Recap

- Everything we do in probability theory is always in reference to an underlying probability space: $(\Omega, \mathcal{F}, P)$ where
  - Ω is the sample space
  - $\mathcal{F} \subset 2^{\Omega}$ set of events; each event is a subset of Ω
  - $P : \mathcal{F} \to [0, 1]$ is a probability (measure) that satisfies the three axioms:
    - A1 $P(A) \geq 0, \forall A \in \mathcal{F}$
    - A2 $P(\Omega) = 1$
    - A3 If $A_i \cap A_j = \phi, \forall i \neq j$ then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

Recap

- When $\Omega = \{\omega_1, \omega_2, \cdots\}$ (is countable), then probability is generally assigned by

$$P(\{\omega_i\}) = q_i, \;\; i = 1, 2, \cdots, \;\text{with } q_i \geq 0, \; \sum_i q_i = 1$$

- When Ω is finite with n elements, a special case is $q_i = \frac{1}{n}, \forall i$. (All outcomes equally likely)

Recap

- Conditional probability of A given (or conditioned on) B is

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

- This gives us the identity: $P(AB) = P(A|B)P(B)$
- This holds for multiple event, e.g., $P(ABC) = P(A|BC)P(B|C)P(C)$
- Given a partition, $\Omega = B_1 + B_2 + \cdots + B_m$, for any event, A,

$$P(A) = \sum_{i=1}^{m} P(A|B_i)P(B_i) \quad \text{(Total Probability rule)}$$

Recap

- Bayes Rule

$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)}$$

- Bayes rule can be viewed as transforming a prior probability into a posterior probability.

Recap: Independence

- Two events A, B are said to be independent if

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

- Suppose $P(A), P(B) > 0$. Then, if they are independent

$$P(A|B) = \frac{P(AB)}{P(B)} = P(A); \text{ similarly } P(B|A) = P(B)$$

- If A, B are independent then so are $A\&B^c$, $A^c\&B$ and $A^c\&B^c$.

Independence of multiple events

- Events $A_1, A_2, \cdots, A_n$ are said to be (totally) independent if for any k, $1 \leq k \leq n$, and any indices $i_1, \cdots, i_k$, we have

$$P(A_{i_1} \cdots A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

- For example, A, B, C are independent if

$$P(AB) = P(A)P(B); \; P(AC) = P(A)P(C);$$

$$P(BC) = P(B)P(C); \; P(ABC) = P(A)P(B)P(C)$$

Pair-wise independence

- Events $A_1, A_2, \cdots, A_n$ are said to be pair-wise independent if

$$P(A_i A_j) = P(A_i)P(A_j), \ \forall i \neq j$$

- Events may be pair-wise independent but not (totally) independent.
- Example: Four balls in a box inscribed with '1', '2', '3' and '123'. Let E_i be the event that number 'i' appears on a radomly drawn ball, $i = 1, 2, 3$.
- Easy to see: $P(E_i) = 0.5, \ i = 1, 2, 3$.
- $P(E_i E_j) = 0.25 \ (i \neq j) \Rightarrow$ pairwise independent
- But, $P(E_1 E_2 E_3) = 0.25 \neq (0.5)^3$

Conditional Independence

- Events A, B are said to be (conditionally) independent given C if
$$P(AB|C) = P(A|C)P(B|C)$$
- If the above holds
$$P(A|BC) = \frac{P(ABC)}{P(BC)} = \frac{P(AB|C)P(C)}{P(BC)}$$
$$= \frac{P(A|C)\ P(B|C)P(C)}{P(BC)} = P(A|C)$$
- Events may be conditionally independent but not independent. (e.g., 'independent' multiple tests for confirming a disease)
- It is also possible that A, B are independent but are not conditionally independent given some other event C.

Use of conditional independence in Bayes rule

- We can write Bayes rule with multiple conditioning events.

$$P(A|BC) = \frac{P(BC|A)P(A)}{P(BC|A)P(A) + P(BC|A^c)P(A^c)}$$

- The above gets simplified if we assume
$P(BC|A) = P(B|A)P(C|A)$,
$P(BC|A^c) = P(B|A^c)P(C|A^c)$
- Consider the old example, where now we repeat the test for the disease.
- Take: $A = D$, $B = T_+^1$, $C = T_+^2$.
- Assuming conditional independence we can calculate the new posterior probability using the same information we had about true positive and false positive rate.

- Let us consider the example with $P(T_+|D) = 0.99$, $P(T_+|D^c) = 0.05$. $P(D) = 0.1$.
- Recall that we got $P(D|T_+) = 0.69$.
- Let us suppose the same test is repeated.

$$P(D \mid T_+^1 T_+^2) = \frac{P(T_+^1 T_+^2 \mid D)P(D)}{P(T_+^1 T_+^2 \mid D)P(D) + P(T_+^1 T_+^2 \mid D^c)P(D^c)}$$

$$= \frac{P(T_+^1 \mid D)P(T_+^2 \mid D)P(D)}{P(T_+^1 \mid D)P(T_+^2 \mid D)P(D) + P(T_+^1 \mid D^c)P(T_+^2 \mid D^c)P(D^c)}$$

$$= \frac{0.99 * 0.99 * 0.1}{0.99 * 0.99 * 0.1 + 0.05 * 0.05 * 0.9} = 0.97$$

Sequential Continuity of Probability

- For function, $f : \Re \rightarrow \Re$, it is continuous at x if and only if $x_n \rightarrow x$ implies $f(x_n) \rightarrow f(x)$.
- Thus, for continuous functions,

$$f\left(\lim_{n\to\infty} x_n\right) = \lim_{n\to\infty} f(x_n)$$

- We want to ask whether the probability, which is a function whose domain is $\mathcal{F}$, is also continuous like this.
- That is, we want to ask the question

$$P\left(\lim_{n\to\infty} A_n\right) \stackrel{?}{=} \lim_{n\to\infty} P(A_n)$$

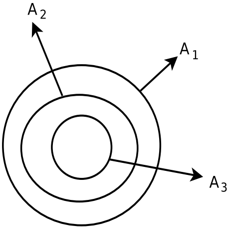
- For this, we need to first define limit of a sequence of sets.

- For now we define limits of only monotone sequences. (We will look at the general case later in the course)
- A sequence, $A_1, A_2, \cdots$, is said to be monotone decreasing if

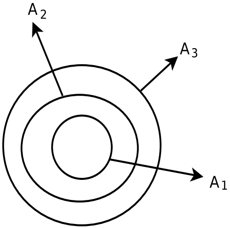
$$A_{n+1} \subset A_n, \ \forall n \quad \text{(denoted as } A_n \downarrow\text{)}$$

- A sequence, $A_1, A_2, \cdots$, is said to be monotone increasing if

$$A_n \subset A_{n+1}, \ \forall n \quad \text{(denoted as } A_n \uparrow\text{)}$$



$A_n \downarrow$



$A_n \uparrow$

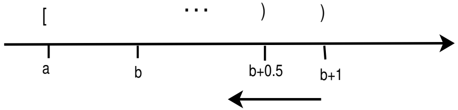
- Let $A_n \downarrow$. Then we define its limit as

$$\lim_{n\to\infty} A_n = \cap_{k=1}^{\infty} A_k$$

- This is reasonable because, when $A_n \downarrow$, we have $A_n \subset A_{n-1} \subset A_{n-2} \cdots$ and hence, $A_n = \cap_{k=1}^{n} A_k$.
- Similarly, when $A_n \uparrow$, we define the limit as

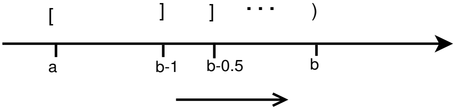
$$\lim_{n\to\infty} A_n = \cup_{k=1}^{\infty} A_k$$

- Let us look at simple examples of monotone sequences of subsets of $\Re$.
- Consider a sequence of intervals:
  $A_n = [a,\ b + \frac{1}{n}), n = 1, 2, \cdots$ with $a, b \in \Re,\ a < b$.



- We have $A_n \downarrow$ and $\lim A_n = \cap_i A_i = [a,\ b]$
- Why? – because
  - $b \in A_n, \forall n \Rightarrow b \in \cap_i A_i$, and
  - $\forall \epsilon > 0,\ b + \epsilon \notin A_n$ after some $n \Rightarrow b + \epsilon \notin \cap_i A_i$.
    For example, $b + 0.01 \notin A_{101} = [a,\ b + \frac{1}{101})$.

- We have shown that $\cap_n [a,\ b + \frac{1}{n}) = [a,\ b]$
- Similarly we can get $\cap_n (a - \frac{1}{n},\ b] = [a,\ b]$
- Now consider $A_n = [a,\ b - \frac{1}{n}]$.



- Now, $A_n \uparrow$ and $\lim A_n = \cup_n A_n = [a,\ b)$.
- Why? – because
  - $\forall \epsilon > 0$, $\exists n$ s.t. $b - \epsilon \in A_n \Rightarrow b - \epsilon \in \cup_n A_n$;
  - but $b \notin A_n, \forall n \Rightarrow b \notin \cup_n A_n$.
- These examples also show how using countable unions or intersections we can convert one end of an interval from 'open' to 'closed' or vice versa.

- To summarize, limits of monotone sequences of events are defined as follows

$$A_n \downarrow \quad \lim_{n\to\infty} A_n = \cap_{k=1}^{\infty} A_k$$

$$A_n \uparrow \quad \lim_{n\to\infty} A_n = \cup_{k=1}^{\infty} A_k$$

- Having defined the limits, we now ask the question

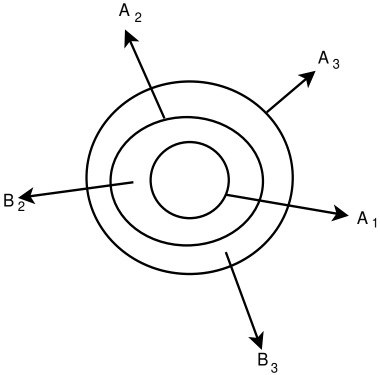
$$P\left(\lim_{n\to\infty} A_n\right) \stackrel{?}{=} \lim_{n\to\infty} P(A_n)$$

where we assume the sequence is monotone.

Theorem: Let $A_n \uparrow$. Then $P(\lim_n A_n) = \lim_n P(A_n)$

- Since $A_n \uparrow$, $A_n \subset A_{n+1}$.
- Define sets B_i, $i = 1, 2, \cdots$, by

$$B_1 = A_1, \;\; B_k = A_k - A_{k-1}, \;\; k = 2, 3, \cdots$$



Theorem: Let $A_n \uparrow$. Then $P(\lim_n A_n) = \lim_n P(A_n)$

- Since $A_n \uparrow$, $A_n \subset A_{n+1}$.
- Define sets B_i, $i = 1, 2, \cdots$, by

$$B_1 = A_1, \;\; B_k = A_k - A_{k-1}, \;\; k = 2, 3, \cdots$$

- Note that B_k are mutually exclusive. Also note that

$$A_n = \cup_{k=1}^n B_k \;\; \text{and hence } P(A_n) = \sum_{k=1}^n P(B_k)$$

- We also have

$$\cup_{k=1}^n A_k = \cup_{k=1}^n B_k, \forall n \;\; \text{and hence } \cup_{k=1}^\infty A_k = \cup_{k=1}^\infty B_k$$

- Thus we get

$$\begin{aligned} P(\lim_n A_n) &= P(\cup_{k=1}^\infty A_k) = P(\cup_{k=1}^\infty B_k) \\ &= \sum_{k=1}^\infty P(B_k) = \lim_n \sum_{k=1}^n P(B_k) = \lim_n P(A_n) \end{aligned}$$

- We showed that when $A_n \uparrow$, $P(\lim_n A_n) = \lim_n P(A_n)$
- We can show this for the case $A_n \downarrow$ also.
- Note that if $A_n \downarrow$, then $A_n^c \uparrow$. Using this and the theorem we can show it. (Left as an exercise)
- This property is known as monotone sequential continuity of the probability measure.

- We can think of a simple example to use this theorem.
- We keep tossing a fair coin. (We take tosses to be independent). We want to show that never getting a head has probability zero.
- The basic idea is simple. $((0.5)^n \rightarrow 0)$
- But to formalize this we need to specify what is our probability space and then specify what is the event (of never getting a head).
- If we toss the coin for any fixed N times then we know the sample space can be $\{0, 1\}^N$.
- But for our problem, we can not put any fixed limit on the number of tosses and hence our sample space should be for infinite tosses of a coin.

- We take Ω as set of all infinite sequences of 0's and 1's:

$$\Omega = \{(\omega_1, \omega_2, \cdots) \; : \; \omega_i \in \{0,1\}, \; \forall i\}$$

- This would be uncountably infinite.
- We would not specify $\mathcal{F}$ fully. But assume that any subset of Ω specifiable through outcomes of finitely many coin tosses would be an event.
- Thus "no head in the first n tosses" would be an event.

- What P should we consider for this uncountable Ω?
  We are not sure what to take.
- So, let us ask only for some consistency.
  For any subset of this Ω that is specified only through outcomes of first n tosses, that event should have the same probability as in the finite probability space corresponding to n tosses.
- Consider an event here;

$$A = \{(\omega_1, \omega_2, ...) \; : \; \omega_1 = \omega_2 = 0\} \subset \Omega$$

  A is the event of tails on first two tosses.
- We are saying we must have $P(A) = (0.5)^2$.
- Now we can complete problem

- For $n = 1, 2, \cdots$, define

$$A_n = \{(\omega_1, \omega_2, ...) \; : \; \omega_i = 0, \; i = 1, \cdots, n\}$$

- A_n is the event of no head in the first n tosses and we know $P(A_n) = (0.5)^n$.
- Note that $\cap_{k=1}^{\infty} A_k$ is the event we want.
- Note that $A_n \downarrow$ because $A_{n+1} \subset A_n$.
- Hence we get

$$P(\cap_{k=1}^{\infty} A_k) = P(\lim_n A_n) = \lim_n P(A_n) = \lim_n (0.5)^n = 0$$

- The Ω we considered can be corresponded with the interval $[0, 1]$.
- Each element of Ω is an infinite sequence of 0's and 1's

$$\omega = (\omega_1, \omega_2, \cdots), \; \omega_i \in \{0, 1\} \; \forall i$$

- We can put a 'binary point' in front and thus consider ω to be a real number between 0 and 1.
- That is, we correspond ω with the real number: $\omega_1 2^{-1} + \omega_2 2^{-2} + \cdots$
- For example, the sequence $(0, 1, 0, 1, 0, 0, 0, \cdots)$ would be the number: $2^{-2} + 2^{-4} = 5/16$.
- Essentially, every number in $[0, 1]$ can be represented by a binary sequence like this and every binary sequence corresponds to a real number between 0 and 1.
- Thus, our Ω can be thought of as interval $[0, 1]$.
- So, uncountable Ω arise naturally if we want to consider infinite repetitions of a random experiment

- The P we considered would be such that probability of an interval would be its length.
- Consider the example event we considered earlier

$$A = \{(\omega_1, \omega_2, ...) \; : \; \omega_1 = \omega_2 = 0\} \subset \Omega$$

- When we view the Ω as the interval $[0, 1]$, the above is the set of all binary numbers of the form $0.00xxxxxx \cdots$.
- What is this set of numbers?
- It ranges from $0.000000 \cdots$ to $0.0011111 \cdots$.
- That is the interval $[0, 0.25]$.
- As we already saw, the probability of this event is $(0.5)^2$ which is the length of this interval

- We looked at this probability space only for an example where we could use monotone sequential continuity of probability.
- But this probability space is important and has lot of interesting properties.

$$\Omega = \{(\omega_1, \omega_2, \cdots) \ : \ \omega_i \in \{0, 1\}, \forall i\}$$

- Here, $\frac{1}{n}\sum_{i=1}^{n} \omega_i$ the fraction of heads in the first n tosses.
- Since we are tossing a fair coin repeatedly, we should expect

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \omega_i = \frac{1}{2}$$

- We expect this to be true for 'almost all' sequences in Ω.
- That means 'almost all' numbers in $[0, 1]$ when expanded as infinite binary fractions, satisfy this property.
- This is called Borel's normal number theorem and is an interesting result about real numbers.

Probability Models

- As mentioned earlier, everything in probability theory is with reference to an underlying probability space: $(\Omega, \mathcal{F}, P)$.
- Probability theory starts with $(\Omega, \mathcal{F}, P)$
- We can say that different P correspond to different models.
- Theory does not tell you how to get the P.
- The modeller has to decide what P she wants.
- The theory allows one to derive consequences or properties of the model.

- Consider the random experiment of tossing a fair coin three times.
- We can take $\Omega = \{0, 1\}^3$ and can use the following P_1.

ω	$P_1(\{\omega\})$
0 0 0	1/8
0 0 1	1/8
0 1 0	1/8
0 1 1	1/8
1 0 0	1/8
1 0 1	1/8
1 1 0	1/8
1 1 1	1/8

- Now probability theory can derive many consequences:
  - The tosses are independent
  - Probability of 0 or 3 heads is 1/8 while that of 1 or 2 heads is 3/8

- Now consider a P_2 (different from P_1) on the same Ω

ω	$P_2(\{\omega\})$
0 0 0	$1/4$
0 0 1	$1/12$
0 1 0	$1/12$
0 1 1	$1/12$
1 0 0	$1/12$
1 0 1	$1/12$
1 1 0	$1/12$
1 1 1	$1/4$

- The consequences now change
  - The probability that number of heads is 0 or 1 or 2 or 3 are all same and all equal $1/4$.
  - The tosses are not independent

- We can not ask which is the 'correct' probability model here.
- Such a question is meaningless as far as probability theory is concerned.
- One chooses a model based on application.
- If we think tosses are independent then we choose P_1. But if we need to model some dependence among tosses, we choose a model like P_2.

- The model P_2 accommodates some dependence among tosses.
- Outcomes of previous tosses affect the current toss.

ω	$P_2(\{\omega\})$
0 0 0	$1/4$ $(= (1/2)(2/3)(3/4))$
0 0 1	$1/12$ $(= (1/2)(2/3)(1/4))$
0 1 0	$1/12$ $(= (1/2)(1/3)(2/4))$
0 1 1	$1/12$
1 0 0	$1/12$
1 0 1	$1/12$
1 1 0	$1/12$
1 1 1	$1/4$

- It is also a useful model.

We next consider the concept of random variables. These allow one to specify and analyze different probability models.

This entire course can be considered as studying different random variables.

Random Variable

- A random variable is a real-valued function on Ω:
  $X : \Omega \to \Re$
- For example, $\Omega = \{H, T\}$, $X(H) = 1, \ X(T) = 0$.
- Another example: $\Omega = \{H, T\}^3$, $X(\omega)$ is numbers of H's.
- A random variable maps each outcome to a real number.
- It essentially means we can treat all outcomes as real numbers.
- We can effectively work with $\Re$ as sample space in all probability models