Week 1: Basic Concepts

Experiment, Outcome, Sample space, Event, Probability

Textbooks for the course

- Probability and Statistics by Siva Athreya, Deepayan Sarkar and Steve Tanner
  - https://www.isibang.ac.in/~athreya/psweur/
  - Work in progress
  - This course will follow this book closely
- Other useful books
  - Think Stats 2 by Allen B Downey (https://greenteapress.com/wp/think-stats-2e/)
  - A Modern Introduction to Probability and Statistics by F. M. Dekking, C. Kraaikamp, H. P. Lopuhaä, L. E. Meester
- There are several other books for Probability and Statistics
  - No book is strictly necessary, but...

Experiment and outcome

- **Experiment**: Process or phenomenon that we wish to study statistically

Images: wikipedia commons, pixy.org

- **Outcome**: Result of the experiment (in as much detail as necessary)

Heads or Tails

1 or 2 or 3 or 4 or 5 or 6
(one die)

Yaml format file
(lots of details)

Sample space

DEFINITION (**Sample space**) A sample space is a set that contains all outcomes of an experiment.

- Sample space is a set, typically denoted S
- Examples
  - Toss a coin: S = { heads, tails }
  - Throw a die: S = { 1, 2, 3, 4, 5, 6 }
  - IPL: too big to write down completely
    - Runs scored in one delivery: { 0, 1, 2, 3, … }
    - Winner in 2021: { CSK, MI, DC, … } (9 teams in 2021)
- Quite often, in practice, it is enough to “imagine” a sample space

More examples and observations

- Draw a marble from an urn (pot or jar) with marbles of many colours
  - Say, for example, 3 each of red, white, blue marbles
  - Sample space: { red, white, blue }
  - What happens when you draw two marbles? With replacement? Without replacement?
    - Sample space starts to get more complicated
- Draw a card from a shuffled pack of 52 cards
  - Sample space: { Spades, Hearts, Diamonds, Clubs } x { 2, 3, …, 10, J, Q, K, A }
    - Recall cartesian product of sets
  - Drawing multiple cards or distributing cards to players?
- As we learn more probability, we will start working without even thinking about the sample space
  - However, sample space is an important theoretical notion
  - When interpreting probabilities becomes confusing, thinking of sample space can be of use

Event

DEFINITION (**Event**) An event is a subset of the sample space. There is a technical restriction on what subsets can be events. We will ignore this restriction for now, and point it out later when necessary.

- Toss a coin: S = { heads, tails }
  - Events: empty set, {heads}, {tails}, { heads, tails }
  - 4 events
- Throw a die: S = { 1,2,3,4,5,6 }
  - Events: empty set, {1}, {2}, {3}, {4}, {5}, {6}, {1,2}, {1,3}, … , {2,3,4,5,6}, {1,2,3,4,5,6}
  - 64 events
  - Some can be described in words: getting an even number, getting a multiple of 3 etc
- Fisherman goes out to fish: Sample space?
  - Event: Catch is more than 100 Kg, Pomfret is in the catch etc
  - Though we have not written down the sample space, we can make sense of above events

Events are central objects in probability theory

- An event is said to have “occurred” if the actual outcome of the experiment belongs to the event.
- Events are sets
    - All set theory notions apply to events
- One event can be contained in another, i.e. $A \subseteq B$
    - Throw a die: A = {2,6}, B = even number
    - If A occurred, B has also occurred
    - If B occurred, A may or may not have occurred
- Complement of an event A, denoted A^c = { outcomes in S not in A } = ($S \setminus A$)
    - Throw a die: A = { 2, 4, 6 } (even), A^c = { 1, 3, 5 } (odd)
    - If A occurred, A^c did not occur
    - If A^c occurred, A did not occur

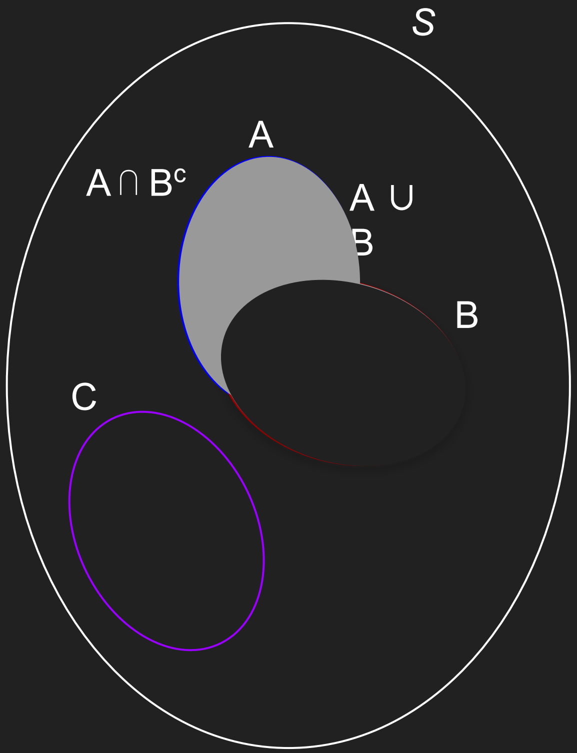
Combining events to create new events

- Since events are subsets, one can do complements, unions, intersections
- Union ("or" in English), denoted $\cup$
  - Throw a die: even number or a multiple of 3, even number or a prime number
    - Can you work out the above events?
  - Fisherman's catch: more than 200 Kg or less than 50 Kg
- Intersection ("and" in English), denoted $\cap$
  - Throw a die: even number and a multiple of 3, even number and a prime number
  - Fisherman's catch: more than 100 Kg and less than 150 Kg
- Many interesting events can be written as intersections and unions
  - Draw 5 cards from a pack without replacement: no aces
    - {1st not ace} $\cap$ {2nd not ace} $\cap$ {3rd not ace} $\cap$ {4th not ace} $\cap$ {5th not ace}
  - IPL: 5 runs off the bat in two legal deliveries
    - $\{1 + 4\} \cup \{2 + 3\} \cup \{3 + 2\} \cup \{4 + 1\}$

Disjoint events

- Two events with an empty intersection are said to be disjoint events
  - Throw a die: even number, odd number are disjoint
  - Fisherman's catch: more than 200 Kg, less than 50 Kg are disjoint
- Suppose A and B are disjoint events
  - If A occurred, B did not occur
  - If B occurred, A did not occur
- Event and its complement
  - A and A^c are disjoint, $A \cap A^c$ = empty set
  - Together, they cover all outcomes, $A \cup A^c = S$
    - Either A occurs or A^c occurs
- Multiple events: E_1, E_2, E_3, …. are disjoint if, for any $i \neq j$, $E_i \cap E_j$ = empty set
  - Card from a pack: Spades, Hearts, Diamond, Clubs
- Large sample space: Partition into disjoint events for study

Events in Venn diagrams



- Union
  - All covered regions
- Intersection
  - Common region
- Disjoint events
  - No overlap in regions
  - A and C are disjoint, B and C are disjoint
  - A ∪ C: both regions, A ∩ C: empty
- A ∩ B^c
  - Region of A outside of B
  - Throw a die: even but not a multiple of 3

Useful skill: Translate “English” to events

- The hats of 5 persons are identical and get mixed up. Each person picks a hat at random.
  - Event A: No person gets their own hat
  - Event B: Every person gets their own hat
  - Event C: At least one person does not get their own hat
  - Event D: At least one person gets their own hat

  [Sample space for this experiment?]

- What is A^c? What is B^c?

- Are A and B disjoint?

- What is $A \cap B^c$?

IPL example

- One over with 6 deliveries; in each delivery, 0, 1, 2, 3, 4 or 6 runs may be scored
  - Event A: no 4s
  - Event B: no 6s
  - Event C: exactly 20 runs scored

- What is $A \cup B$? Over had no 4s or over had no 6s
- What is $A \cap B$? No 4s and no 6s in the over
- Can $A \cap B \cap C$ occur?

- What is the complement of $A \cup B$ and $A \cap B$? Use De Morgan's laws
  - $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$

Probability

Events and chance

- Given an event, we may have an idea of how likely it is to occur
  - Toss a coin: we expect (over several tosses) heads occurring about half the time
  - Distribute a pack of cards to 4 players: the chance of getting all 13 cards of one suit is very very low
  - Football world cup: chance of India winning?
  - Cricket world cup: chance of India winning?
- Probability theory
  - “Mathematical” theory to assign chances to events of an experiment
- What is a “mathematical” theory?
  - Define the basic objects of interest precisely
  - **Assume a few things - these are called axioms**
  - Deduce everything else with logical proof
- Successful theories have “natural” axioms that respect intuition and end up having a wide range of applications

Probability

DEFINITION (**Probability**) "Probability" is a function P that assigns to each event a real number between 0 and 1. The entire probability space (sample space, events and probability function) should satisfy two axioms, which will be specified soon.

- Value assigned by the probability function is supposed to represent the "chance" of the event occurring
  - Higher value means higher chance
  - 0 means event cannot occur and 1 means event always occurs
  - Often, probability is mentioned as percentage
  - Warning: This kind of "meaning" of an object is not important for the mathematical theory
- All ingredients of the theory have been defined now
  - Sample space, events, probability - these together are called "probability space"

Towards axioms for probability

- Probability function
  - Assigns a value that represents chance of occurence of the event
  - Axioms will place conditions on the probability function
- The axioms should be intuitive with respect to combination of events and relationships between events
  - For instance, if A is a subset of B, P(B) should be higher than P(A)?
  - Given P(A) and P(B), what can one say about $P(A \cup B)$?
- A few basic scenarios
  - P(empty set) = ?
  - P(S) = ?
  - If A and B are disjoint, are there constraints on P(A), P(B)?

Probability space axioms

DEFINITION (**Probability**) "Probability" is a function *P* that assigns to each event a real number between 0 and 1 and satisfies the following two axioms:

1. P(*S*) = 1 (probability of the entire sample space equals 1)

2. If E_1, E_2, E_3, ... are disjoint events (how many events? Could be infinitely many),

$$P(E_1 \cup E_2 \cup E_3 \cup \ldots) = P(E_1) + P(E_2) + P(E_3) + \ldots$$

- Axioms are intuitive and restrict the functions that can be probability functions
- They result in several intuitive deductions on probabilities of combination of events, probabilities of subsets of events etc.

Examples

- Toss a coin: S = { H, T }
  - Valid as per axioms: P(empty) = 0, P({H}) = 0.5, P({T}) = 0.5, P({H,T}) = 1
  - Valid as per axioms: P(empty) = 0, P({H}) = p, P({T}) = 1 - p, P({H,T}) = 1 ($0 \leq p \leq 1$)
  - Invalid: P(empty) = 0, P({H}) = 0.5, P({T}) = 0.6, P({H,T}) = 1 (why?)
- Throw a die: S = { 1, 2, 3, 4, 5, 6 }
  - There are 64 events. How to specify the probability function?
  - How to ensure that the axioms are satisfied?
  - How will valid probability functions look?
- We will answer the above questions for several types of sample spaces in the ensuing lectures
  - Only the axioms will be used as assumptions
  - Everything else can be deduced from them!

Basic properties

Recap

- Physical setting
  - Experiment - Process to be studied statistically
  - Outcome - result of experiment
- Probability space
  - Sample space *S* - set of outcomes
  - Events - subsets of sample space (technical condition)
  - Probability P - function from events to the closed interval [0,1]
  - Axioms
    - $P(S) = 1$
    - $E_1, E_2, \ldots$ disjoint: $P(E_1 \cup E_2 \cup \ldots) = P(E_1) + P(E_2) + \ldots$
- This lecture
  - Basic properties of probability space
  - These are critical for computing probabilities

Property 1: Empty set

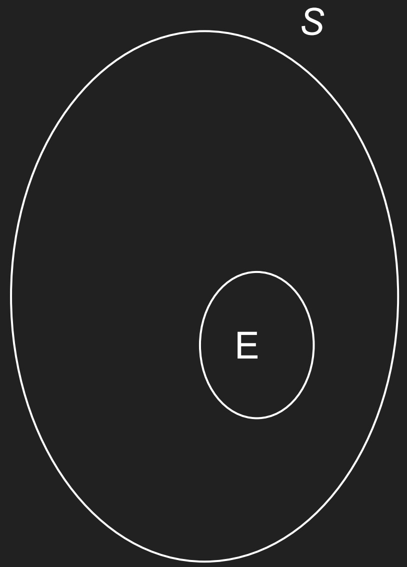
Probability of the empty set (denoted Φ) equals 0.
$P(\Phi) = 0$

- Proof
  - $\Phi^c = S$ and Φ, S are disjoint and $\Phi \cup S = S$
  - By Axiom 2, $P(\Phi \cup S) = P(\Phi) + P(S)$ or $P(S) = P(\Phi) + P(S)$ or $P(\Phi) = 0$

Property 2: Complement

Let E^c be the complement of Event E. Then,
$P(E^c) = 1 - P(E)$

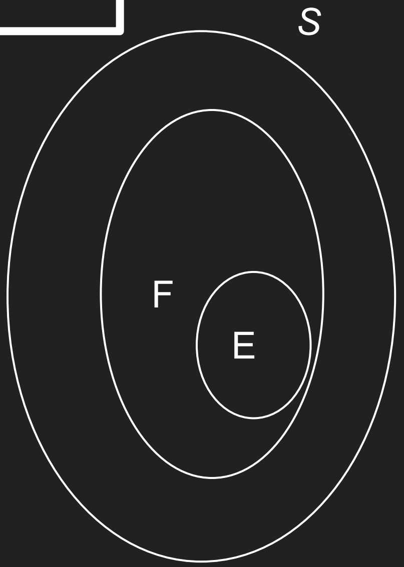
- Proof
  - E and E^c are disjoint and $E \cup E^c = S$
  - By Axiom 2, $P(E \cup E^c) = P(E) + P(E^c)$ or $P(S) = P(E) + P(E^c)$
  - By Axiom 1, $1 = P(S) = P(E) + P(E^c)$
  - So, $P(E^c) = 1 - P(E)$



Property 3: Subset

If Event E is a subset of Event F, i.e. $E \subseteq F$, then
$P(F) = P(E) + P(F \setminus E)$,
which implies that $P(E) \leq P(F)$

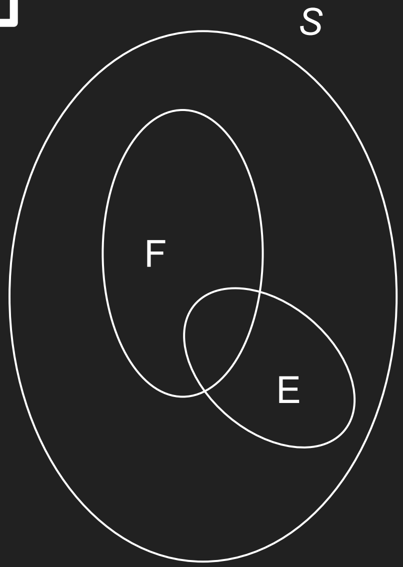
- Proof
  - $F \setminus E = F \cap E^c$ (outside of E and inside of F)
  - E and $F \setminus E$ are disjoint and $E \cup (F \setminus E) = F$
  - By Axiom 2, $P(E \cup (F \setminus E)) = P(E) + P(F \setminus E)$ or $P(F) = P(E) + P(F \setminus E)$



Property 3a: Difference and intersection

If E and F are events, then
$P(E) = P(E \cap F) + P(E \setminus F)$,
$P(F) = P(E \cap F) + P(F \setminus E)$

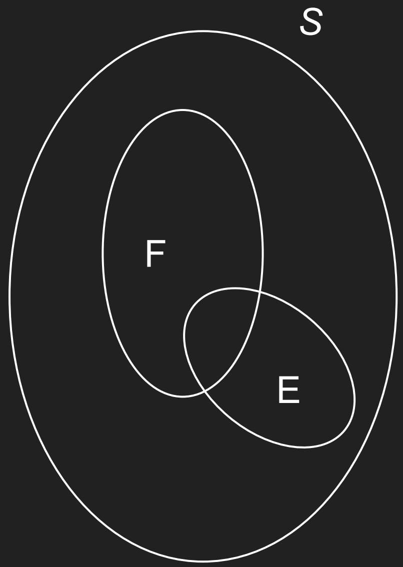
- Proof
    - $E \cap F$ is a subset of E
    - By subset property, $P(E) = P(E \cap F) + P(E \setminus (E \cap F))$
    - Now, $E \setminus (E \cap F) = E \setminus F$
    - So, $P(E) = P(E \cap F) + P(E \setminus F)$



Property 4: Union and intersection

If E and F are events, then
$P(E \cup F) = P(E) + P(F) - P(E \cap F)$

- Proof
  - $E \cup F = (E \setminus F) \cup (E \cap F) \cup (F \setminus E)$ and the 3 events on RHS are disjoint
  - $P(E \cup F) = P(E \setminus F) + P(E \cap F) + P(F \setminus E)$
  - Use $P(E \setminus F) = P(E) - P(E \cap F)$ and $P(F \setminus E) = P(F) - P(E \cap F)$



Working with probability spaces

Toss a coin

- Sample space: $S = \{ H, T \}$
- Events: Φ, $E = \{H\}$, $F = \{T\}$, $S = \{H, T\}$
- By Axiom 1, $P(S) = 1$
- By empty set property, $P(\Phi) = 0$
- By complement property, since $F = E^c$, $P(F) = 1 - P(E)$
- Probability function has to be the following
  - Let $p = P(E)$ (probability of heads is p) for some $0 \leq p \leq 1$
  - $P(\Phi) = 0$, $P(E) = p$, $P(F) = 1 - p$, $P(S) = 1$
- In an actual application, we need to choose p. How?
  - Suppose coin is fair. Then, probability of heads = probability of tails
  - So, $p = 1 - p$, which results in $p = 0.5$

Restaurant hiring

A waiter and a cashier are to be hired. There are 4 applicants - David and Megha from Delhi, and Rajesh and Veronica from Mumbai. The restaurant hires one person at random as waiter, and another from the remaining as cashier.

- Write out the sample space
- Event A: Cashier is from Delhi
- Event B: Exactly one position is filled by a Delhiite
- Event C: Neither position is filled by a Delhiite

Fishing town

- Town with a few fishing boats going out to catch fish everyday.
- Over the years, folks have observed the following:
  - Chance of catching more than 400 Kg of fish in a day is 35%
  - Chance of catching more than 500 Kg of fish in a day is 10%
- What is the chance of catching between 400 and 500 Kg of fish in a day?

Weather forecast

- Supposing you hear the following forecast for rain and temperature
  - Chance of rain tomorrow: 60%
  - Chance of max temperature above 30 deg: 70%
  - Chance of rain and max temperature above 30 deg: 40%
- What is the chance of no rain and max temperature below 30 deg?

Manipulating events to get more events

Suppose $S = \{a,b,c,d,e\}$ and let $E = \{a,b,e\}$ and $F = \{b,c\}$ be events. What other events can be obtained starting from E and F by various operations?

Distributions

Example: Throw a die

Sample space $S = \{ 1, 2, 3, 4, 5, 6 \}$

- Suppose that all individual outcomes are events
  - This implies that all subsets are events (take unions of the individual outcomes in the subset)
- Let
  $p_1 = P(\{1\})$, $p_2 = P(\{2\})$, $p_3 = P(\{3\})$, $p_4 = P(\{4\})$, $p_5 = P(\{5\})$, $p_6 = P(\{6\})$
  Each p_i is between 0 and 1
- $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$: disjoint events, $\{1\} \cup \{2\} \cup \{3\} \cup \{4\} \cup \{5\} \cup \{6\} = S$
- By Axiom 2, $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = P(S) = 1$
- Fair die
  All p_i are equal. So, $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$
  Called “equally likely outcomes”

Example: Throw a die (continued)

- Probability of any subset is easy to compute

  $P(\{1,3,5\}) = P(\{1\} \cup \{3\} \cup \{5\}) = P(\{1\}) + P(\{3\}) + P(\{5\}) = p_1 + p_3 + p_5$

  Similarly, one can compute P(E) for any event E

- Fair die: equally likely outcomes

  $p_i = 1/6$

  P(E) = (no of outcomes in E) / (no of outcomes in S)

  $P(\{1,3,5\}) = 3/6$, $P(\{2,4\}) = 2/6$ etc

Distributions

- Assign probabilities to individual outcomes
- When is this possible?
    - When the outcomes can be enumerated as first, second, third and so on (called countable sample space)
    - Finite sample space - yes, this is a special case of countable
- Example - Uniform distribution on a finite sample space
    - S = { finite number of outcomes }
    - Equally likely outcomes - assign the same probability to each outcome
    - P(one outcome) = 1 / $|S|$, i.e. probability of an outcome = 1 / (no of outcomes in S)
    - P(event) = (no of outcomes in event) / (no of outcomes in S)

Uniform distribution on a finite sample space

P(event) = (no of favourable outcomes) / (total no of outcomes)

Example: Marbles in an urn

5 red and 8 blue marbles in an urn. Pick a marble from the urn at random.

Example: Throw two dice

Throw two dice. What is the probability that the sum of the two numbers is 8?

Example: Lost keys

Key to an apartment has been lost and the person living there is given 50 possible keys by security personnel. One key after another is tried till the matching key is found. How will you describe the outcomes and the sample space? Is the uniform distribution a reasonable one?

Example: Mixed-up hats

The hats of 3 persons are identical and get mixed up. Each person picks a hat at random. What is the probability that none of the persons get their own hat?

Conditional Probability

Motivation

- Experiment contains a sequence of steps - one after another
  - Ex: Toss a coin three times, throw a die 2 times, IPL over (one delivery after another)
- Initial probability space - no step has been completed
- Suppose you observe the result after the first step. What happens to the probability space?
  - Example: Toss a coin three times
    - Initial sample space: {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
    - Let Event B: first toss results in tails. Suppose Event B occurs.
    - Can we get a smaller probability space to work on for the second and third tosses? Yes
      - Sample space becomes B: {THH, THT, TTH, TTT}
      - Events and probability function: can be redefined to account for occurrence of B

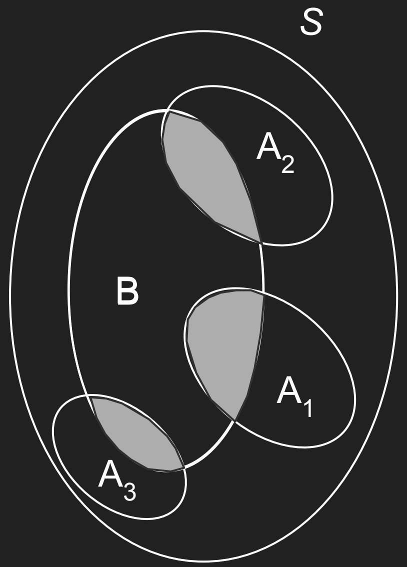
Initial probability space: splits into Event that occurred and "Conditional" probability space given that the event occurred

Conditional Probability Space

Consider a probability space: Sample space *S*, Collection of Events, Probability function P. Let B be an event with P(B) > 0.

DEFINITION (**Conditional Probability Space Given B**)
Sample space: B
Events: $A \cap B$ for every event A in original space
Probability function: $P(A \cap B) / P(B)$
(denoted $P(A \mid B)$ and called *conditional probability of* A given B)

For any event A in original space,
$P(A \cap B) = P(B)\, P(A \mid B)$



Example: Throw a die

$S = \{ 1,2,3,4,5,6 \}$, uniform distribution

$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = 1/6$

Event $E = \{ 2,4,6 \}$, $P(E) = 1/2$

Conditional probability space given E

$$P(\{2\} \mid E) = P(\{2\} \cap E) / P(E) = P(\{2\}) / P(E) = 1/3$$

$$P(\{4\} \mid E) = P(\{6\} \mid E) = 1/3$$

$$P(\{1\} \mid E) = P(\{1\} \cap E) / P(E) = P(\Phi) / P(E) = 0$$

$$P(\{3\} \mid E) = P(\{5\} \mid E) = 0$$

$$P(\{2,5\} \mid E) = 1/3$$

$$P(\{2,3,4\} \mid E) = 2/3$$

Example: Two urns with coloured marbles

Pick an urn at random and then pick a marble at random from the chosen urn.

P(red | urn 1) = 7/13, P(red | urn 2) = 5/13

P(blue | urn 1) = 6/13, P(blue | urn 2) = 8/13

Example: 3 students from a class

Consider a class with 15 students - 4 from State 1, 8 from State 2 and 3 from State 3. Three different students are chosen at random one after another. What is the probability that the selected three are from State 1, State 3 and State 1 again in that order?

Example: Family with two children

A family has two children. What is the probability that both are girls, given that at least one is a girl? Notice how the word “conditional” is dropped….