RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Probability & Uncertainty

Edgar Granados

Task Environments

(Remembering...)

Performance, Environment, Actuator, Sensors

- Observability: Full vs Partial
- Number: Single vs Multi
- Deterministic vs Stochastic
- Episodic vs Sequential
- Discrete vs Continuous
- Known vs Unknown

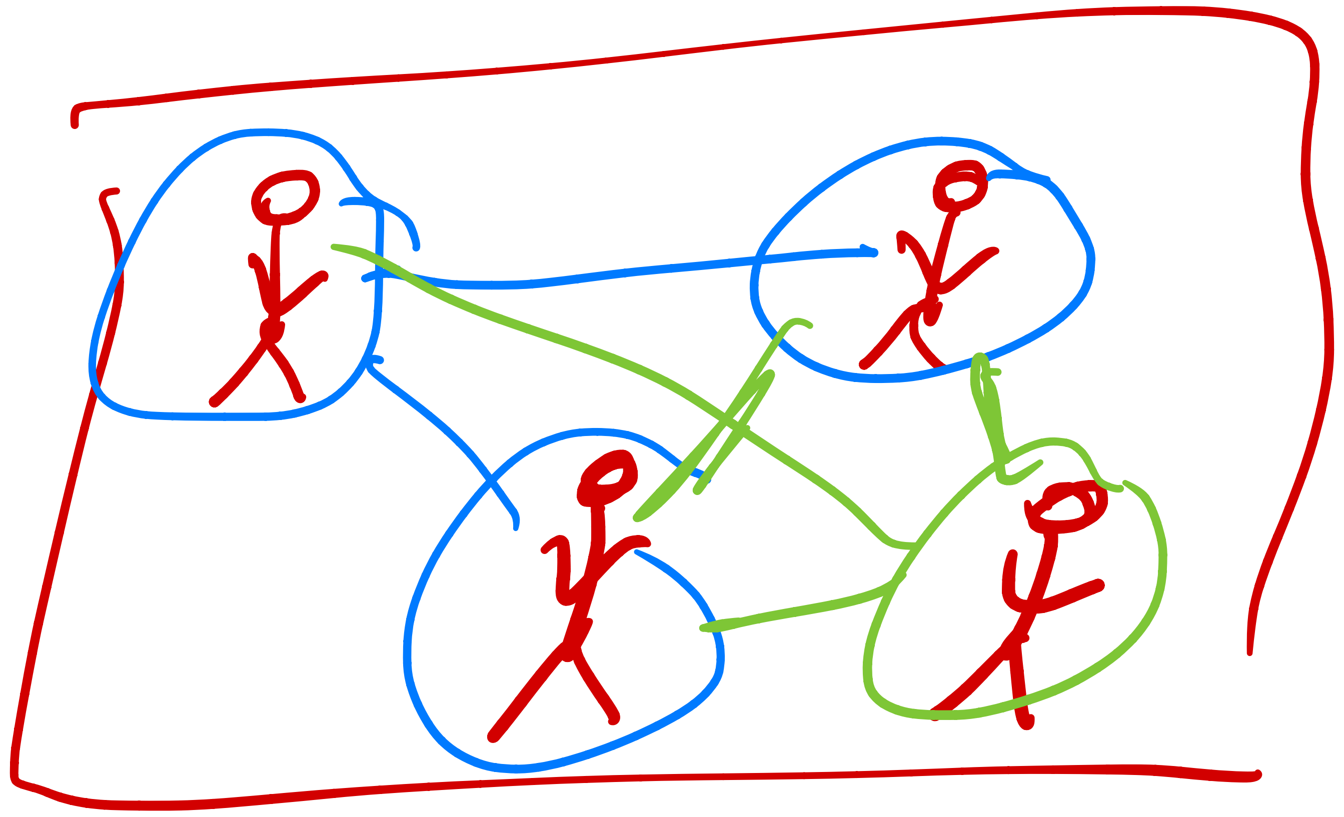
Example

Going to the Airport

- Option A_{120}
- Option A_{180}

More Examples

- Warranties
- Resupply/Restock
- The birthday problem



P persons

50% >

$P(2 \text{ people} \dots) \geq 50\%$

P

23

Probability

- Is a way of dealing with **uncertainty**
- Reason despite laziness or ignorance

Probability

- **Frequentist**
- **Subjectivism**

degrees of belief

Probability

Preliminaries

- The generalized basic principle of counting

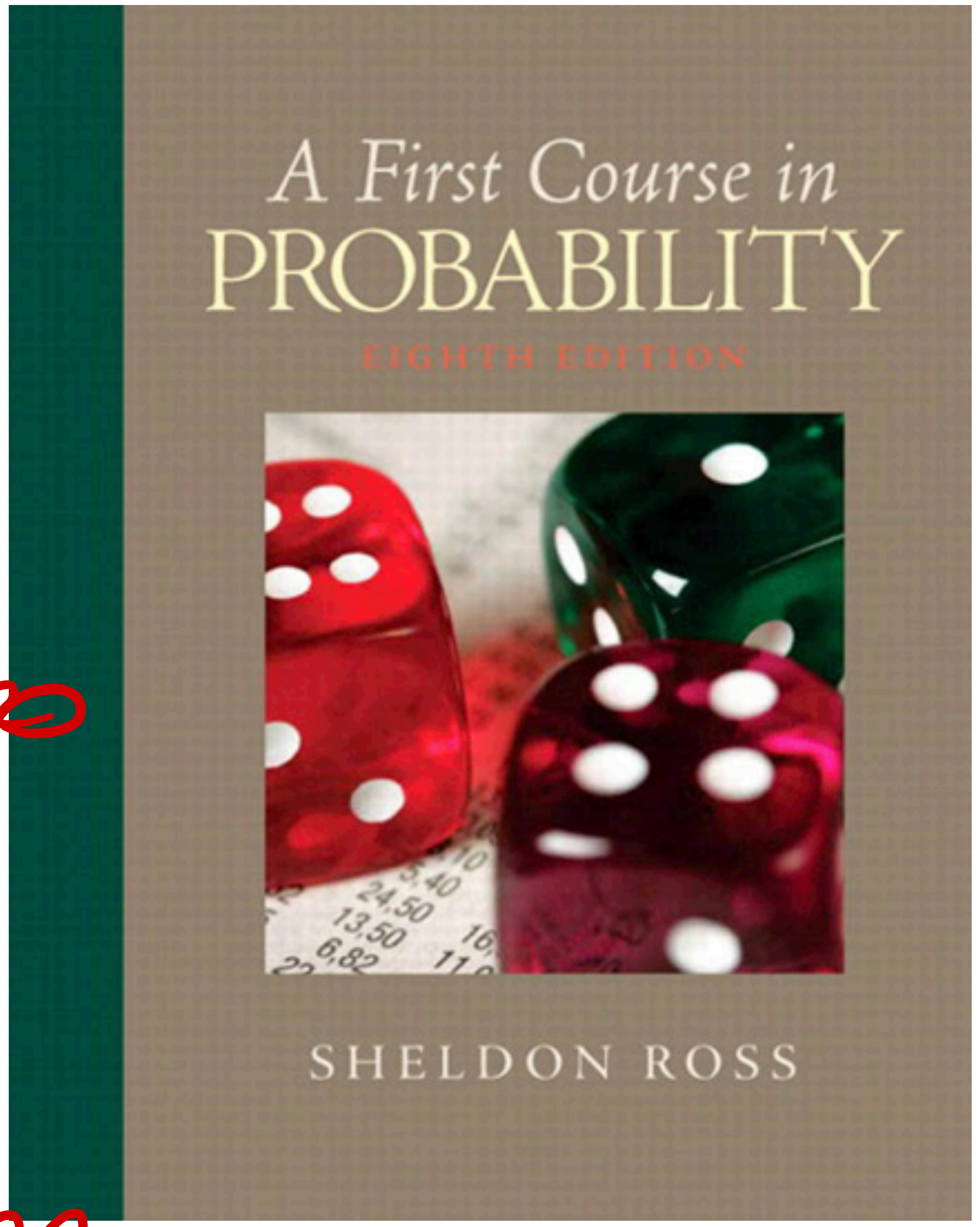
How many different 7-place license plates are possible if the first 3 places are to be occupied by letters and the final 4 by numbers?

26

$X_1 X_2 X_3 N_1 N_2 N_3 N_4$

$26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10$

$= 175,760,000$



How many license plates would be possible if repetition among letters or numbers were prohibited?

$X_1 X_2 X_3 N_1 N_2 N_3 N_4$

$26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 \cdot 7 = 78,624,000$

Probability

Preliminaries

- Permutations
  - Different permutations of the n objects $$\boxed{n(n-1)(n-2)\cdots 3\cdot 2\cdot 1 = n!}$$
  - Different permutations of n objects, of which n_1 are alike, n_2 are alike, ..., n_r are alike. $$\frac{n!}{n_1!\,n_2!\,\cdots\,n_r!}$$ $\rightarrow 2\cdot 2\cdot 2$
- Combinations
  - Number of possible combinations of n objects taken r at a time $$\binom{n}{r} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{(n-r)!\,r!}$$

Probability

Definitions

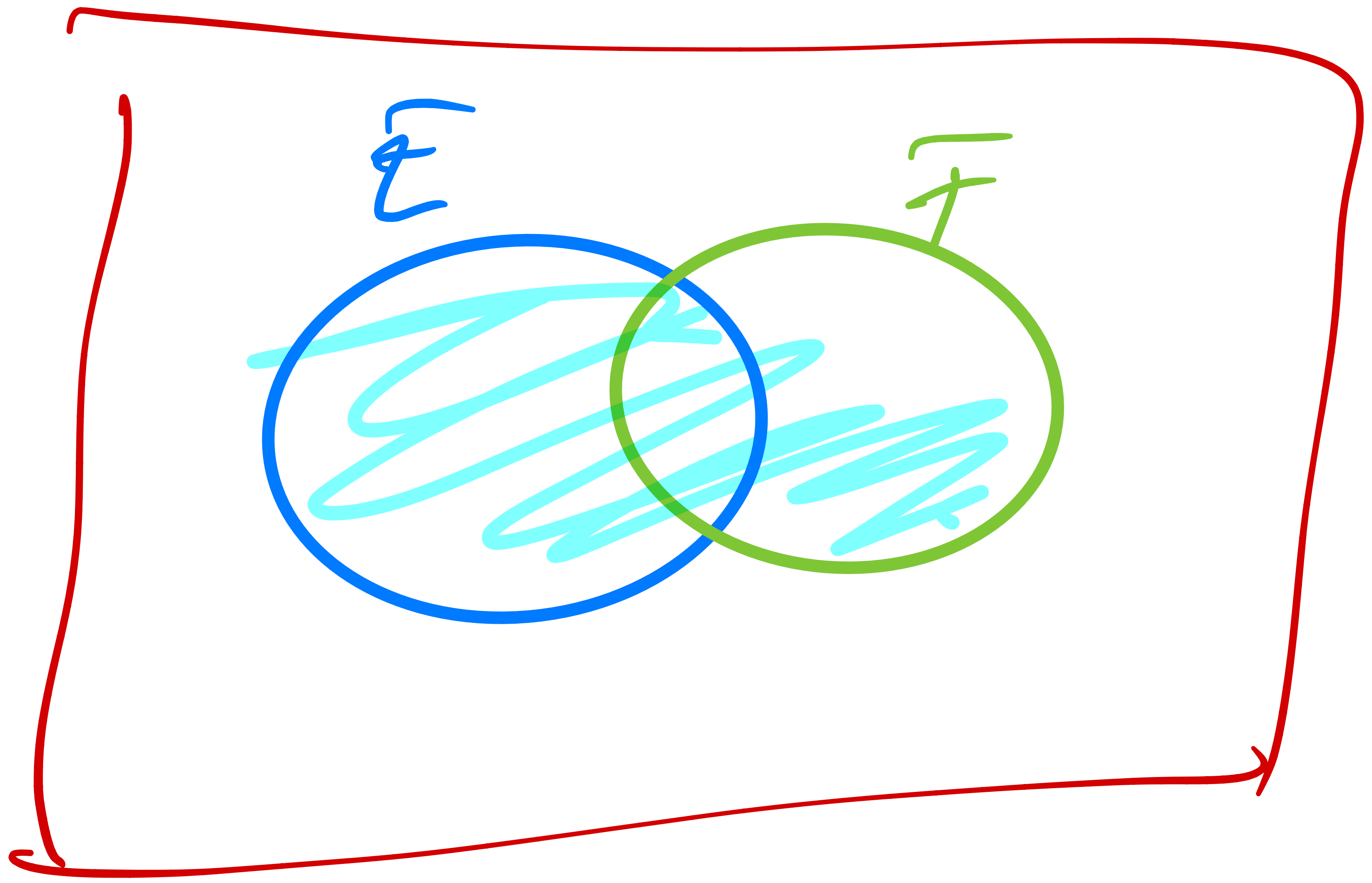
- **Sample Space:** set of all possible outcomes of an experiment
- **Event:** Any subset of the sample space

Probability

Definitions

- **Sample Space:** set of all possible outcomes of an experiment
- **Event:** Any subset of the sample space

Commutative laws	$E \cup F = F \cup E$	$EF = FE$
Associative laws	$(E \cup F) \cup G = E \cup (F \cup G)$	$(EF)G = E(FG)$
Distributive laws	$(E \cup F)G = EG \cup FG$	$EF \cup G = (E \cup G)(F \cup G)$



Probability

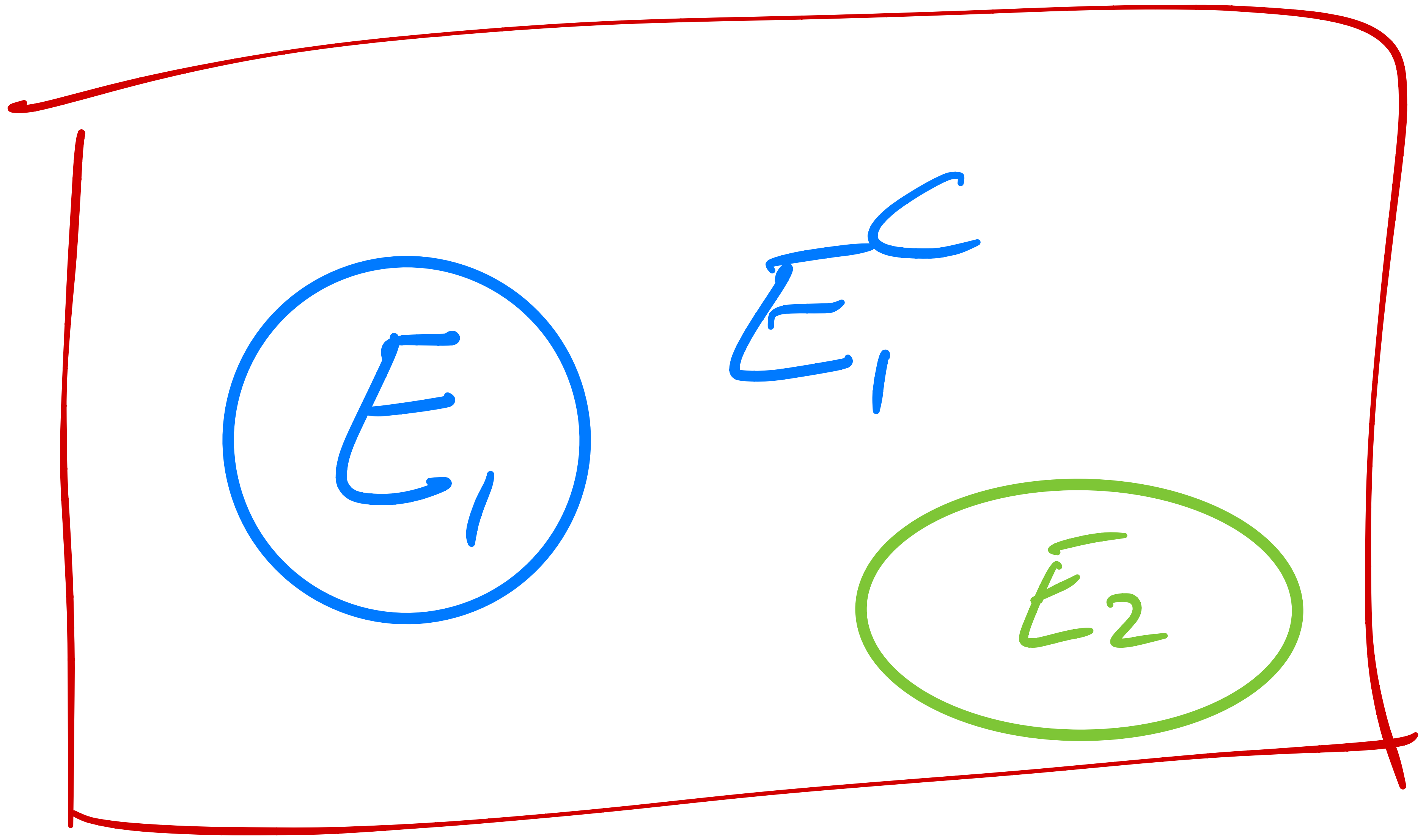
Definitions: DeMorgan's laws

- **Sample Space:** set of all possible outcomes of an experiment
- **Event:** Any subset of the sample space

E_i

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$$

$$\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$



Probability

Definitions

- **Sample Space:** set of all possible outcomes of an experiment
- **Event:** Any subset of the sample space
- **Probability:** $$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

$E \sim H$

Probability

Axioms

- **Axiom 1:**

100%

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

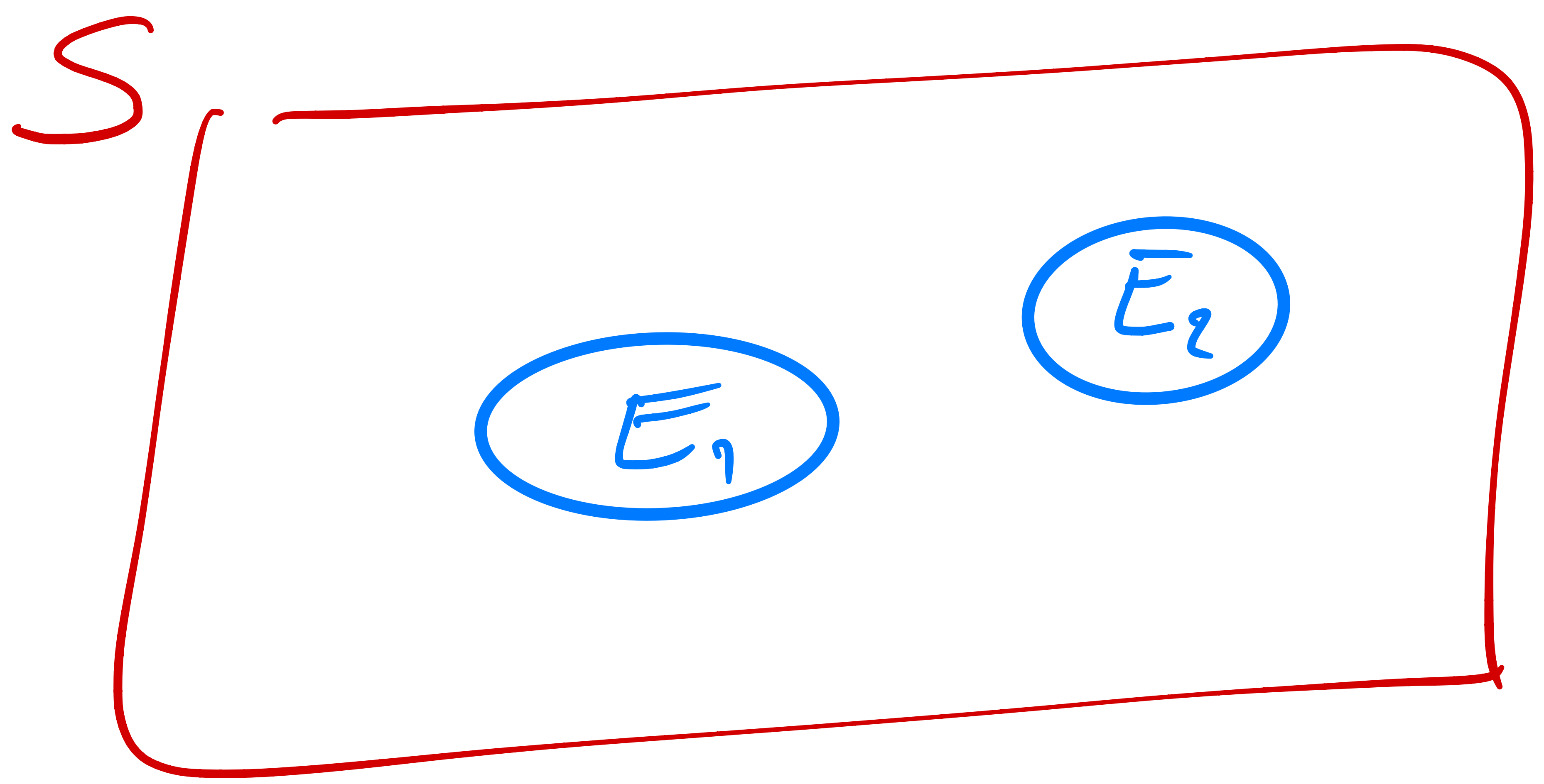
- **Axiom 2:**

$$P(S) = 1$$

$S = E \cup E^c$

- **Axiom 3:**

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

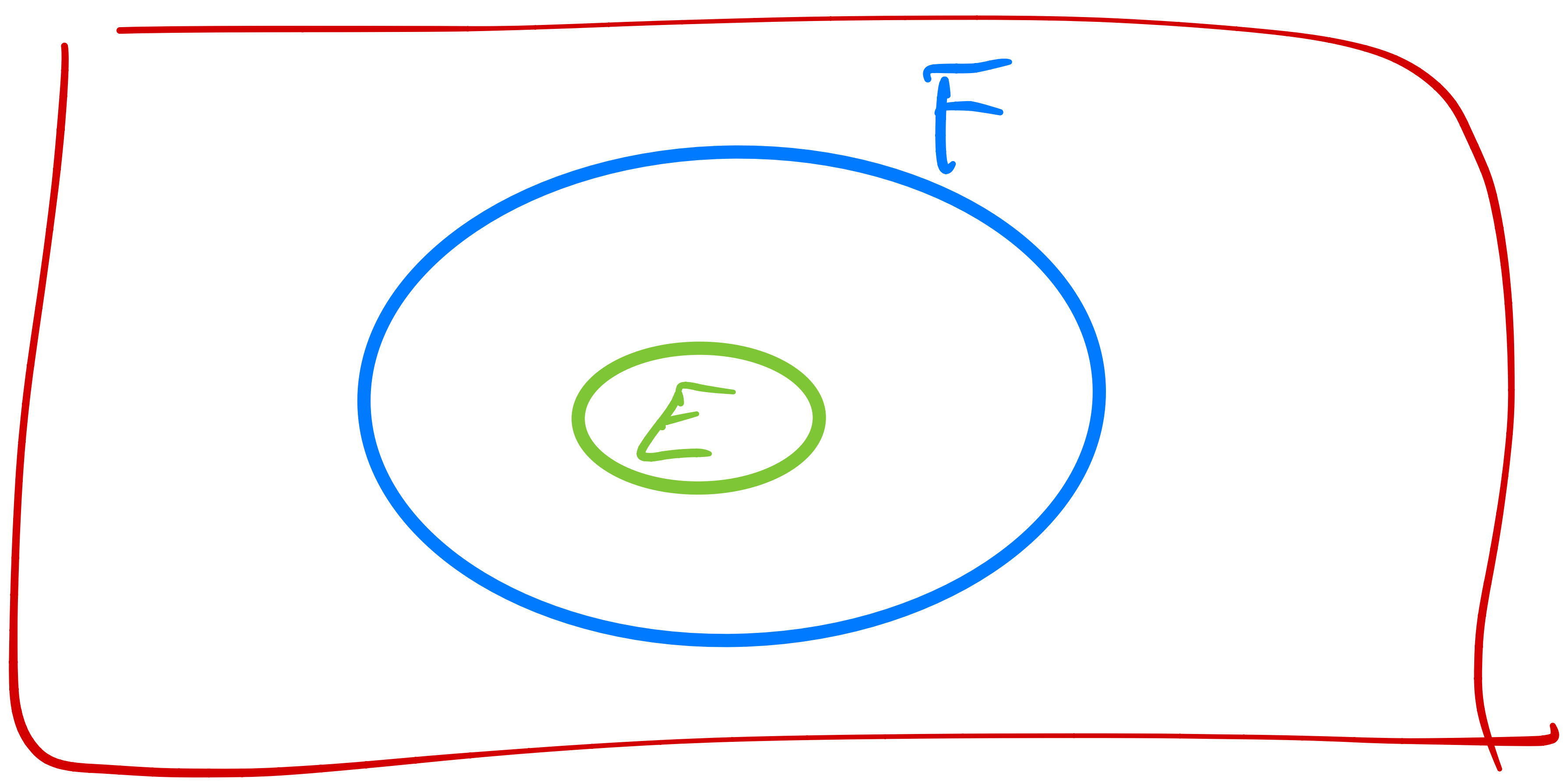


Probability

Propositions

- **Axiom 1:** $0 \leq P(E) \leq 1$
- **Axiom 2:** $P(S) = 1$ $\quad S = E \cup E^c$
- **Axiom 3:**

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$



- Propositions:

  1. $P(E^c) = 1 - P(E)$
  2. if $E \subset F \Rightarrow P(E) \leq P(F)$

Don't Include this

Probability

Probability Space

Event Space

- **A triple** (Ω, E, P)

Sample Space

Probability Measure

$P(\emptyset) = 0$

$P(\Omega) = 1$

$P(S) = 1$

Probability Space

Example

- What is the probability space (Ω, E, P) of tossing 1 coin

$\Omega: \{H, T\}$

$E: \{\{H\}, \{T\}\}$

$P(H) = 0.5$

$P(T) = 0.5$

Probability Space

Example

- What is the probability space (Ω, E, P) of tossing $\overset{n}{3}$ coins

$\to \Omega = \{HHH, HHT, HTH, (\dots), TTT\} \rightsquigarrow \underline{8}$

$E = \{\{HHH\}, (\dots)\}$

$P(HHH) = \frac{1}{8}$

$P(\text{At least } 2\ H)$

Probability Space

Example

- What is the probability space (Ω, E, P) of tossing 1 dice

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$E = \{1, 2, 3, 4, 5, 6\}$$

$$i \in [1, 6] \quad P(E_i) = \frac{1}{6}$$

Probability Space

Example

- What is the probability space (Ω, E, P) of tossing 2 dice

$\Omega = \{(1,1), (1,2), (\ldots), (6,6)\}$

$E = \{(1 \text{ dice} > 5)\}$

$\boxed{P((1,1)\ (5,5))} = \frac{1}{36}$

Inclusion-Exclusion Principle

Example

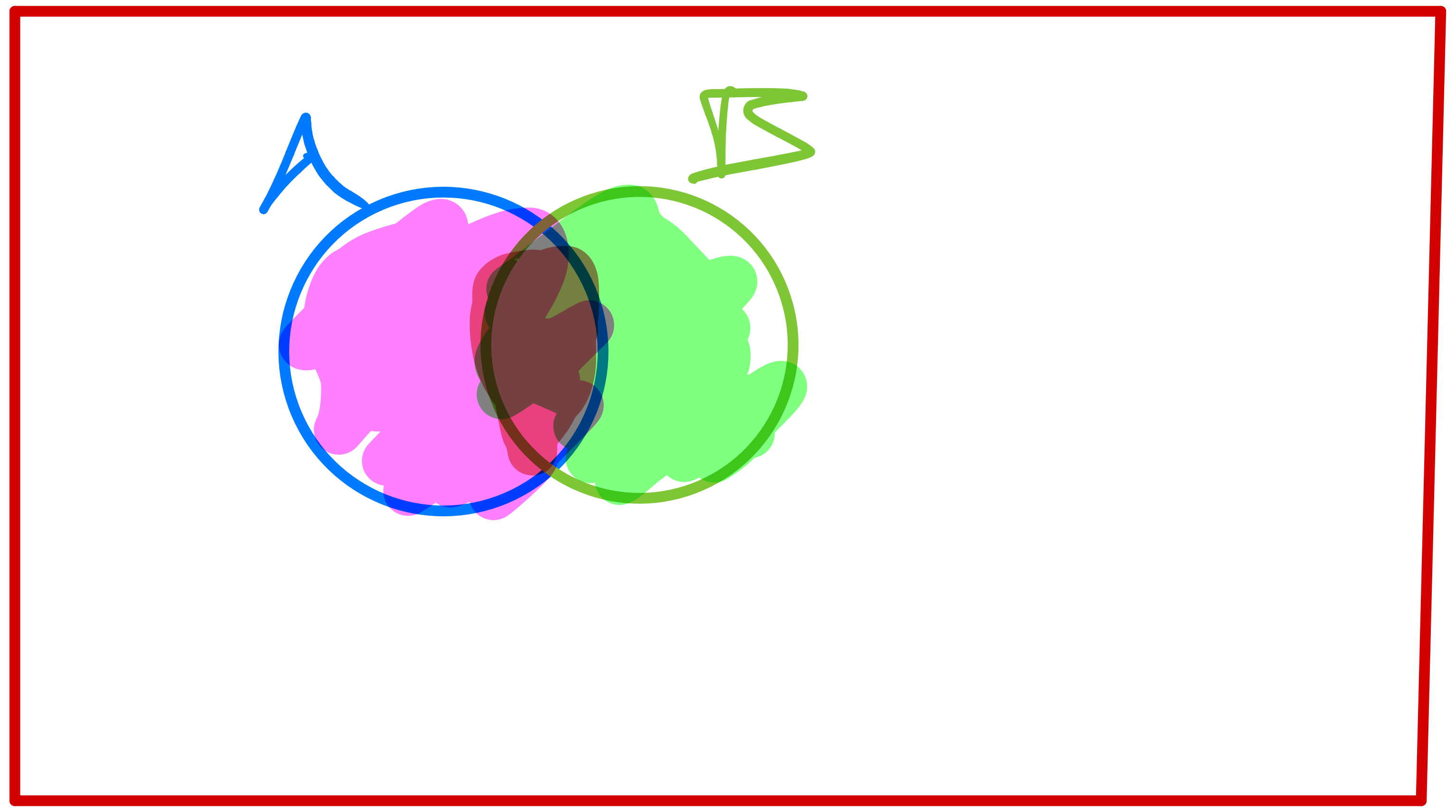
- Tossing **two** dice, what is the probability of
  - **Event**: at least one dice gives a number of 5

A: d_1 yields 5

B: d_2 yields 5

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

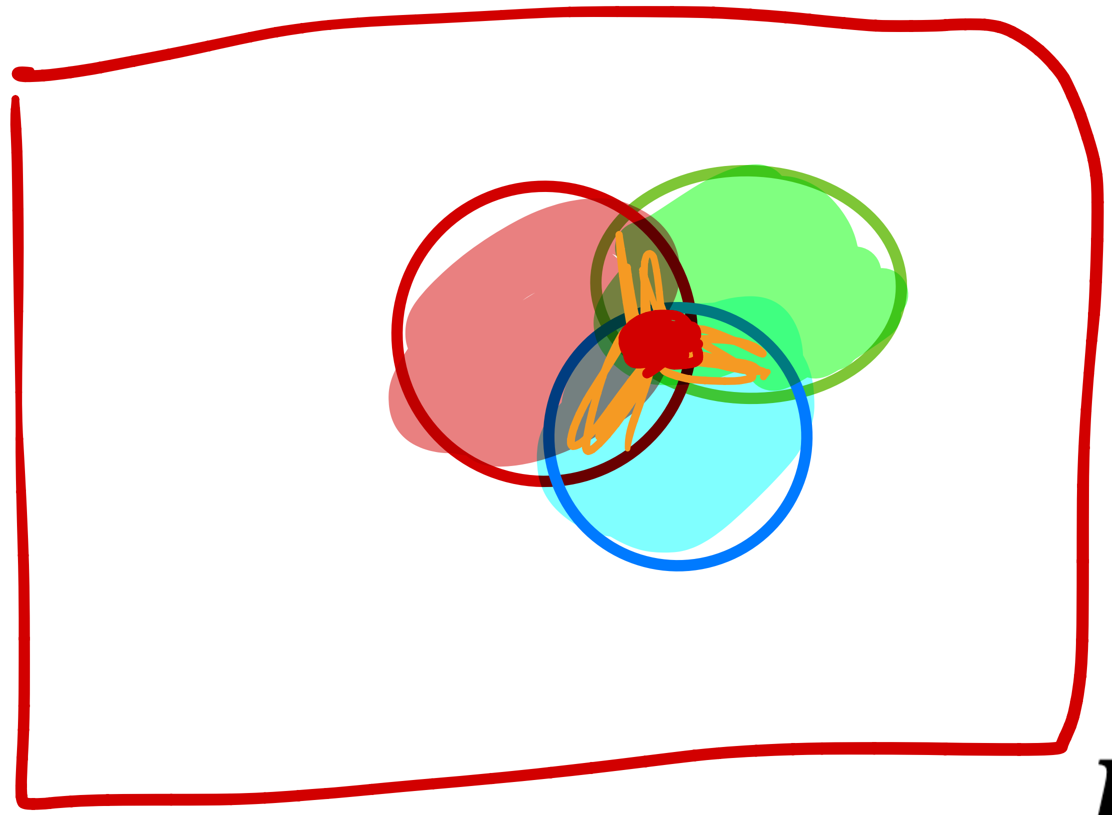
$$\frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36}$$



(5,5)

Inclusion-Exclusion Principle

Definition



$P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)$
$-P(B\cap C)$
$+P(A\cap B\cap C)$

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

$$P(A_1\cup\cdots\cup A_n) = \sum_{1\le i\le n} P(A_i) - \sum_{1\le i\le j\le n} P(A_i\cap A_j) + \sum_{1\le i\le j\le k\le n} P(A_i\cap A_j\cap A_k) + \cdots$$

Inclusion-Exclusion Principle

Example

- Tossing **THREE** dice, what is the probability of
  - **Event**: at least one dice gives a number of 5

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} - \frac{1}{36} \cdot 3 + \frac{1}{216} = \boxed{\frac{91}{216}}$$

$$\frac{1}{6} \cdot \frac{1}{6} \qquad \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$$

Probability

Continuous Domains

- (Ω, E, P) Still applies!
- Continuous dice?

$F(x) = \lfloor x \rfloor \quad x \in [1, 7]$

$G(x) = x$

$P(G(x) \leq 5)$

$P(4 \leq G(x) \leq 5)$

Probability

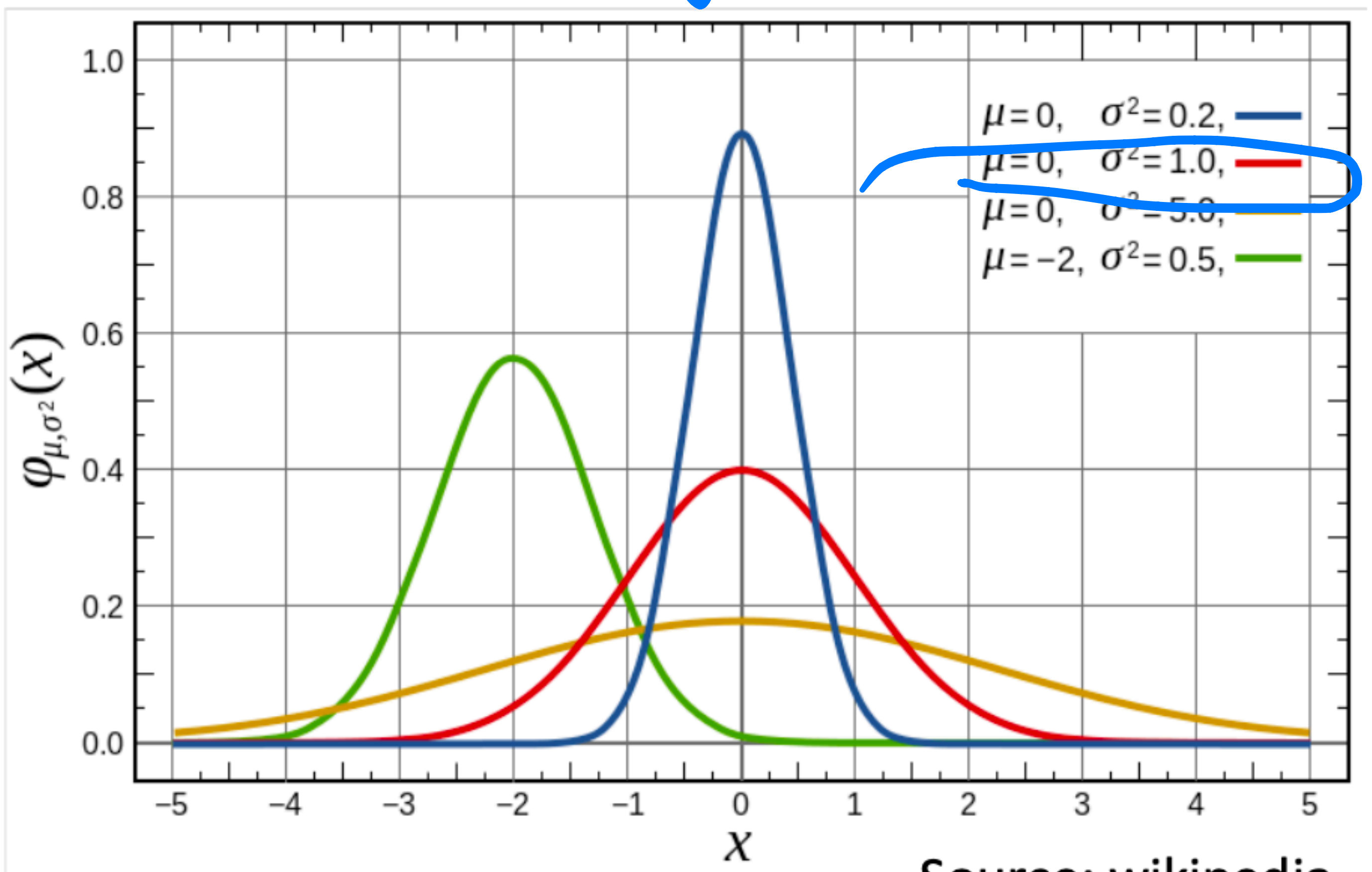
Continuous Domains

- (Ω, E, P) Still applies!

$N(0, 1)$

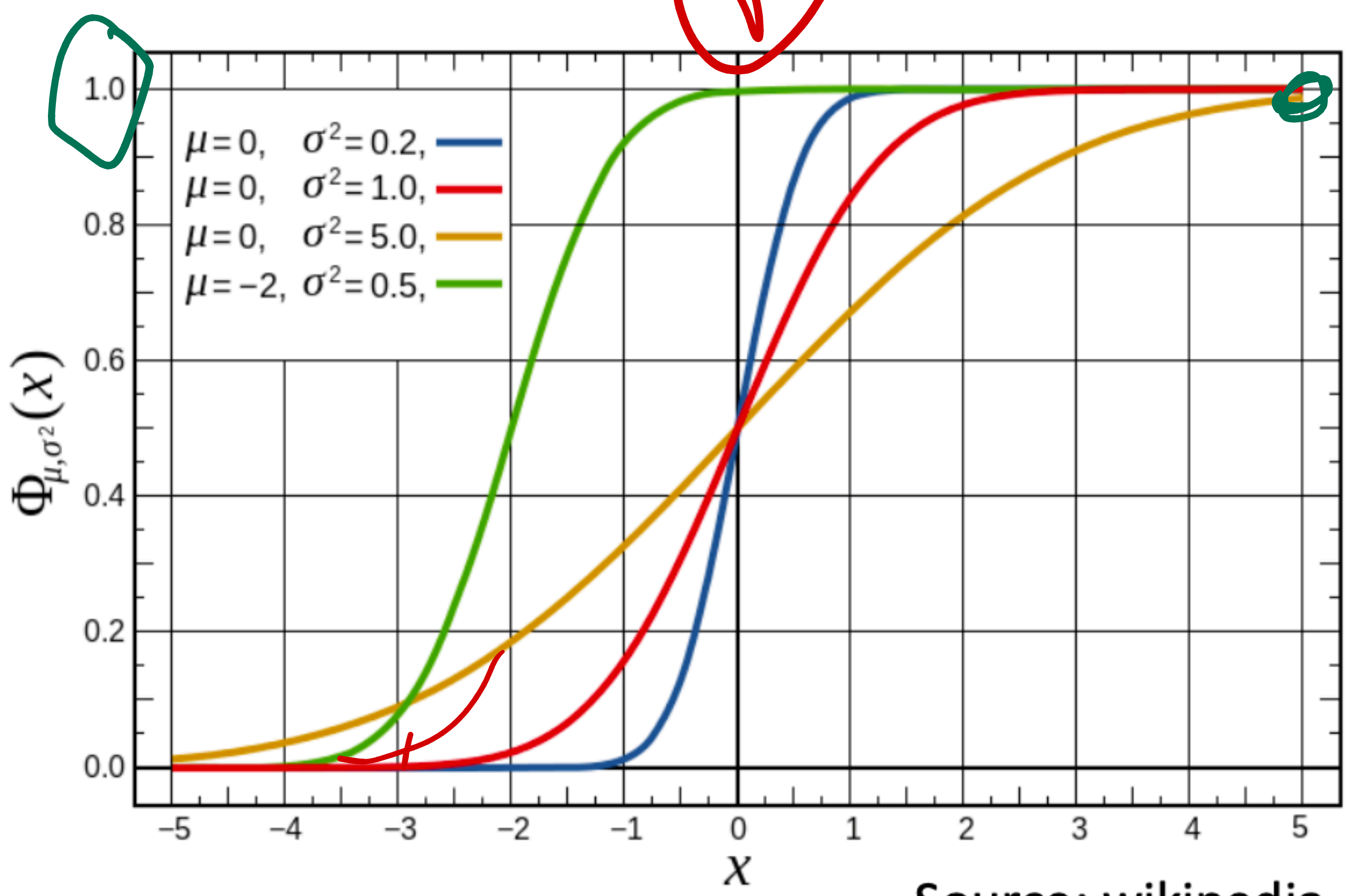
$N(\mu, \sigma^2)$

$N(\mu, \Sigma)$



Source: wikipedia

Probability density function (PDF)

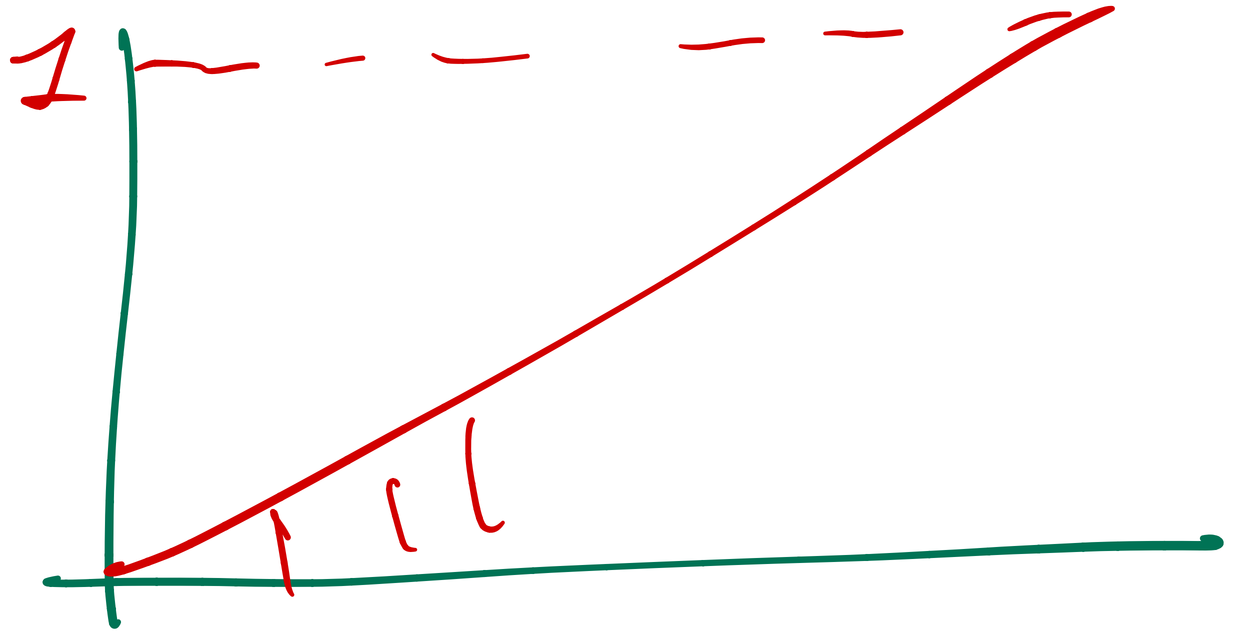


Source: wikipedia

Cumulative distribution function (CDF)

Uniform (PDF)

Uniform (CDF)



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

$P : P(a \leq x \leq b) = \int_a^b f(x|\mu, \sigma^2)\, dx$

Random Variable and Expectation

RV $\quad X : \Omega \to E$

- A random variable is a function from the sample space to some other space
- Sum, difference, (...) of RVs?

$X+Y$

$X-Y$

αX

XY

$f(X)+g(Y)$

$$X_c(\omega) = \begin{cases} 1, & \omega = \text{head} \\ 0, & \omega = \text{tail} \end{cases}$$

$X_d(\omega) = i$ for face i

$\Omega = \{f_1, f_2, \ldots, f_6\}$

Random Variable and Expectation

- A random variable is a function from the sample space to some other space
- Expected value of a random variable

$$\mathbb{E}[X] = \sum x_i P(X = x_i) = x_1 P(x_1) + (\ldots) + x_n P(x_n)$$

$$E[X] = \sum_{x:p(x)>0} x p(x)$$

Random Variable and Expectation

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

Example

- Tossing one fair dice, what is the expected face value? → mean

$$E[X] = \frac{1}{6}(1) + \frac{1}{6}(2) + (\ldots) + \frac{1}{6}6 = 3.5$$

Random Variable and Expectation

$$E[X] = \boxed{\sum_{x:p(x)>0} xp(x)}$$

Example

- Tossing **two** fair dice, what is the expected face value?

$$E[X_1 + X_2] = E[X_1] + E[X_2]$$

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

$$E[X_{d_1} + X_{d_2}] = 3.5 + 3.5 = 7$$

Random Variable and Expectation

$$E[X] = \sum_{x:p(x)>0} x p(x)$$

Example

- Expected number of tosses to get both sides of a fair coin? $\rightarrow 3$

$$\boxed{P(E) = p \quad \frac{1}{p}}$$

$$1 + \underbrace{2}_{} = 3$$

$$\mathbb{E}[\text{tosses} \ldots] = \mathbb{E}[1^{st}] + \mathbb{E}[2^{nd}]$$

Random Variable and Expectation

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

Example

- Expected number of tosses to get all sides of a dice?

$$\mathbb{E}[\text{tosses to get 6s}] = \mathbb{E}[1^{st}] + \mathbb{E}[2^{nd}] + (\ldots) + \mathbb{E}[6^{th}]$$

$$= \boxed{\frac{6}{6}} + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \boxed{\frac{6}{1}}$$

Expectation

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

Example

$P(A_{720}) = .99$

- Airport example

$1 - .85$ $\quad$ $1 -$

$$\boxed{P(A_{120}\ works) = 0.85},\ P(\boxed{A_{180}\ works}) = \boxed{0.975}$$

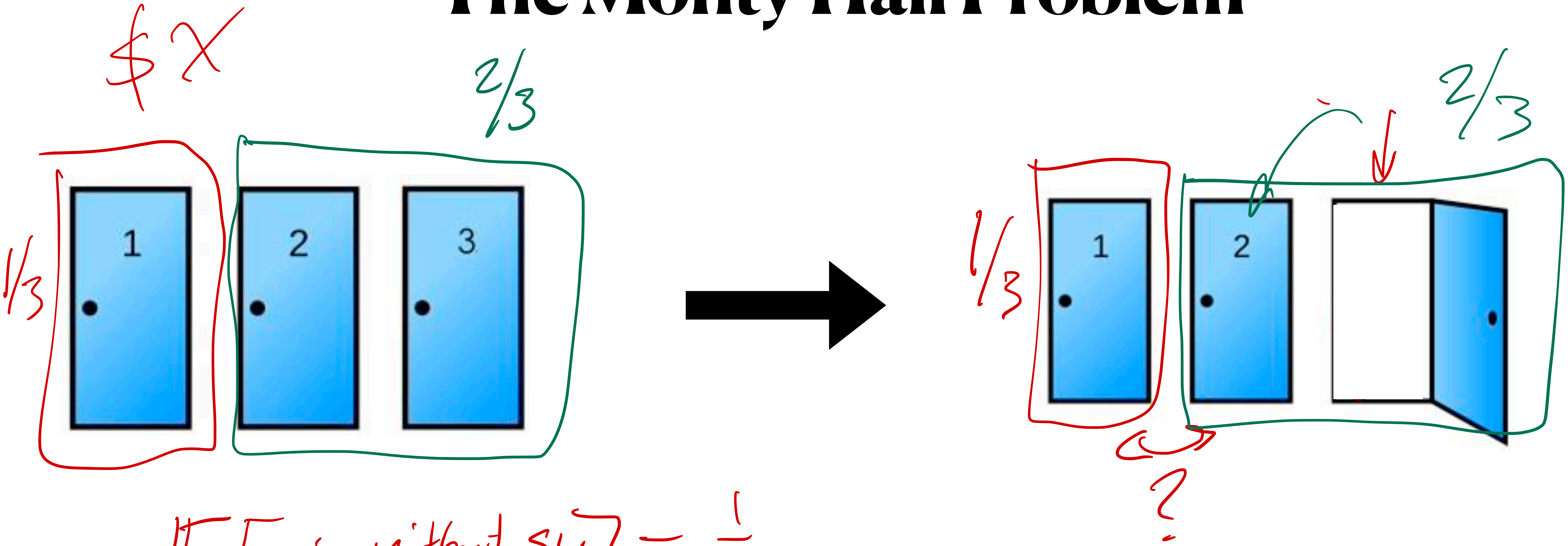
- Ticket cost: $\boxed{\$2k}$
- Extra hour at airport: $\boxed{\$200}$

$1 \cdot 200$

$\mathbb{E}[\text{loss } A_{120}] = 2k \cdot (.15) + = 300$

$\mathbb{E}[\text{loss } A_{180}] = 2k \cdot (.025) + \cancel{1}\ 2 \cdot 200 = 250$

The Monty Hall Problem



$$\mathbb{E}[\text{win without sw}] = \frac{1}{3}$$

$$\mathbb{E}[\text{win sw}] = \frac{2}{3}$$

Joint Probability Distributions

atomic event — Ω

- A joint probability distribution assigns probabilities to all atomic events, e.g., for toothache, cavity, and catch (positive diagnosis), a joint probability distribution is

$P(t, catch, \neg cavity)$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Joint Probability Distributions

Notation

- $P(X_1 = x_1, \ldots, X_n = x_n)$ represents the probability of a single atomic event
- Can also write as $P(x_1, \ldots, x_n)$ when the context is clear
- $\boxed{P(X_1, \ldots, X_n)}$ refers to the general probability distribution

X x

Marginal Probability Distributions

- A marginal distribution is the (single or joint) distribution over a $\boxed{\text{subset}}$ of random variables of a joint distribution

$P(A, B, C, D, E) \rightarrow$ Marginal dist $P(A, B, C)$, $P(A)$, $P(C, D)$

- From a full joint distribution, all marginal distributions can be easily computed using marginalization

$P(X_1, (\ldots), X_m)$ from $P(X_1, (\ldots), X_m, \boxed{X_{m+1}, (\ldots), X_n})$ $\quad m < n$

$$P(X_1 = x_1, \ldots, X_m = x_m) = \sum_{\text{all choices of } x_{m+1}, \ldots, x_n} P(\boxed{X_1 = x_1, \ldots, X_m = x_m}\, \boxed{X_{m+1} = x_{m+1}, \ldots, X_n = x_n})$$

Marginal Probability Distributions

Example

$X_2 = \neg cavity = \neg X_2$

- Say X_1, X_2, X_3 represents the RVs for toothache, cavity, and catch $\begin{cases} X_3 = catch \\ X_3 = \neg catch \end{cases}$
- Compute $P(X_1, X_2) \Rightarrow P(X_1 = t, X_2 = cavity) = P(X_1 = t, X_2 = cav.,$ $= 0.12$

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$.108 + .012 = .12$

$P(X_1 = t)$

Conditional Probability

- The probability of some events assuming that some other (not necessarily mutually exclusive) events happen

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

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

$$P(A|\underbrace{B,C}_{D}) = \frac{P(A \cap D)}{P(D)}$$

Conditional Probability

- The probability of some events assuming that some other (not necessarily mutually exclusive) events happen
- Example: compute P(toothache | Cavity) $= \frac{P(X_1 \cap X_2)}{P(X_2)} = \frac{.12}{.12 + .08} = .6$

	t	$\neg t$
Cav	.12	.08
$\neg$Cav	.08	.72

Chain Rule or Product Rule

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

- The probability of some events assuming that some other (not necessarily mutually exclusive) events happen
- Example: compute P(toothache | Cavity)

$$P(X_1, \dots, X_n) = P(X_1)P(X_2|X_1)\dots P(X_n|X_1, \dots, X_{n-1})$$

$$= \prod_{i=1}^{n} P(X_i|X_1, \dots, X_{i-1})$$

Independence

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

$$\boxed{P(X,Y) = P(X)P(Y)}$$

$$\frac{1}{6} \cdot \frac{1}{6} \cdot$$

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

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

Independence

$$P(X,Y) = P(X)P(Y)$$

Conditional Independence

$P(X,Y) = P(X)P(Y)$

- $\boxed{X, Y}$ are conditionally independent given Z

$$P(X,Y|Z) = \boxed{P(X|Z)P(Y|Z)}$$

$$P(X|Y,Z) = P(X|Z)$$

$$P(Y|X,Z) = P(Y|Z)$$

Bayes' Rule

$P(X,Y) = P(X)P(Y|X) = P(Y)P(X|Y)$

$$\boxed{P(X|Y)} = \frac{P(X,Y)}{P(Y)} = \frac{P(Y|X)P(X)}{P(Y)}$$

$$\boxed{P(H|D)} = \frac{\boxed{P(D|H)}\,\boxed{P(H)}}{P(D)}$$

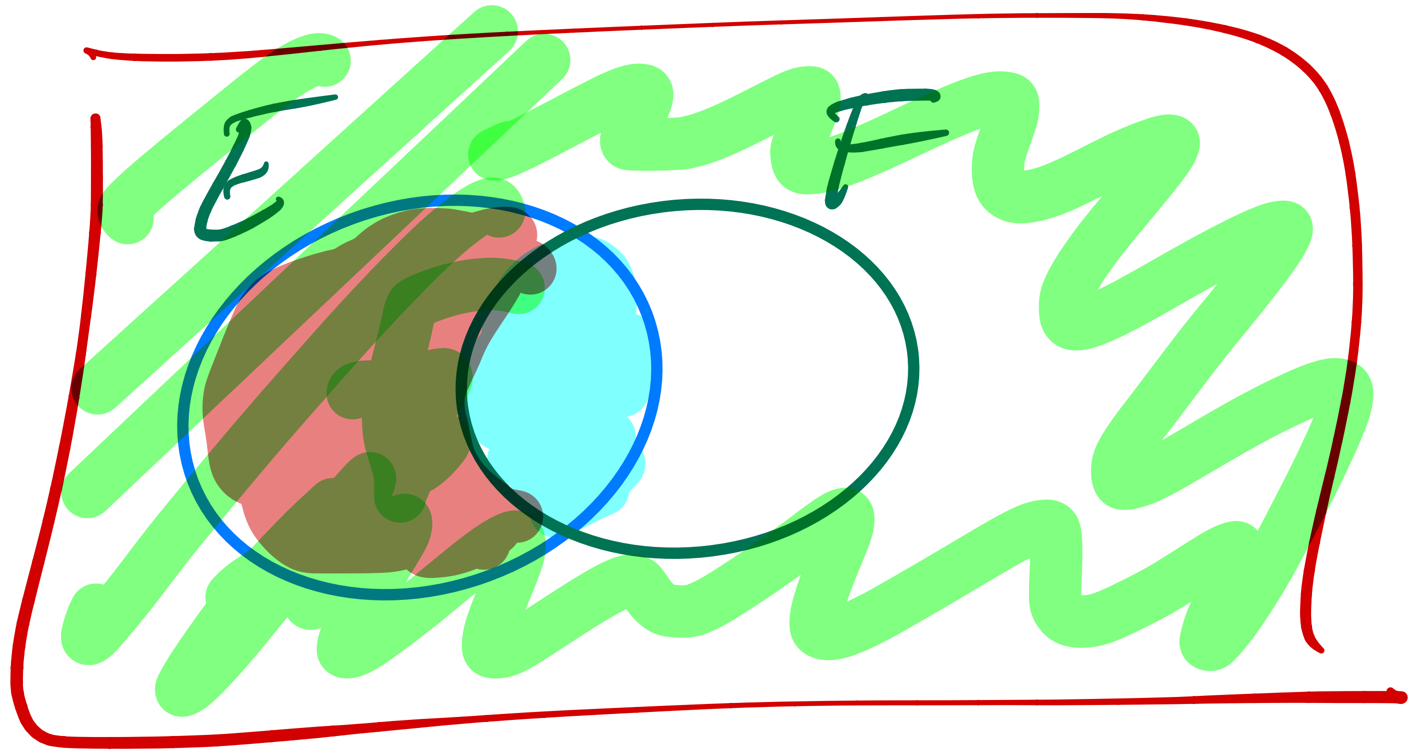
Hypothesis based on Data

Current Belief

Data Collected Given a Hypothesis

Bayes' Rule

$$P(X|Y) = \frac{P(X,Y)}{P(Y)} = \frac{P(Y|X)P(X)}{P(Y)}$$



- How to compute $P(\cancel{Y}E)$?

$E = EF \cup EF^c$

$P(E) = \underline{P(EF)} + P(EF^c)$

$= P(E|F)\cdot P(F) + P(E|F^c)P(F^c)$

$= P(E|F)\cdot P(F) + P(E|F^c)(1 - P(F))$

The Importance Bayes' Rule

Hypothesis based on Data

$$\boxed{P(H \mid D)} = \frac{\boxed{P(D \mid H)}\,\boxed{P(H)}}{P(D)}$$

Current Belief

Data Collected Given a Hypothesis

Don't include this

$P(H)$ ~ Prior

$P(H \mid D)$ ~ Posterior

$P(D \mid H)$ ~ likelihood

$P(D)$ ~

$P(H \mid D) \propto P(D \mid H)\, P(H)$ Max a Posteriori

$P(H \mid D) \propto P(D \mid H)$ Maximum Likelihood