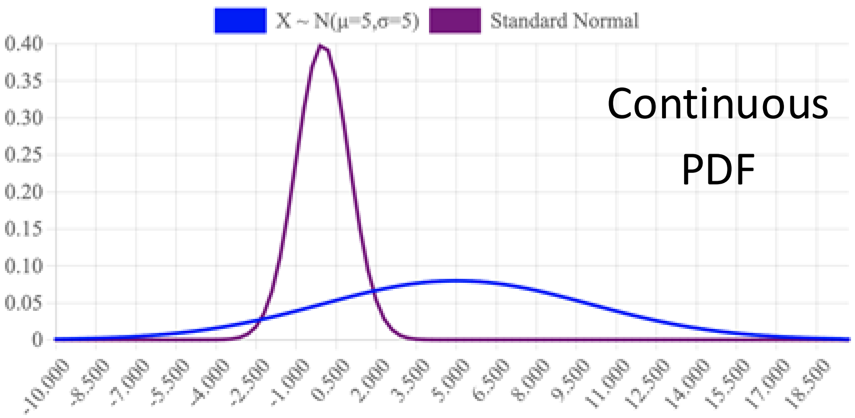


Random Variables

A random variable is a variable whose value is uncertain.

Most Important Property: The probability distribution (PMF or PDF) – all possible values of the RV and their probabilities (or densities).





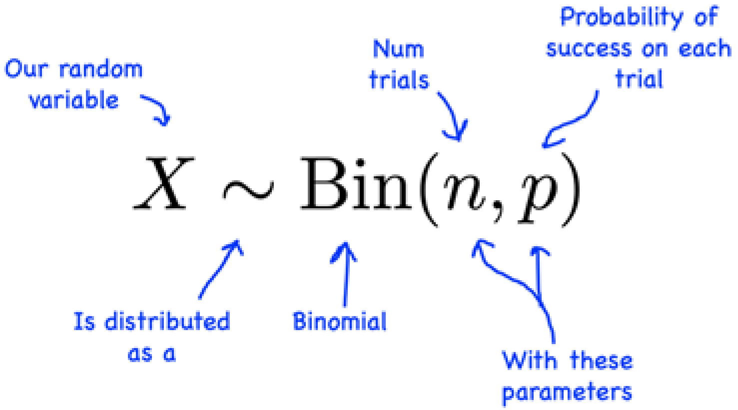
A random variable taking on a value is an event! (X = 10)

Random Variables: Define, Then Use!

The recipe for solving random variable problems:

Step 1 - Define your random variable

"Let X be…"

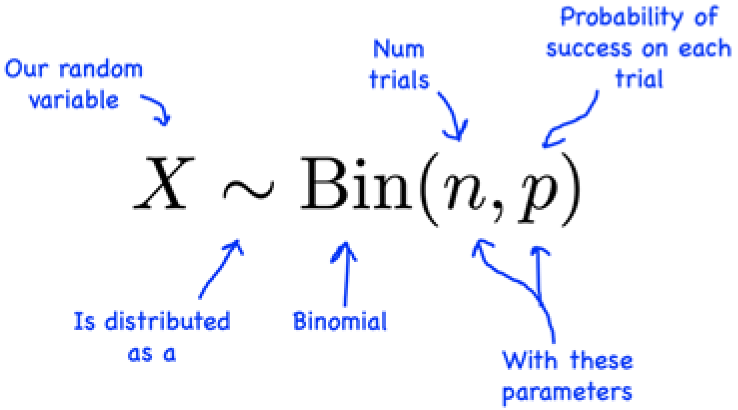


Random Variables: Define, Then Use!

The recipe for solving random variable problems:

Step 1 - Define your random variable

"Let X be..."



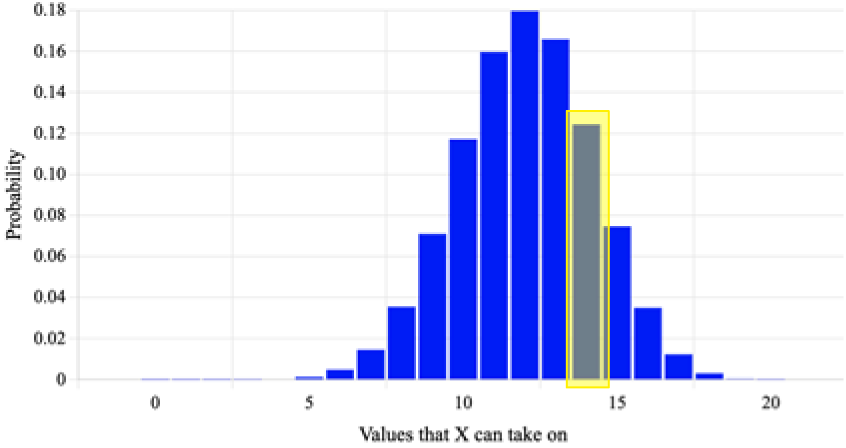
Step 2 - Profit off the PMF/PDF/CDF

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Discrete Random Variables: Using PMFs

To find P(X = x):

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

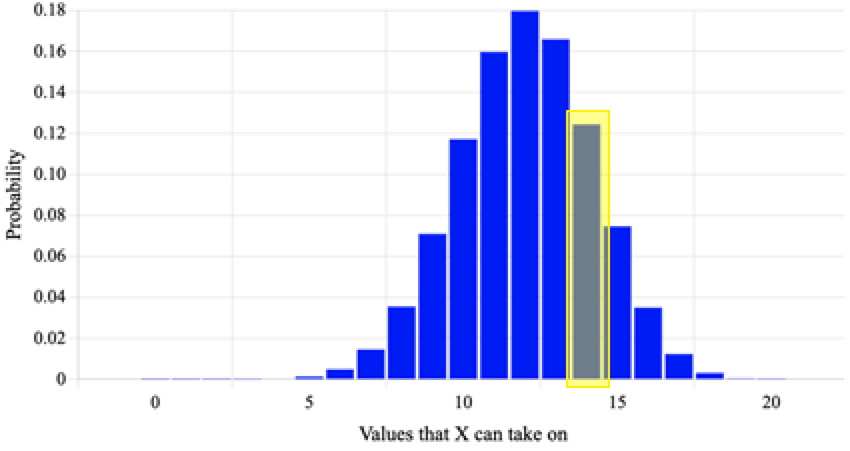


One bar of the PMF

Discrete Random Variables: Using PMFs

To find P(X = x):

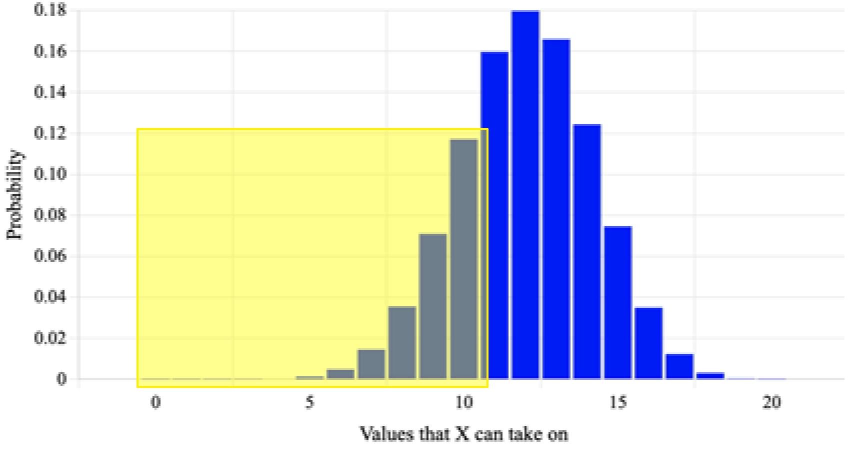
$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$



One bar of the PMF

To find P(X < x), P(X > x), etc:

$$P(X \leq x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i}$$



Add up bars of the PMF

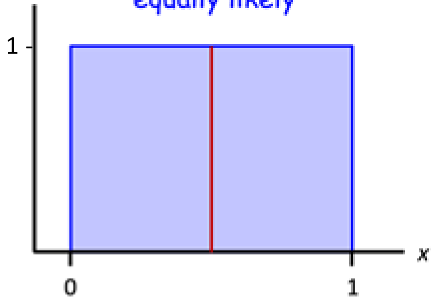
Continuous Random Variables: Using PDFs

To find P(X = x):

It's just zero! :O

For Uniform(0, 1):

All values are
equally likely



Possible values are
between 0 and 1

$$P(0 \leq X \leq 1) = 1$$

$$P(0.5 \leq X \leq 1) = 0.5$$

$$P(0.5 \leq X \leq 0.6) = 0.1$$

$$P(0.5 \leq X \leq 0.5001) = 0.0001$$

$$P(X = 0.5) = 0$$

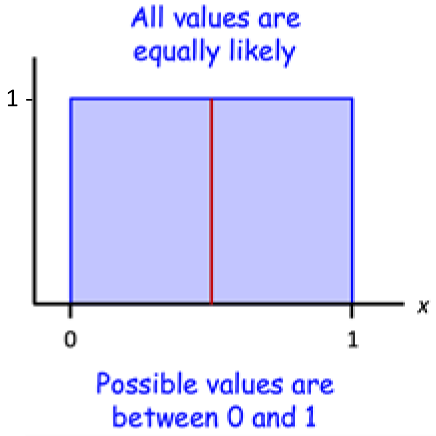
The probability of any exact outcome,
with infinite precision...is zero

Continuous Random Variables: Using PDFs

To find P(X = x):

It's just zero! :O

For Uniform(0, 1):



$$P(0 \le X \le 1) = 1$$

$$P(0.5 \le X \le 1) = 0.5$$

$$P(0.5 \le X \le 0.6) = 0.1$$

$$P(0.5 \le X \le 0.5001) = 0.0001$$

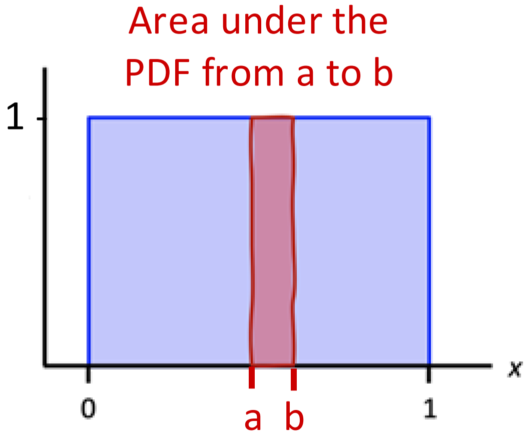
$$P(X = 0.5) = 0$$

The probability of any exact outcome,
with infinite precision...is zero

To find P(X < x), P(X > x), etc:

$$P(a < X < b) = \int_{x=a}^{b} f(X = x)dx$$

(PDF)



Continuous Random Variables: Using CDFs

To find $P(X = x)$:

The CDF can't give you this :)

Continuous Random Variables: Using CDFs

To find P(X = x):

The CDF can't give you this :)

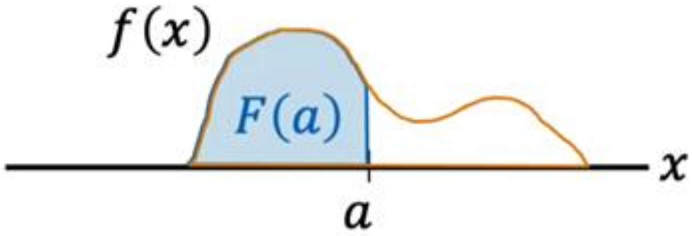
The CDF is the integral of the PDF,
done for you!

To find P(X < x):

Plug into the CDF!

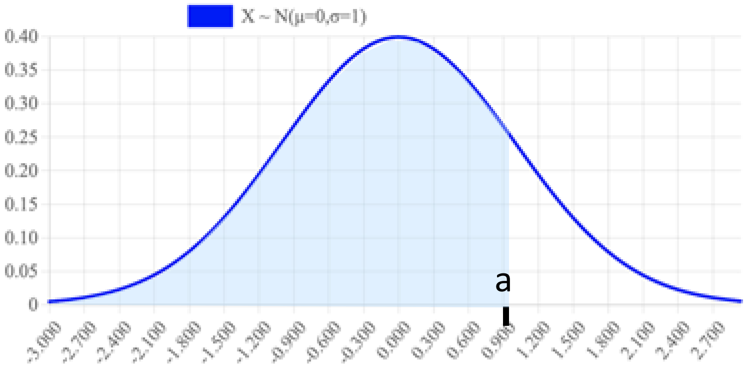
(PDF)

$$F(a) = \int_{-\infty}^{a} f(x)dx$$



Continuous Random Variables: Any Range With CDF

How to get any probability from a CDF:



Probability Query	Solution	Explanation
$\text{P}(X < a)$	$F(a)$	That is the definition of the CDF
$\text{P}(X \leq a)$	$F(a)$	Trick question. $\text{P}(X = a) = 0$
$\text{P}(X > a)$	$1 - F(a)$	$\text{P}(X < a) + \text{P}(X > a) = 1$
$\text{P}(a < X < b)$	$F(b) - F(a)$	$F(a) + \text{P}(a < X < b) = F(b)$

Random Variables: Expected Value and Variance

General Formulas for Expectation:

Discrete

$$\mathrm{E}[X] = \sum_x x \cdot \mathrm{P}(X = x)$$

Continuous

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Random Variables: Expected Value and Variance

General Formulas for Expectation:

Discrete

$$\mathrm{E}[X] = \sum_x x \cdot \mathrm{P}(X = x)$$

Continuous

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

General Formulas for Variance:

$$\mathrm{Var}(X) = E[(X-\mu)^2] = E[X^2] - (E[X])^2$$
$$\mathrm{Var}(aX+b) = a^2 \mathrm{Var}(X)$$

Properties of Expectation:

Property: Linearity of Expectation

$$\mathrm{E}[aX + b] = a\,\mathrm{E}[X] + b$$

Property: Expectation of the Sum of Random Variables

$$E[X+Y] = E[X] + E[Y]$$

Property: Law of Unconcious Statistician

$$E[g(X)] = \sum_x g(x)\,\mathrm{P}(X = x)$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Properties of Variance:

Property: Variance of Sum of ***Independent*** Vars

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Common Distributions: Binomial

Classic RVs have "conditions" – the problem scenario has to fit the RV's metaphor.

Binomial (*n*, *p*)

- "Let X be…" the number of successes in *n* trials
- Fixed number of independent trials, *n*
- Each trial is a success or failure (coin flip)
- Each trial has same probability *p* of success

Common Distributions: Binomial, Geometric

Classic RVs have "conditions" – the problem scenario has to fit the RV's metaphor.

Binomial (*n*, *p*)

- "Let X be…" the number of successes in *n* trials
- Fixed number of independent trials, *n*
- Each trial is a success or failure (coin flip)
- Each trial has same probability *p* of success

Geometric (*p*)

- "Let X be…" the number of trials until a success
- *We don't know the number of trials!*
- Each trial is a success or failure (coin flip)
- Each trial has same probability *p* of success

Common Distributions: Binomial, Geometric

Classic RVs have "conditions" – the problem scenario has to fit the RV's metaphor.

Binomial (*n*, *p*)	**Geometric (*p*)**
- "Let X be…" the number of successes in *n* trials	- "Let X be…" the number of trials until a success
- Fixed number of independent trials, *n*	- *We don't know the number of trials!*
- Each trial is a success or failure (coin flip)	- Each trial is a success or failure (coin flip)
- Each trial has same probability *p* of success	- Each trial has same probability *p* of success

Scenario: flipping a coin

Question: how many heads in n flips?

Question: how many flips *until* heads?

Common Distributions: Binomial, Geometric

Classic RVs have "conditions" – the problem scenario has to fit the RV's metaphor.

Binomial (*n*, *p*)

- "Let X be…" the number of successes in *n* trials
- Fixed number of independent trials, *n*
- Each trial is a success or failure (coin flip)
- Each trial has same probability *p* of success

Geometric (*p*)

- "Let X be…" the number of trials until a success
- *We don't know the number of trials!*
- Each trial is a success or failure (coin flip)
- Each trial has same probability *p* of success

Scenario: Tinder Dating

Question: You swipe right on 100 people. How many matches will you get?

Question: How many times do you have to swipe right until your first match?

Common Distributions: Binomial, Geometric

Classic RVs have "conditions" – the problem scenario has to fit the RV's metaphor.

Binomial (*n*, *p*)
- "Let X be…" the number of successes in *n* trials
- Fixed number of independent trials, *n*
- Each trial is a success or failure (coin flip)
- Each trial has same probability *p* of success

Geometric (*p*)
- "Let X be…" the number of trials until a success
- *We don't know the number of trials!*
- Each trial is a success or failure (coin flip)
- Each trial has same probability *p* of success

Poisson (λ)
- "Let X be…" # events over a fixed time interval
- Events happen independently
- The rate of events occurring (λ) is constant

Exponential (λ)
- "Let X be…" time until the next event
- Events happen independently
- The rate of events occurring (λ) is constant

Common Distributions: Binomial, Geometric

Classic RVs have "conditions" – the problem scenario has to fit the RV's metaphor.

Scenario: Uber requests

Question: how many requests in next 5 min?

Question: how long until the next request?

Poisson (λ)

- "Let X be…" # events over a fixed time interval
- Events happen independently
- The rate of events occurring (λ) is constant

Exponential (λ)

- "Let X be…" time until the next event
- Events happen independently
- The rate of events occurring (λ) is constant

Common Distributions: Binomial, Geometric

Classic RVs have "conditions" – the problem scenario has to fit the RV's metaphor.

Scenario: Phones going off

Question: how many times will a phone ding during a 2 hour exam?

Question: how many seconds until the next phone ding?

Poisson (λ)

- "Let X be…" # events over a fixed time interval
- Events happen independently
- The rate of events occurring (λ) is constant

Exponential (λ)

- "Let X be…" time until the next event
- Events happen independently
- The rate of events occurring (λ) is constant

Distributions: Poisson & Exponential

During an exam, each student's phone will ding at a rate of 0.3 dings per hour. **300 students**

Poisson Approach

What is the probability of exactly 3 dings in an hour?

Exponential Approach

What is the probability of going 5 mins without a ding?

Poisson (λ)	**Exponential (λ)**
- The RV is defined as: *# events over a unit of time*	- The RV is defined as: *time until next event*

Distributions: Poisson & Exponential

During an exam, each student's phone will ding at a rate of 0.3 dings per hour. **300 students**

Poisson Approach

What is the probability of exactly 3 dings in an hour?

Let X be the number of dings in the first hour.

$\lambda = 0.3$ dings/hour $* \ 300 = 90$ dings in **1 hour**

$X \sim \text{Poisson}(\lambda = 90)$

$\text{P}(X = 3) =$ [PMF of Poisson, plug in x=0]

Exponential Approach

What is the probability of going 5 mins without a ding?

Poisson (λ)	**Exponential (λ)**
- The RV is defined as: *# events over a unit of time*	- The RV is defined as: *time until next event*

Distributions: Poisson & Exponential

During an exam, each student's phone will ding at a rate of 0.3 dings per hour. **300 students**

Poisson Approach

What is the probability of exactly 3 dings in an hour?

Let X be the number of dings in the first hour.

$\lambda = 0.3$ dings/hour * $300 = 90$ dings in **1 hour**

$X \sim \text{Poisson}(\lambda = 90)$

$P(X = 3) =$ [PMF of Poisson, plug in x=0]

Exponential Approach

What is the probability of going 5 mins without a ding?

We have to choose a time interval first!

Then, choose lambda

Poisson (λ)	**Exponential (λ)**
- The RV is defined as: *# events over a unit of time*	- The RV is defined as: *time until next event*

Distributions: Poisson & Exponential

During an exam, each student's phone will ding at a rate of 0.3 dings per hour. **300 students**

Poisson Approach

What is the probability of exactly 3 dings in an hour?

Let X be the number of dings in the first hour.

$\lambda = 0.3$ dings/hour * $300 = 90$ dings in **1 hour**

$X \sim \text{Poisson}(\lambda = 90)$

$P(X = 3) =$ [PMF of Poisson, plug in x=0]

Exponential Approach

What is the probability of going 5 mins without a ding?

Time interval: **1 hour**

Let Y be # time intervals until the next ding.

$\lambda = 0.3$ dings/second * $300 = 90$ dings in **1 hour**

$Y \sim \text{Exponential}(\lambda = 90)$

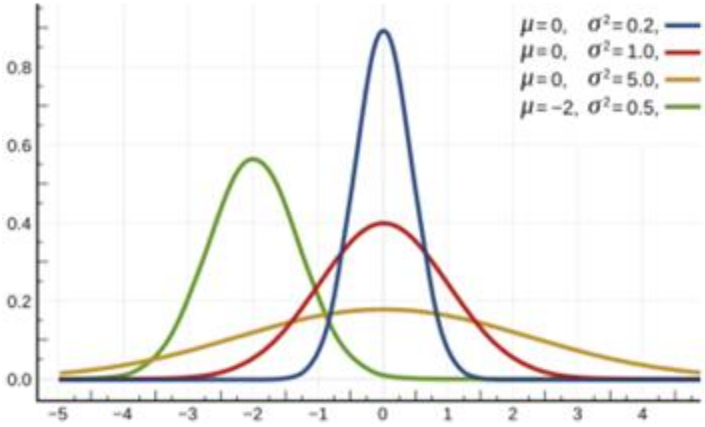
$P(Y > 1/12) = 1$ - [CDF of Exponential, plug in 1]

Poisson (λ)	**Exponential (λ)**
- The RV is defined as: *# events over a unit of time*	- The RV is defined as: *time until next event*

The Normal Distribution

mean variance

$$X \sim \mathcal{N}(\mu, \sigma^2)$$



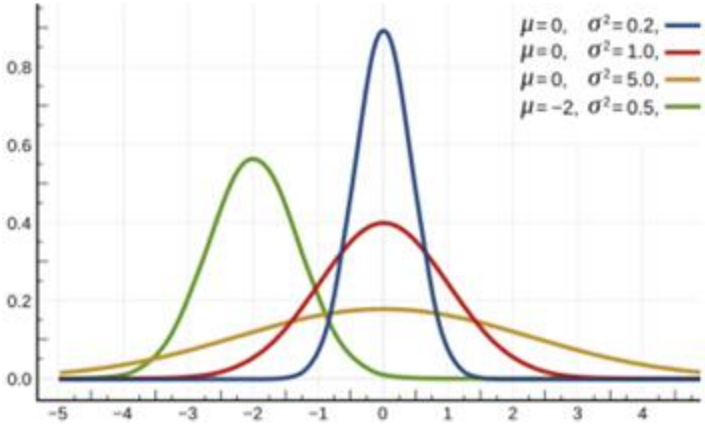
PDF: $$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Yikes. We can't integrate that

The Normal Distribution

mean → μ, variance → σ^2

$$X \sim \mathcal{N}(\mu, \sigma^2)$$



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

Yikes. We can't integrate that

…so we use the CDF instead!

CDF: $$F(x) = \phi\left(\frac{x-\mu}{\sigma}\right)$$

Phi is the Standard Normal (Z) CDF.

We transform our Normal to Z, then use Phi

Approximating A Binomial: Poisson, Normal

If n is large (> 200), need to approximate

Binomial

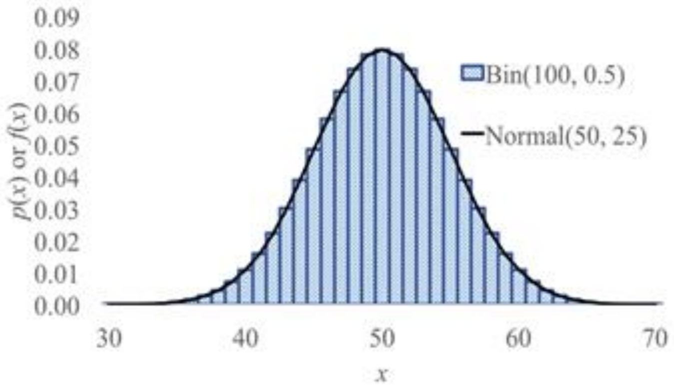
Why approximate? Because computing this is a yikes:

$$\sum_{i=5500}^{10000} \binom{10000}{i} p^i (1-p)^{10000-i}$$

Approximating A Binomial: Poisson, Normal

If n is large (> 200),
need to approximate

Binomial



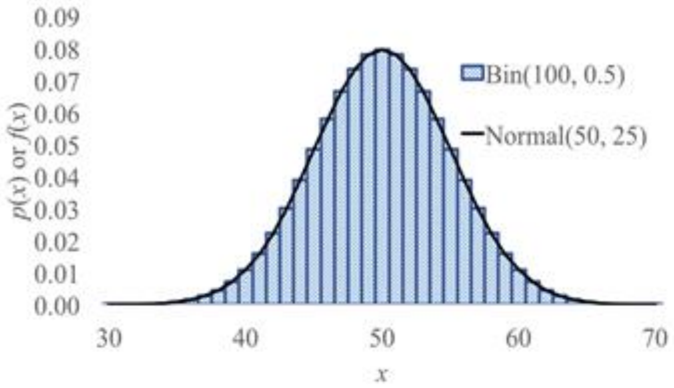
Goal: center a new distribution on the Binomial, so the shapes are the same

Approximating A Binomial: Poisson, Normal

If n is large (> 200), need to approximate

Binomial

- if p is small: → **Poisson**, $\lambda = np$
- if p is moderate: → **Normal**, $\mu = np,\ \sigma^2 = np(1-p)$



Goal: center a new distribution on the Binomial, so the shapes are the same

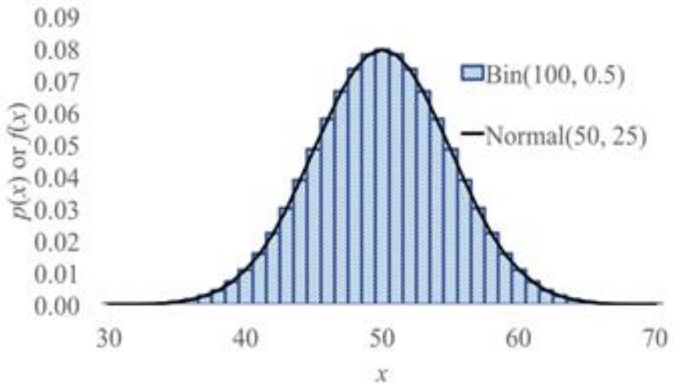
Approximating A Binomial: Poisson, Normal

If n is large (> 200), need to approximate

Binomial

if p is small: → **Poisson**, $\lambda = np$

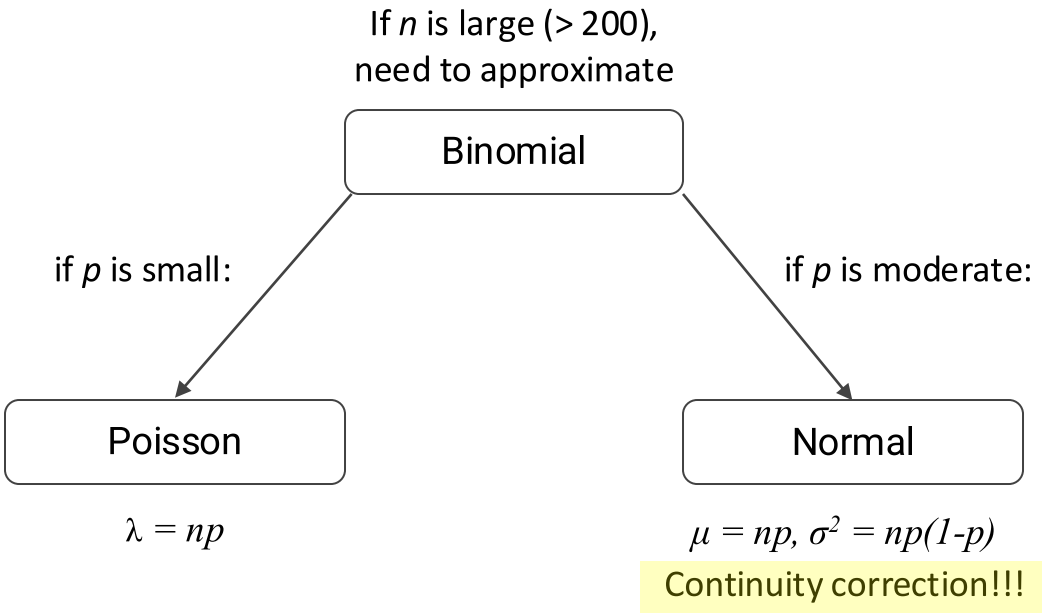
if p is moderate: → **Normal**, $\mu = np$, $\sigma^2 = np(1\text{-}p)$

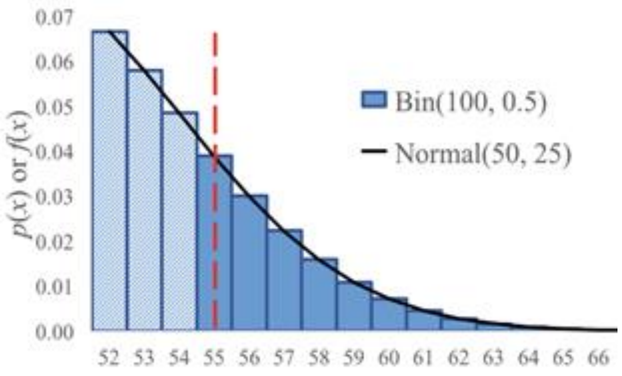


Goal: center a new distribution on the Binomial, so the shapes are the same

Tip: always write out the original binomial first, then switch to the approximation.

Approximating A Binomial: Poisson, Normal





Discrete (e.g., Binomial) probability question	Continuous (Normal) probability question
$P(X = 6)$	$P(5.5 < X < 6.5)$
$P(X \geq 6)$	$P(X > 5.5)$
$P(X > 6)$	$P(X > 6.5)$
$P(X < 6)$	$P(X < 5.5)$
$P(X \leq 6)$	$P(X < 6.5)$

Can you get creative with expectation?

m strings are hashed (not necessarily uniformly) into a hash table with n buckets. Each string's hash is an independent trial, and the probability that a string hashes to bucket i is p_i, with $\sum_{i=1}^{n} p_i = 1$.

What is the expected number of buckets that are not empty?

Can you calculate variance?

7. Estimating Course Size [19 points]

In order to hire the correct number of TAs, Stanford needs to estimate final enrollment in CS109 two weeks before the start of the quarter. Two weeks before the start of the quarter 300 students are enrolled.

For the last 10 offerings of CS109 Stanford has recorded the ratio:

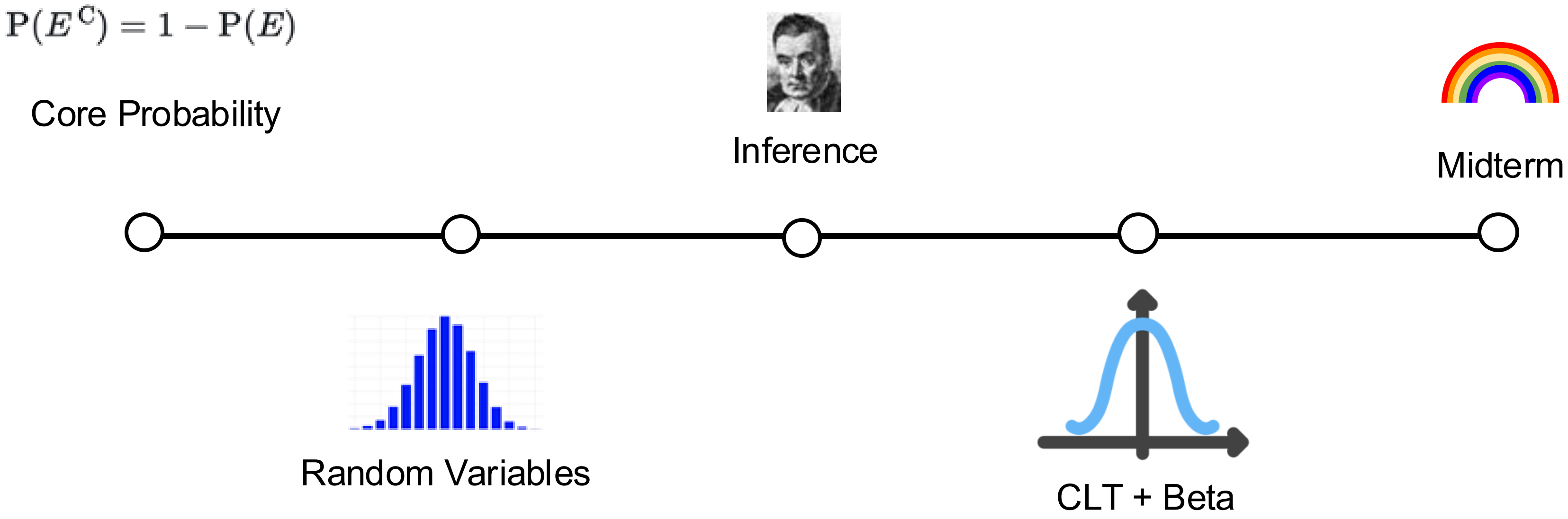
$$r_i = \frac{\text{(final enrollment for offering } i\text{)}}{\text{(enrollment two weeks before start for offering } i\text{)}}$$

which you can access as a list of values $[r_1, r_2, \ldots, r_{10}]$. Assume that each r_i is an i.i.d. sample from the true ratio random variable R. The number of students who end up taking the class this quarter, T, will be $T = 300 \cdot R$. From historical analysis we know that both R and T are normally distributed.

a. (5 points) Estimate $E[T]$, the expected class size. Provide your answer as a math expression.

b. (6 points) Estimate $\text{Var}(T)$. Provide your answer as a math expression.

Roadmap So Far



Inference: Fancy Bayes Theorem!

Posterior belief

Likelihood of evidence

Prior belief

$$P(B|E) = \frac{P(E|B) \quad P(B)}{P(E)}$$

Normalization constant

Inference: Fancy Bayes' Theorem

It's just Bayes' Theorem!*

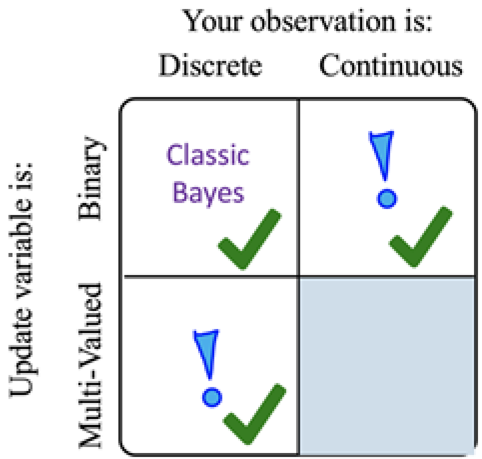
* except, instead of only binary events, we have random variables

Inference: Fancy Bayes' Theorem

It's just Bayes' Theorem!*

* except, instead of only binary events, we have random variables

* except, we have whole distributions to apply Bayes over in a "for loop"



Inference: Fancy Bayes' Theorem

It's just Bayes' Theorem!*

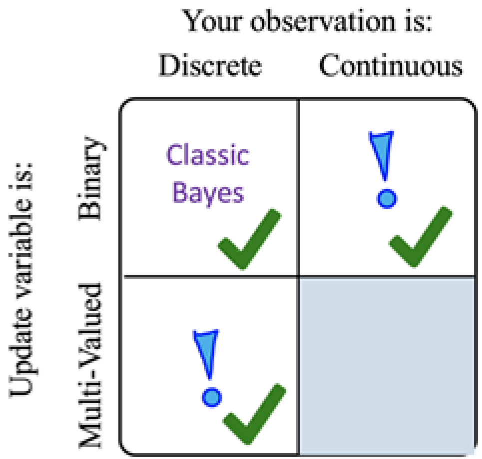
* except, instead of only binary events, we have random variables

* except, we have whole distributions to apply Bayes over in a "for loop"

* except, sometimes we use continuous PDFs ($\mathrm{P} \rightarrow f$)

$$f(X = x | N = n) = \frac{\mathrm{P}(N = n | X = x) f(X = x)}{\mathrm{P}(N = n)}$$

$$\mathrm{P}(N = n | X = x) = \frac{f(X = x | N = n)\, \mathrm{P}(N = n)}{f(X = x)}$$



Inference: Fancy Bayes' Theorem

It's just Bayes' Theorem!*

```
def update_belief_carbon_dating(m = 900):
  # pr_A[i] is P(Age = i| m = 900).
  pr_A = {}
  for i in range(100,10000+1):
    prior = 1 / n_years # P(A = i)
    likelihood = calc_likelihood(m, i) #P(M=m|A=i)
    pr_A[i] = likelihood * prior
  # implicitly computes the normalization constant
  normalize(pr_A)
  return pr_A
```

Inference: Fancy Bayes' Theorem

It's just Bayes' Theorem!*

```
def update_belief_carbon_dating(m = 900):
  # pr_A[i] is P(Age = i| m = 900).
  pr_A = {}
  for i in range(100,10000+1):
    prior = 1 / n_years # P(A = i)
    likelihood = calc_likelihood(m, i) #P(M=m|A=i)
    pr_A[i] = likelihood * prior
  # implicitly computes the normalization constant
  normalize(pr_A)
  return pr_A
```

We return a whole PMF as a dict → pr_A = {}

For loop over all values of the "updating" RV → for i in range(100,10000+1):

Numerator of Bayes → pr_A[i] = likelihood * prior

Denominator of Bayes → normalize(pr_A)

Inference: Fancy Bayes' Theorem

It's just Bayes' Theorem!*

```
def update_belief_carbon_dating(m = 900):
  # pr_A[i] is P(Age = i| m = 900).
  pr_A = {}
  for i in range(100,10000+1):
    prior = 1 / n_years # P(A = i)
    likelihood = calc_likelihood(m, i) #P(M=m|A=i)
    pr_A[i] = likelihood * prior
  # implicitly computes the normalization constant
  normalize(pr_A)
  return pr_A
```

We return a whole PMF as a dict → pr_A = {}

For loop over all values of the "updating" RV → for i in range(100,10000+1):

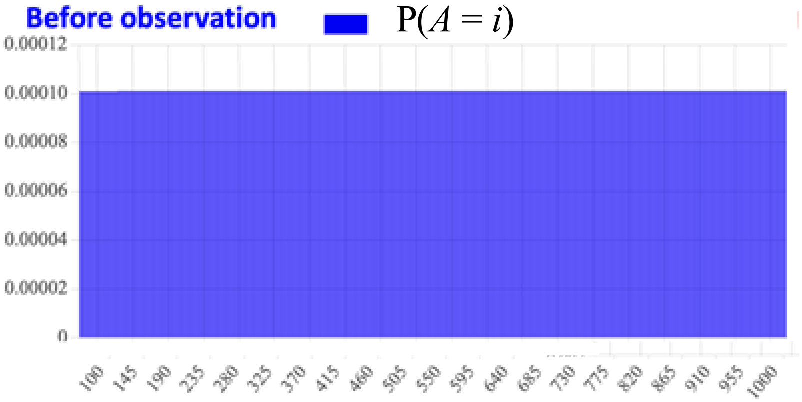
The likelihood is usually the hard part

Numerator of Bayes → pr_A[i] = likelihood * prior

Denominator of Bayes → normalize(pr_A)

Inference: Updating Belief For Random Variables

Prior belief distribution



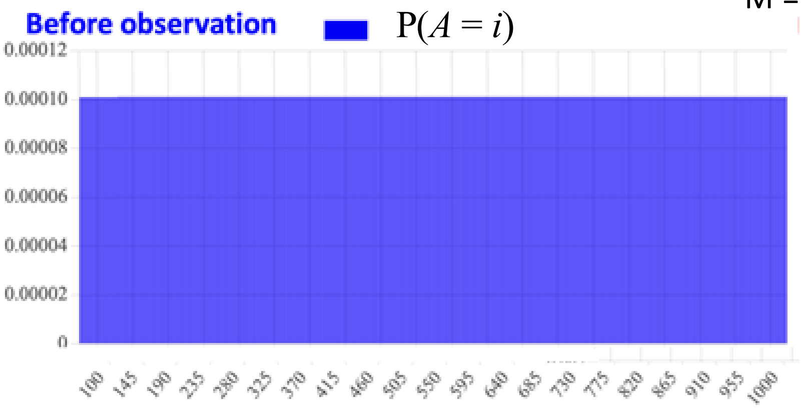
1. We start with a distribution of the random variable *before* observing anything.

Inference: Updating Belief For Random Variables

Prior belief distribution

We observe that M = 900

M = # of C14 molecules remaining



1. We start with a distribution of the random variable *before* observing anything.

2. We observe some new information.

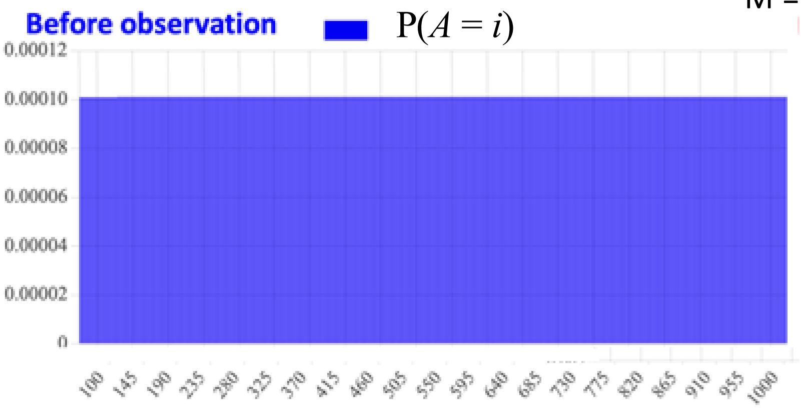
Inference: Updating Belief For Random Variables

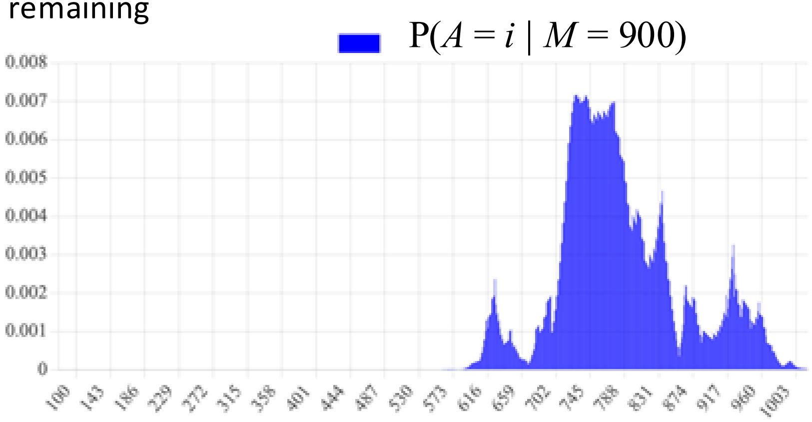
Prior belief distribution

Posterior belief distribution

We observe that M = 900

M = # of C14 molecules remaining





1. We start with a distribution of the random variable *before* observing anything.

2. We observe some new information.

3. Using Bayes Theorem, we update our probability distribution.

Inference: You Already Did It

Prior belief distribution

$P(L = \text{location i})$

0.05	0.10	0.05	0.05
0.05	0.10	0.05	0.05
0.05	0.05	0.10	0.05
0.05	0.05	0.10	0.05

$P(B = 1 | L = \text{location i})$

0.75	0.95	0.75	0.05
0.05	0.75	0.95	0.75
0.01	0.05	0.75	0.95
0.01	0.01	0.05	0.75

Posterior belief distribution

$P(L = \text{location i} | B = 1)$

0.07	0.18	0.07	0.01
0.01	0.14	0.09	0.07
0.01	0.01	0.14	0.09
0.01	0.01	0.01	0.07

1. We start with a distribution of the random variable *before* observing anything.

2. We observe some new information.

3. Using Bayes Theorem, we update our probability distribution.

Example: Cell Phone Location

Goal: Find P(Location | Observed signal)

- Let B be a Bernoulli RV for if the cell phone observes a signal ($B = 1$ or $B = 0$).
- Let L represent the location of the cell phone (there are 16 possibilities).

$$P(L = \text{location i} | B = 1) = \frac{P(B = 1 | L = \text{location i}) P(L = \text{location i})}{P(B = 1)}$$

$P(L = \text{location i})$

0.05	0.10	0.05	0.05
0.05	0.10	0.05	0.05
0.05	0.05	0.10	0.05
0.05	0.05	0.10	0.05

$P(B = 1 | L = \text{location i})$

0.75	0.95	0.75	0.05
0.05	0.75	0.95	0.75
0.01	0.05	0.75	0.95
0.01	0.01	0.05	0.75

$P(L = \text{location i} | B = 1)$

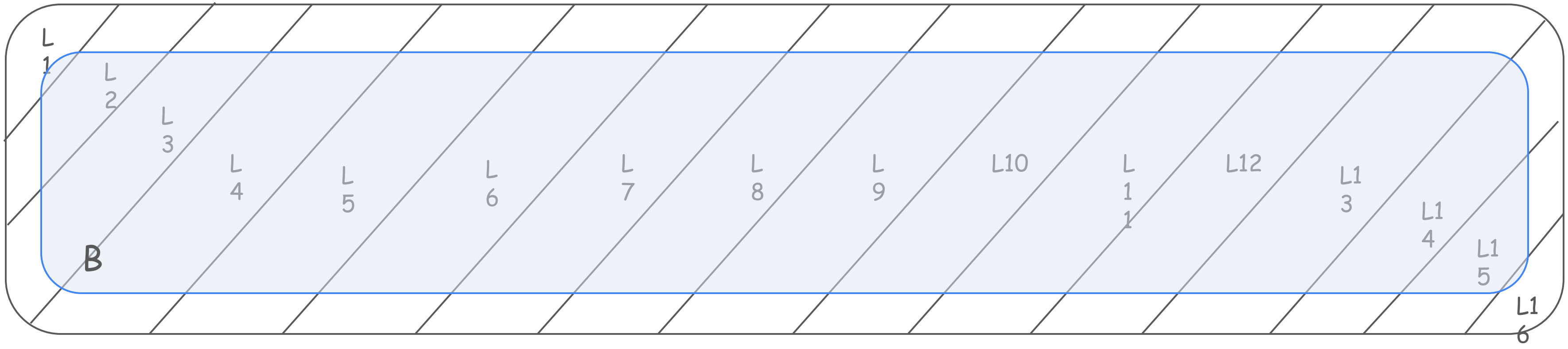
0.07	0.18	0.07	0.01
0.01	0.14	0.09	0.07
0.01	0.01	0.14	0.09
0.01	0.01	0.01	0.07

Example: Cell Phone Location

Goal: Find P(Location | Observed signal)

- Let B be a Bernoulli RV for if the cell phone observes a signal ($B = 1$ or $B = 0$).
- Let L represent the location of the cell phone (there are 16 possibilities).

$$P(L = \text{location i} | B = 1) = \frac{P(B = 1 | L = \text{location i}) P(L = \text{location i})}{P(B = 1)}$$



Example: Cell Phone Location

Goal: Find P(Location | Observed signal)

- Let B be a Bernoulli RV for if the cell phone observes a signal ($B = 1$ or $B = 0$).
- Let L represent the location of the cell phone (there are 16 possibilities).

$$P(L = \text{location i} | B = 1) = \frac{P(B = 1 | L = \text{location i}) P(L = \text{location i})}{P(B = 1)}$$

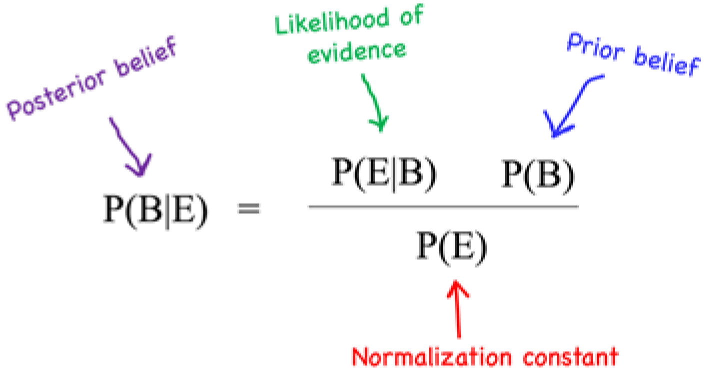
$$= \frac{P(B = 1 | L = \text{location i}) P(L = \text{location i})}{\sum_{i=1}^{16} P(B = 1 | L = \text{location i}) P(L = \text{location i})}$$

General form of LOTP!

Inference: Problem Wording

All inference problems have to give you:

1. A prior
2. An observation
3. A way to find the likelihood
4. A question asking you to update your belief



Inferring Age From C14

You observe a measurement of 900 C14 molecules in a sample. You assume that the sample originally had 1000 C14 molecules when it died. Infer $P(A = i|M = 900)$ where $A = i$ is the event that the sample organism died i years ago. Note that age is a discrete random variable which takes on whole numbers of years.

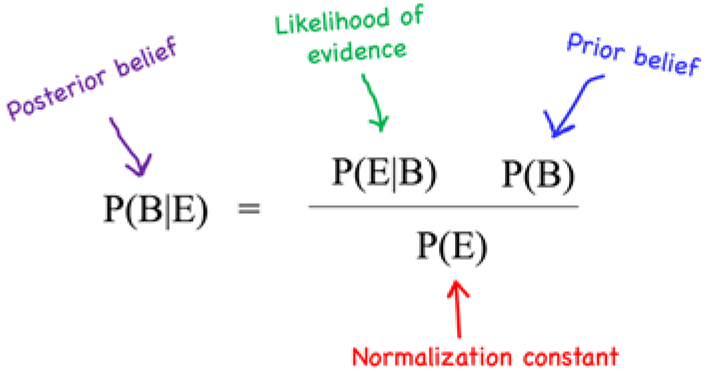
For your prior belief you know that the sample must be between $A = 100$ and $A = 10000$ inclusive and you assume that every year in that range is equally likely.

What is your belief distribution over the age of the sample?

Inference: Problem Wording

All inference problems have to give you:

1. A prior
2. An observation
3. A way to find the likelihood
4. A question asking you to update your belief



Inferring Age From C14

You observe a measurement of 900 C14 molecules in a sample. You assume that the sample originally had 1000 C14 molecules when it died. Infer $P(A = i|M = 900)$ where $A = i$ is the event that the sample organism died i years ago. Note that age is a discrete random variable which takes on whole numbers of years.

For your prior belief you know that the sample must be between $A = 100$ and $A = 10000$ inclusive and you assume that every year in that range is equally likely.

What is your belief distribution over the age of the sample?

Tricky problem

Let $X \sim \text{Exp}(\lambda = 1)$

Let $Y \sim \text{Poi}(\lambda = X)$

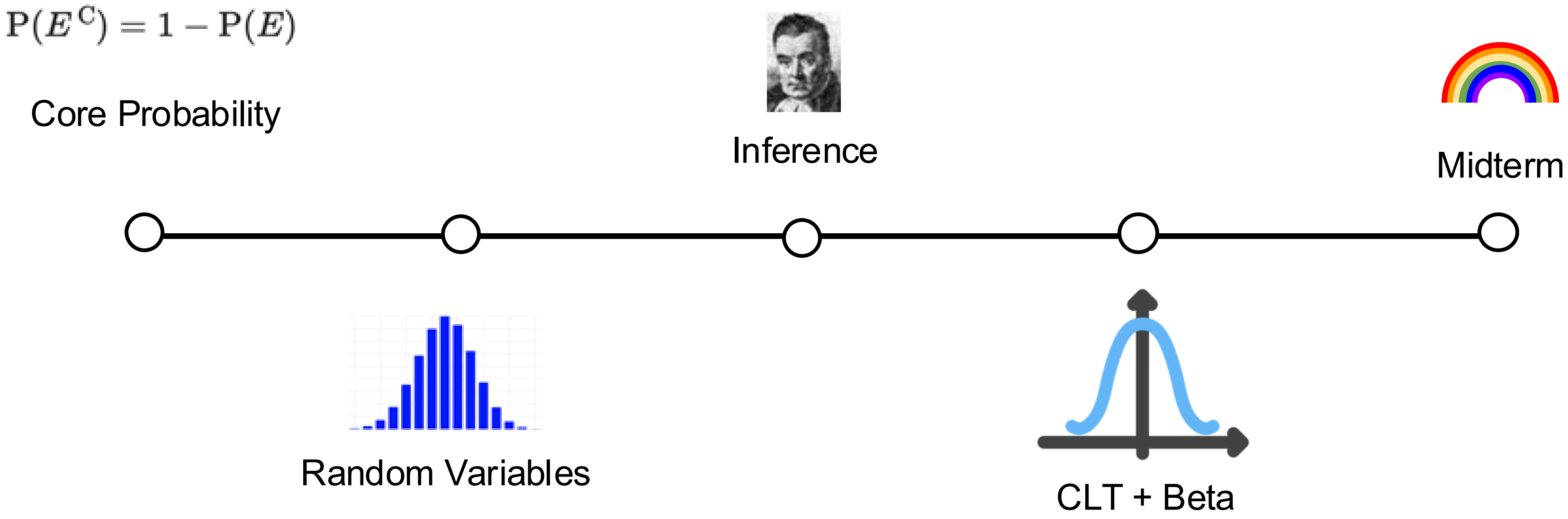
What is $P(Y = 2)$?

What is $P(X \mid Y = 2)$?

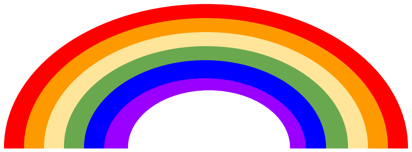
It is helpful to know this integral rule

$$\int_0^\infty x^{k-1} e^{-ax} dx = \frac{k!}{a^k}, \quad a > 0, k > 0$$

Roadmap So Far



The End



You got this :)