Probability

Jonathan Gai

February 19, 2022

Contents

Lecture 1: Probability Space

20 Jan. 11:00

Example. If we have a die with outcomes 1, 2, . . . , 6.

1. $\mathbb{P}(2) = \frac{1}{6}$
2. $\mathbb{P}(\text{multiple of } 3) = \frac{2}{6} = \frac{1}{3}$
3. $\mathbb{P}(\text{pair or a multiple of } 3) = \frac{4}{6} = \frac{2}{3}$

1 Formal Setup

We try to define a probability space rigorously in this section.

Definition 1.1 (Probability Space). We have the following,

1. Sample space Ω, a set of outcomes.
2. $\mathcal{F}$, a collection of subsets of Ω (called events).
3. $\mathcal{F}$ is a σ-algebra if
   (a) **F1**: $\Omega \in \mathcal{F}$
   (b) **F2**: if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$
   (c) **F3**: For all countable collections $\{A_n\}$ in $\mathcal{F}$, $\cup_n A_n \in \mathcal{F}$.

Given σ-algebra $\mathcal{F}$ on Ω, function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure if

1. **P1**: The probability function is nonnegative.
2. **P2**: $\mathbb{P}(\Omega) = 1$
3. **P3**: For all countable collection $\{A_n\}$ of disjoint events in $\mathcal{F}$, we have $\mathbb{P}(\cup_n A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.

Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Problem. Why $\mathbb{P} : \mathcal{F} \to [0, 1]$, not $\mathbb{P} : \Omega \to [0, 1]$?

We will justify the definition in the following examples.

Example. When Ω is finite or countable,

1. In general: $\mathcal{F} = \mathcal{P}(\Omega)$.
2. $\mathbb{P}(2)$ is shorthand for $\mathbb{P}(\{2\})$.
3. $\mathbb{P}$ is determined by $\mathbb{P}(\{w\}), \forall w \in \Omega$.

Remark. When Ω is uncountable, a probability space behaves differently, as shown in the following example.

Example. If $\Omega = [0, 1]$, and we want to choose a real number, all equally likely.

If $\mathbb{P}\{0\} = \alpha > 0$, then $\mathbb{P}(\{0, 1, \frac{1}{2}, \ldots, \frac{1}{n}\} = n\alpha)$. This cannot happen if n large, because we would have $\mathbb{P} > 1$. So $\mathbb{P}(\{0\}) = 0$ or undefined.

Example. When Ω is infinitely countable (e.g., $\Omega = \mathbb{N}$ or $\Omega = \mathbb{Q} \cap [0, 1]$), however, it is not possible to choose uniformly. Suppose it is possible, there are two possibilities

- If $\mathbb{P}(\{\omega\}) = \alpha \quad \forall \omega \in \Omega$,

  then $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \infty$. $\lightning$
- If $\mathbb{P}(\{\omega\}) = 0 \quad \forall \omega \in \Omega$,

  then $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 0$. $\lightning$

So it is not possible to have one such uniform probability space. But that's fine as there exists many other interesting probability measures on a infinite countably set.

Property. From the axioms, we want to prove the following properties of a probability space.

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

   Proof. A, A^c disjoint. $A \cup A^c = \Omega$. So $\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(\Omega) = 1$ ■

2. $\mathbb{P}(\varnothing) = 0$
3. If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

1.1 Examples of Probability Spaces

Example. Here we list some concrete examples of probability spaces.

1. Ω finite, $\Omega = \{w_1, \ldots, w_n\}$, $\mathcal{F}$ = all subsets under uniform choice.

   $\mathbb{P} : \mathcal{F} \to [0,1], \mathbb{P}(A) = \frac{|A|}{|\Omega|}$. In particular: $\mathbb{P}(\{w\}) = \frac{1}{|\Omega|} \forall w \in \Omega$.

2. If we are choosing without replacement n indistinguishable marbles that are labelled $\{1, \ldots, n\}$. Pick $k \leq n$ marbles uniformly at random.

   Here we have $\Omega = \{A \subseteq \{1, \ldots, n\}$, $|A| = k$, $|\Omega| = \binom{n}{k}$.

3. If we have a well-shuffled deck of cards, and we uniformly chose permutation of 52 cards.

   $\Omega = \{$all permutations of 52 cards$\}$. $|\Omega| = 52!$.

   Then we have

   $$\mathbb{P}(\text{first three cards have the same suit}) = \frac{52 \cdot 12 \cdot 11 \cdot 49!}{52!} = \frac{22}{425}.$$

Lecture 2: Finite Probability Space

22 Jan. 11:00

Example (Coincidental Birthday). There we have n people, what is the probability that at least two share a birthday? To be precise, we first make the following assumptions,

- No leap years; (365 days in a year)
- All birthdays are equally likely.

We have the probability space

$$\begin{aligned}
\Omega &= \{1, \ldots, 365\}^n \\
\mathcal{F} &= \mathcal{P}(\Omega) \\
A &= \{\text{at least 2 people share birthday}\} \\
A^c &= \{\text{all } n \text{ birthdays are different}\}.
\end{aligned}$$

So we have the probability

$$\mathbb{P}(A^c) = \frac{365 \times 364 \times \ldots \times (365 - n - 1)}{365^n},$$
$$\mathbb{P}(A) = 1 - \frac{365 \times 364 \times \ldots \times (365 - n - 1)}{365^n}.$$

Remark.

- We note several special n values,

$$\begin{aligned} n = 22 &\quad : \quad \mathbb{P}(A) \approx 0.479 \\ n = 23 &\quad : \quad \mathbb{P}(A) \approx 0.507 \\ n \geq 366 &\quad : \quad \mathbb{P}(A) = 1 \end{aligned}$$

- The probability of birthday is not equal in real life though. It is more likely to be born about 9 months after christmas.
- Sometimes it would be easier to calculate the probability of the complement of an event.

1.2 Combinatorial Analysis

If Ω is a finite set such that $|\Omega| = n$,

Problem. How many ways to partition Ω into k disjoint subsets $\Omega_1, \ldots \Omega_k$ with $|\Omega_i| = n_i$ $(\sum\limits_{i=1}^{k} n_i = n)$?

The total number of ways M is

$$\begin{aligned} M &= \binom{n}{n_i}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-n_2\cdots-n_{k-1}}{n_k} \\ &= \binom{n}{n_i}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n_k}{n_k} \\ &= \frac{n!}{n!(n-n_1)!} \times \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \times \cdots \times \frac{(n-n_1-n_2-\cdots-n_{k-1})!}{x_k!0!} \\ &= \frac{n!}{n_1!n_2!\cdots n_k!} \\ &= \binom{n}{n_1, n_2, \ldots, n_k} \end{aligned}$$

which is called the *multinomial coefficient*, and denoted by the last term in the equations.

Remark. The ordering of the subsets do matter in this setting.

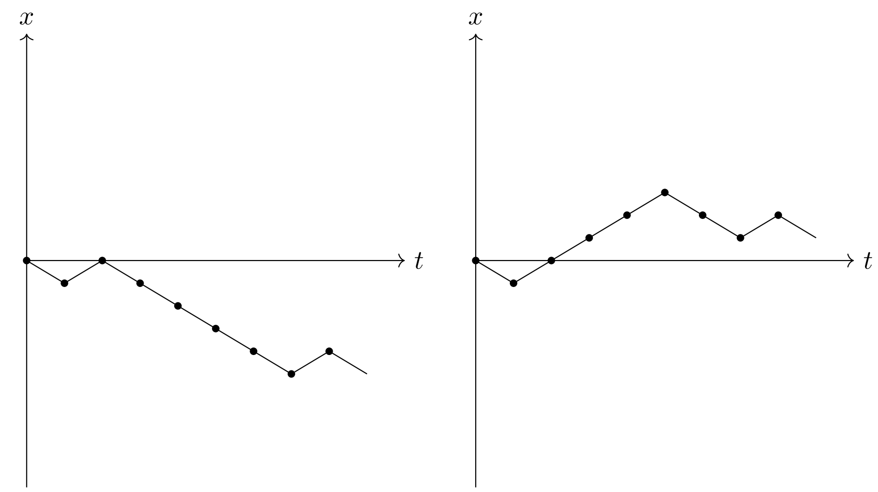

Figure 1: Random Walks

1.3 Random Walks

We have the following uniform probability space

$$\Omega = \{(x_0, x_1, \dots, x_n) \mid x_0 = 0, |x_k - x_{k-1}| = 1, k = 1, \dots, n\},$$
$$|\Omega| = 2^n.$$

Problem. What's $\mathbb{P}(x_n = 0)$ and $\mathbb{P}(x_n = n)$?

We have $\mathbb{P}(x_n = n) = \frac{1}{2^n}$.

When n is odd, $\mathbb{P}(x_n = 0) = 0$ because after every step the value changes parity. To find the probability when n is even, we need to choose $\frac{n}{2}$ ks for which $x_k = x_{k-1} + 1$, and the rest $x_k = x_{k-1} - 1$. So

$$\begin{aligned}\mathbb{P}(x_n = 0) &= 2^{-n}\binom{n}{n/2} \\ &= \frac{n!}{2^n[(\frac{n}{2})!]^2}.\end{aligned}$$

Problem. What happens when n is large?

We next present Stirling's Formula, and we adopt the following notation for the time being.

Notation. If (a_n), b_n are two sequences, we say $a_n \sim b_n$ as $n \to \infty$ if $\frac{a_n}{b_n} \to 1$ as $n \to \infty$.

Theorem 1.1 (Stirling's Formula).

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} \quad \text{as} \quad n \to \infty.$$

We also have the weaker version

$$\log(n!) \sim n \log n.$$

Lecture 3

25 Jan. 11:00

Proof. We have

$$\log(n!) = \log 2 + \log 3 + \ldots + \log n.$$

So

$$\int_1^n \log x dx \leq \log(n!) \leq int_1^{n+1} \log x dx$$

$$\underbrace{n \log n - n + 1}_{n \log n} \leq \log(n!) \leq \underbrace{(n+1)\log(n+1) - n}_{n \log n}.$$

$\log(n!)$ is sandwiched between the lower and upper integrals, so $\log(n!)$ must be approximately $n \log n$ as well. In this calculation, these facts helped

1. $\log x$ is increasing, so it's easier to bounded by the integrals.
2. $\log x$ has a nice integral. So the integrals have closed forms.

■

(Ordered) Compositions

Definition 1.2. A *composition* of m with k parts is sequence $(m1, \ldots, m_k)$ of non-negative integers with $\sum_{i=1}^{k} m_i = m$.

We use stars and bars. There are m stars and $k-1$ bars, and

$$\#\text{Compositions} = \binom{m+k-1}{m}.$$

1.4 Properties of Probability Measures

Recall Definition (1.1). We prove the following properties.

Property.

1. Countable sub-additivity

   Let $(A_n)_{n\geq 1}$ sequence of events in $\mathcal{F}$. Then

$$\mathbb{P}(\cup_{n\geq 1} A_n) \leq \sum_{n\geq 1} \mathbb{P}(A_n).$$

Proof. We rewrite $\cup_{n\geq 1}$ as a disjoint union.

Define $B_1 = A_1$ and $B_n = A_n \setminus (A_1 \cup \ldots \cup A_{n-1})$.

So

- $\cup_{n\geq 1} B_n = \cup_{n\geq 1} A_n$,
- $(B_n)_{n\geq 1}$ disjoint (by construction),
- $B_n \subseteq A_n \implies \mathbb{P}(B_n) \leq \mathbb{P}(A_n)$.

And we have

$$\mathbb{P}(\cup_{n\geq 1} A_n) = \mathbb{P}(\cup_{n\geq 1} B_n) = \sum_{n\geq 1} \mathbb{P}(B_n) = \sum_{n\geq 1} \mathbb{P}(A_n).$$

■

2. Continuity $(A_n)_{n\geq 1}$ increasing sequence of events in $\mathcal{F}$ that is $A_n \subseteq A_{n+1}$ for all n.

   In fact, $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\cup_{n\geq 1} A_n)$.

   Proof. We reuse the B_ns, and we have

   - $\sqcup_{k=1}^{n} B_k = A_n$,
   - $\cup_{n\geq 1} B_n = \cup_{n\geq 1} A_n$.

   So we have

   $$\mathbb{P}(A_n) = \sum_{k=1}^{n} \mathbb{P}(B_k) \to \sum_{k\geq 1} \mathbb{P}(B_k) = \mathbb{P}(\cup_{n\geq 1} B_n) = \mathbb{P}(\cup_{n\geq 1} A_n).$$

   ■

3. Inclusion-Exclusion Principle

   Background: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

   Similarly, for $A, B, C \in \mathcal{F}$,

   $$\begin{aligned}\mathbb{P}(A \cup B \cup C) =& \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(B \cap C) \\ &- \mathbb{P}(C \cap A) + \mathbb{P}(A \cap B \cap C).\end{aligned}$$

   The full Inclusion-Exclusion principle statement is the following. Let $A_1, \ldots, A_n \in \mathcal{F}$, then

   $$\begin{aligned}\mathbb{P}(\cup_{i=1}^{n} A_i) &= \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1\leq i_1 < i_2 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \ldots \\ &\quad + (-1)^{n+1} \mathbb{P}(A_1 \cap \ldots \cap A_n) \\ &= \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ I \neq \varnothing}} (-1)^{|I|+1} \mathbb{P}(\cap_{i\in I} A_i).\end{aligned}$$

Lecture 3: Inclusion-Exclusion Principle

27 Jan. 2022

Proof. We used induction. The $n = 2$ case is proved in the example sheet.

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \mathbb{P}\left((\bigcup_{i=1}^{n-1} A_i)\bigcup A_n\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^{n-1} A_i\right) + \mathbb{P}(A_n) - \mathbb{P}\left((\bigcup_{i=1}^{n-1} A_i)\bigcap A_n\right).\end{aligned}$$

Note that for $J \subseteq \{1, \ldots, n-1\}$,

$$\bigcap_{i\in J}(A_i \cap A_n) = \bigcap_{i\in J\cup\{n\}} A_i.$$

The inductive statement tells us

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{\substack{J\subseteq\{1,\ldots,n-1\}\\ J\neq\varnothing}} (-1)^{|J|+1}\mathbb{P}(\bigcap_{i\in J} A_i) + \mathbb{P}(A_n) \\ &\quad - \sum_{\substack{J\subseteq\{1,\ldots,n-1\}\\ J\neq\varnothing}} (-1)^{|J|+1}\mathbb{P}\left(\bigcap_{i\in J\cup\{n\}} A_i\right) \\ &= \sum_{\substack{I\subseteq\{1,\ldots,n-1\}\\ I\neq\varnothing}} (-1)^{|I|+1}\mathbb{P}(\bigcap_{i\in I} A_i) + \mathbb{P}(A_n) \\ &\quad + \sum_{\substack{I\subseteq\{1,\ldots,n-1\}\\ n\in I, |I|\geq 2}} (-1)^{|I|+1}\mathbb{P}\left(\bigcap_{i\in I} A_i\right) \\ &= \sum_{\substack{I\subseteq\{1,\ldots,n\}\\ I\neq\varnothing}} (-1)^{|I|+1}\mathbb{P}\left(\bigcap_{i\in I} A_i\right).\end{aligned}$$

■

1.5 Bonferroni Inequalities

Problem. What if you truncate Inclusion-Exclusion Principle?

Recall countable subadditivity states that $\mathbb{P}(\cup A_i) \leq \sum \mathbb{P}(A_i)$, also known as union bound. We have the following inequalities.

- $\mathbb{P}(\cup_{i=1}^{n} A_i) \leq \sum_{k=1}^{r} (-1)^{k+1} \sum_{i_1<\ldots<i_k} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k})$ when r is odd;
- $\mathbb{P}(\cup_{i=1}^{n} A_i) \geq \sum_{k=1}^{r} (-1)^{k+1} \sum_{i_1<\ldots<i_k} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k})$ when r is even.

Problem. When is it good to truncate at, for example, $r = 2$?

Proof. We induct on r and n. When r is odd

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) =& \mathbb{P}\left(\bigcup_{i=1} A_i\right) + \mathbb{P}(A_n) - \mathbb{P}\left(\bigcup_{i=1}^{n-1}(A_i \cap A_n)\right) \\
\leq& \sum_{\substack{J\subseteq\{1,\dots,n-1\} \\ 1\leq |J| \leq r}} (-1)^{|J|+1}\mathbb{P}\left(\bigcap_{i\in J} A_i\right) + \mathbb{P}(A_n) \\
&- \sum_{\substack{J\subseteq\{1,\dots,n-1\} \\ 1\leq |J| \leq r-1}} (-1)^{|J|+1}\mathbb{P}\left(\bigcap_{i\in J\cup\{n\}} A_i\right) \\
\leq& \sum_{\substack{I\subseteq\{1,\dots,n\} \\ 1\leq |I| \leq r}} (-1)^{|I|+1}\mathbb{P}\left(\bigcap_{i\in I} A_i\right).
\end{aligned}$$

And a similar argument follows when r is even. ■

1.6 Counting with IEP

Inclusion Exclusion Principle gives up a route to solve questions that do not have a closed form answer.

When we have a uniform probability measure on Ω with $|\Omega| < \infty$,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} \ \forall A \subseteq \Omega.$$

Then $\forall A_1, \dots, A_n \subseteq \Omega$,

$$|A_1 \cup \ldots \cup A_n| = \sum_{k=1}^{n}(-1)^{n+1} \sum_{i_1 < \ldots < i_k} |A_{i_1} \cap \ldots \cap A_{i_k}|,$$

and similarly for Bonferroni inequalities.

Example. We count the number of surjections $f : \{1, \dots, n\} \to \{1, \dots, m\}$ with $n \geq m$.

We have the probability space and event

$$\begin{aligned}
\Omega &= \{f : \{1, \dots, n\} \to \{1, \dots, m\}\}, \\
A &= \{f : \operatorname{Im}(f) = \{1, \dots, m\}\}.
\end{aligned}$$

For all $i \in \{1, \dots, m\}$, let $B_i = \{f \in \Omega \mid i \notin \operatorname{Im}(f)\}$. We have the following key observations:

- $A = B_1^c \cap \ldots B_m^c = (B_1 \cup \ldots \cup B_m)^c$.
- $|B_{i_1} \cap \ldots \cap B_{i_k}|$ is nice to calculate, and we have

$$|B_{i_1} \cap \ldots \cap B_{i_k}| = |\{f \in \Omega \mid i_1, \dots, i_k \notin \operatorname{Im}(f)\}| = (m-k)^n.$$

So by IEP, we have

$$\begin{aligned}|B_1 \cup \ldots \cup B_m| &= \sum_{k=1}^{m} (-1)^{k+1} \sum_{i_1 < \ldots < i_k} |B_{i_1} \cap \ldots \cap B_{i_k}| \\ &= \sum_{k=1}^{m} (-1)^{k+1} \binom{m}{k} (m-k)^n.\end{aligned}$$

So $|A| = m^n - \sum_{k=1}^{m} (-1)^{k+1} \binom{m}{k} (m-k)^n = \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n$.

Lecture 5: Independence

29 Jan. 2022

Example (Derangements). We try to find the number of permutations with no fixed points, for a Secret Santa for example. We have the sample space and event

$$\begin{aligned}\Omega &= \{\text{permutations of } \{1, \ldots, n\}\}, \\ D &= \{\sigma \in \Omega \mid \sigma(i) \neq i \ \forall i = 1, \ldots, n\}.\end{aligned}$$

For all $i \in 1, \ldots, n$, let $A_i = \{\sigma \in \Omega \mid \sigma(i) = i\}$.

Problem. Is $\mathbb{P}(D)$ large or small when $n \to \infty$.

Similar to the last example, $D = A_1^c \cap \ldots \cap A_n^c = (\cup_{i=1}^n A_i)^c$, and

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \frac{(n-k)!}{n!}.$$

So by IEP, we have

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1 < \ldots < i_k} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) \\ &= \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!}.\end{aligned}$$

So $\mathbb{P}(D) = 1 - \mathbb{P}(\cup_{i=1}^n A_i) = 1 - \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$.

In fact, when $n \to \infty$, $\mathbb{P}(D) \to \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = e^{-1} \approx 0.37$.

Note. What if instead $\Omega' = \{\text{all functions } f : \{1, \ldots, n\} \to \{1, \ldots, n\}\}$?

We have $D = \{f \in \Omega' \mid f(i) \neq i \ \forall i = 1, \ldots, n\}$, and

$$\mathbb{P}(D) = \frac{(n-1)^n}{n^n} = (1 - \frac{1}{n})^n \to e^{-1}.$$

Can we just say $\mathbb{P}(D) = (\frac{n-1}{n})^n$? We would need independence to say that.

Also note that $f(i)$ is a random quantity associated to Ω. We will study these later as a random variable.

We are allowed to toss a fair coin n times, but we can't toss an unfair coin n times so far.

1.7 Independence

Definition 1.3. Events $A, B \in \mathcal{F}$ are *independent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \text{ (denoted as } A \perp\!\!\!\perp B)$$

A countable collection of events (A_n) is *independent* if for all distinct $i_1, \ldots, i_k$, we have

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j}).$$

Remark. *Pairwise independence* does not imply independence.

Example. If we have the uniform probability space

$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\},$$

and $\mathbb{P}(\{\omega\}) = \frac{1}{4}$ for all $\omega \in \Omega$. And we define the following events

$$\begin{aligned} A &= \text{first coin } H = \{(H,H),(H,T)\} \\ B &= \text{second coin } H = \{(H,H),(T,H)\} \\ C &= \text{same outcome} = \{(H,H),(T,T)\} \end{aligned}$$

Note that probability of each of these happening is $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$, and $A \cap B = A \cap C = B \cap C = \{(H,H)\}$, so they are pairwise independent. But

$$\mathbb{P}(A \cap B \cap C) = \frac{1}{4} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

The three events are not independent.

Example.

- If we have $\Omega' = \{\text{all functions } f : \{1, \ldots, n\} \to \{1, \ldots, n\}\}$, and let $A_i = \{f \in \Omega' \mid f(i) = i\}$. Then,

$$\begin{aligned} \mathbb{P}(A_i) &= \frac{n^(n-1)}{n^n} = \frac{1}{n} \\ \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) &= \frac{n^{n-k}}{n^n} = \frac{1}{n^k} = \prod_{j=1}^{k} \mathbb{P}(A_{i_j}). \end{aligned}$$

  Here, (A_i) are independent events.

- If we have $\Omega = \{\sigma \mid \text{permutation of } \{1, \ldots, n\}\ \}$, and let $A_i = \{\sigma \in \Omega \mid \sigma(i) = i\}$. Then,

$$\begin{aligned} \mathbb{P}(A_i) &= \frac{n^(n-1)}{n^n} = \frac{1}{n} \\ \mathbb{P}(A_i \cap A_j) &= \frac{(n-1)!}{n!} = \frac{1}{n(n-1)} \neq \mathbb{P}(A_i)\mathbb{P}(A_j). \end{aligned}$$

  Here, (A_i) are not independent.

Property.

1. If A is independent of B then A is also independent of B^c.

   Proof.
   $$\begin{aligned}\mathbb{P}(A\cap B^c) &= \mathbb{P}(A)-\mathbb{P}(A\cap B)\\ &= \mathbb{P}(A)-\mathbb{P}(A)\mathbb{P}(B)\\ &= \mathbb{P}(A)(1-\mathbb{P}(B))\\ &= \mathbb{P}(A)\mathbb{P}(B^c).\end{aligned}$$
   ∎

2. A is independent of $B=\Omega$ and of $C=\varnothing$.

   Proof. $\mathbb{P}(A\cap\Omega)=\mathbb{P}(A)=\mathbb{P}(A)\mathbb{P}(\Omega)$, and $A \perp\!\!\!\perp \varnothing$ by part 1. ∎

3. $\mathbb{P}(B)=0$ or 1 Then A is independent of B.

1.8 Conditional Probability

Definition 1.4 (Conditional Probability). If we have a probability space $(\Omega,\mathcal{F},\mathbb{P})$ as before. Consider $B\in\mathcal{F}$ with $\mathbb{P}(B)>0$, and we have $\mathbb{P}(A)$, The *conditional probability of A given B* is

$$\mathbb{P}(A\mid B)\coloneqq\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}.$$

We can interpret this informally as the probability of A if we know B happened.

Example. If A,B are independent events,

$$\mathbb{P}(A\mid B)=\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}=\frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)}=\mathbb{P}(A).$$

Informally, we know that if A,B are independent, then knowing where B happened doesn't affect probability of A.

Lecture 6

1 Feb. 2022

Property.

1. $\mathbb{P}(A\mid B)\geq 0$.
2. $\mathbb{P}(B\mid B)=\mathbb{P}(\Omega\mid B)=1$.
3. (A_n) disjoint events in $\mathcal{F}$, we claim

$$\mathbb{P}(\cup_{n\geq 1}A_n\mid B)=\sum_{n\geq 1}\mathbb{P}(A_n\mid B).$$

Proof. $$\begin{aligned}\mathbb{P}(\cup_{n\geq 1}A_n \mid B) &= \frac{\mathbb{P}((\cup_n A_n)\cap B)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(\cup_n (A_n \cap B))}{\mathbb{P}(B)} \quad \text{numerator is a disjoint union} \\ &= \frac{\sum_n \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n\geq 1}\mathbb{P}(A_n \mid B).\end{aligned}$$
To prove it, we used the definition, and applied **P1**, **P2**, **P3** to numerator.
■

4. $\mathbb{P}(\cdot \mid B)$ is a function from $\mathcal{F} \to [0,1]$ that satisfies the rules to be a probability measure in Ω. It is often useful to restrict the function to

$$\begin{aligned}\Omega' &= B \\ \mathcal{F}' &= \mathcal{P}(B),\end{aligned}$$

especially in finite/ countable setting. Then $(\Omega', \mathcal{F}', \mathbb{P}(\cdot \mid B))$ also satisfies rules to be a probability measure on Ω'.

We have

$$\begin{aligned}\mathbb{P}(A\cap B) =& \mathbb{P}(A)\mathbb{P}(B \mid A) \\ \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) =& \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2) \\ & \cdots \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1})\end{aligned}$$

Example. Uniform permutation $(\sigma(1), \sigma(2), \ldots, \sigma(n)) \in \Sigma_n$. We claim that

$$\begin{aligned}&\mathbb{P}(\sigma(k) = i_k \mid \sigma(1) = i, \ldots, \sigma(k-1) = i_{k-1}) \\ =&\begin{cases} 0, & \text{if } i_k \in \{i, \ldots, i_{k-1}\} \\ \dfrac{1}{n-k+q}, & \text{if otherwise}\end{cases}\end{aligned}$$

Proof. We have

$$\begin{aligned}&\mathbb{P}(\sigma(k) = i_k \mid \sigma(1) = i_1, \ldots, \sigma(k-1) = i_{k-1}) \\ =&\frac{\mathbb{P}(\sigma(1) = i_1, \ldots, \sigma(k) = i_k)}{\mathbb{P}(\sigma(1) = i_1, \ldots, \sigma(k-1) = i_{k-1})} \\ =&\frac{\frac{(n-k)!}{n!}}{\frac{(n-k+1)!}{n!}} = \frac{1}{n-k+1}.\end{aligned}$$

■

1.9 Law of Total Probability & Bayes' Formula

Definition 1.5. $(B_1, B_2, \ldots) \subseteq \Omega$ is a *partition* of Ω if $\Omega = \cup_n B_n$ and (B_n) are disjoint.

Theorem 1.2. (B_n) a finite or countable partition of Ω with $B_n \in \mathcal{F}$ for all n such that $\mathbb{P}(B_n) > 0$. Then for all $A \in \mathcal{F}$:

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A \mid B_n)\mathbb{P}(B_n).$$

This is also called "Partition Theorem".

Proof. Note that $\cup_n (A \cap B_n) = A$. So we have

$$\mathbb{P}(A) = \sum_{n \geq 1} \mathbb{P}(A \cap B_n) = \sum_n \mathbb{P}(A \mid B_N)\mathbb{P}(B_n).$$

■

Theorem 1.3 (Bayes' Formula). With the same setup as above, we have

$$\mathbb{P}(B_n \mid A) = \frac{\mathbb{P}(A \cap B_N)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_n)\mathbb{P}(B_n)}{\sum\limits_m \mathbb{P}(A \mid B_m)\mathbb{P}(B_m)}.$$

Rephrasing for $n = 2$, we have $\mathbb{P}(B \mid A)\underbrace{\mathbb{P}(A)}_{given} = \underbrace{\mathbb{P}(A \mid B)\mathbb{P}(B)}_{given} = \mathbb{P}(A \cap B)$.

Example. Lecture course has $\frac{2}{3}$ of the lectures on weekdays and $\frac{1}{3}$ on weekends. We have

$$\mathbb{P}(\text{forget notes} \mid \text{weekday}) = \frac{1}{8}$$
$$\mathbb{P}(\text{forget notes} \mid \text{weekend}) = \frac{1}{2}$$

What is $\mathbb{P}(\text{weekend} \mid \text{forget notes})$?

We have $B_1 = \{\text{weekday}\}$ and $B_2 = \{\text{weekend}\}$ and $A = \{\text{forget notes}\}$. So we have

$$\mathbb{P}(A) = \frac{2}{3} \cdot \frac{1}{8} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{12} + \frac{1}{6} = \frac{1}{4}.$$

And by Bayes' Formula, we have

$$\mathbb{P}(B_2 \mid A) = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{1}{4}} = \frac{2}{3}.$$

Example (Disease testing). If p are infected and $1 - p$ are not, and we have

$$\mathbb{P}(\text{positive} \mid \text{infected}) = 1 - \alpha$$
$$\mathbb{P}(\text{positive} \mid \text{not infected}) = \beta.$$

Ideally, you want both α, β to be small. Of course, we want p to be small as well. We want to find $\mathbb{P}(\text{infected} \mid \text{positive})$. By LTP, we have

$$\mathbb{P}(\text{positive}) = p(1 - \alpha) + (1 - p)\beta.$$

Using Bayes', we have

$$\mathbb{P}(\text{infected} \mid \text{positive}) = \frac{p(1-\alpha)}{p(1-\alpha)+(1-p)\beta}.$$

Suppose $p \ll \beta$, we have $p(1-\alpha) \ll (1-p)\beta$. The probability is approximately $\frac{p(1-\alpha)}{(1-p)\beta} \sim \frac{p}{\beta}$ which is small.

Example (Simpson's Paradox). If the scientists want to know if jelly beans make your tongue change color? Studies give results:

Oxford	Change	No change	% change
Blue	15	22	41 %
Green	5	8	38 %

Cambridge	Change	No change	% change
Blue	10	3	77 %
Green	23	14	62 %,

but if you add them up, you get

Total	Change	No change	% change
Blue	25	25	50 %
Green	28	22	56 %.

Lecture 7

3 Feb. 2022

We continue from the Simpson's Paradox example. Let $A = \{\text{change color}\}$, $B = \{\text{blue}\}$, $B^c = \{\text{green}\}$, $C = \{\text{Cambridge}\}$ and $C^c = \{\text{Oxford}\}$. We have

$$\begin{aligned}\mathbb{P}(A \mid B \cap C) &> \mathbb{P}(A \mid B^c \cap C)\\ \mathbb{P}(A \mid B \cap C^c) &> \mathbb{P}(A \mid B^c \cap C^c).\end{aligned}$$

But it is not true that $\mathbb{P}(A \mid B) > \mathbb{P}(A \mid B^c)$. LTP for conditional probabilities is the following. Suppose $C_1, C_2, \ldots$ is a partition of B, and we have

$$\begin{aligned}\mathbb{P}(A \mid B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A \cap (\cup_n C_n))}{\mathbb{P}(B)}\\ &= \frac{\mathbb{P}(\cup_n (A \cap C_n))}{\mathbb{P}(B)} = \frac{\sum\limits_n \mathbb{P}(A \cap C_n)}{\mathbb{P}(B)}\\ &= \frac{\sum_n \mathbb{P}(A \mid C_n)\mathbb{P}(C_n)}{\mathbb{P}(B)} = \sum_n \mathbb{P}(A \mid C_n)\frac{\mathbb{P}(B \cap C_n)}{\mathbb{P}(B)}\end{aligned}$$

So in conclusion, we have

$$\mathbb{P}(A \mid B) = \sum_n \mathbb{P}(A \mid C_n)\mathbb{P}(C_n \mid B).$$

Special Case:

- If all $\mathbb{P}(C_n)$ are equal, then $\mathbb{P}(C_n \mid B)$ are all equal.
- If $\mathbb{P}(A \mid C_n)$ are all equal. Note that $\sum_n \mathbb{P}(C_n \mid B) = 1$. Then we have

$$\mathbb{P}(A \mid B) = \mathbb{P}(A \mid C_n).$$

Example. Uniform permutation $(\sigma(1), \sigma(2), \dots, \sigma(52)) \in \Sigma_{52}$ ("well-shuffled cards"). We call $\{1,2,3,4\}$ the aces. We consider $A = \{\sigma(1), \sigma(2) \text{ aces}\}$, and $B = \{\sigma(1) \text{ ace}\} = \{\sigma(1) \leq 4\}$, $C_i = \{\sigma(1) = i\}$.

Note $\mathbb{P}(A \mid C_i) = \mathbb{P}(\sigma(2) \in \{1,2,3,4\} \mid \sigma(1) = i) = \frac{3}{51}$ for $i \leq 4$ by previous example. And we have $\mathbb{P}(C_i) = \frac{1}{52}$. So we have $\mathbb{P}(A \mid B) = \frac{3}{51}$. In total, we have

$$\mathbb{P}(A) = \mathbb{P}(B) \times \mathbb{P}(A \mid B) = \frac{4}{52} \times \frac{3}{51}.$$

2 Discrete Random Variables

Motivation: Roll two dices. $\Omega = \{1, \dots, 6\}^2 = \{(i,j) \mid 1 \leq i,j \leq 6\}$. If we restrict attention to first dice $\{(i,j) \mid i = 3\}$; sum of dices $\{(i,j) \mid i + j = 8\}$; max of dice $\{(i,j) \mid i,j \leq 4, i \text{ or } j = 4\}$.

Goal: "Random real-valued measurements".

Definition 2.1. A *discrete random variable* X (often denoted by RV) on a probability space $(\Omega, \mathcal{F}, \mathbb{P}())$ is a function $X : \Omega \to \mathbb{R}$ such that

1. $\{\omega \in \Omega \mid X(\omega) = x\} \in \mathcal{F}$.
2. $\mathrm{Im}(X)$ is finite or countable (subset of $\mathbb{R}$).

We can write $\{\omega \in \Omega \mid X(\omega) = x\}$ as $\{X = x\}$. So $\mathbb{P}(X = x)$ is valid. And the image is often $\mathbb{Z}$ or $\{0,1\}$ for example, instead of $\{\text{Heads}, \text{Tails}\}$.

If Ω is finite or countable, and $\mathcal{F} = \mathcal{P}(\Omega)$, both requirements hold automatically.

Example (Part II Applied Probability). If we consider the arrival problem, we have $\Omega = \{\text{countable subsets } (a_1, a_2, \dots) \text{ of } (0, \infty)\}$. Then,

$$\begin{aligned} N_t &= \text{number of arrivals by time t} \\ &= |\{a_i \mid a_i \leq t\}| \in \{0, 1, 2, \dots\} \end{aligned}$$

is a discrete RV for each time t.

Definition 2.2. The *probability mass function* (p.m.f.) of discrete RV X is the function $p_X : \mathbb{R} \to [0,1]$ given by

$$p_X(x) = \mathbb{P}(X = x) \quad \forall x \in \mathbb{R}.$$

Note.

- If $x \notin \mathrm{Im}(X)$ (that is, $X(\omega)$ never takes value x), then

$$p_X(x) = \mathbb{P}(\omega \in \Omega \mid X(\omega) = x) = \mathbb{P}(\varnothing) = 0.$$

- $$\begin{aligned} \sum_{x \in \mathrm{Im}(X)} p_X(x) &= \sum_{x \in \mathrm{Im}(x)} \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x\}) \\ &= \mathbb{P}\big(\cup_{x \in \mathrm{Im}(X)} \{\omega \in \Omega \mid X(\omega) = x\}\big) = \mathbb{P}(\Omega) = 1 \end{aligned}$$

Example (Indicator Function). Event $A \in \mathcal{F}$, define $\mathbf{1}_A : \omega \to \mathbb{R}$ by

$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \notin A \end{cases}$$

called the *indicated function* of A. $\mathbf{1}_A$ is a discrete RV with $\mathrm{Im}(\mathbf{1}) = \{0, 1\}$. The probability mass function is

$$p_{\mathbf{1}_A}(1) = \mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(A)$$
$$p_{\mathbf{1}_A}(0) = \mathbb{P}(\mathbf{1}_A = 0) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$
$$p_{\mathbf{1}_A}(x) = 0 \quad \forall x \notin \{0, 1\}.$$

It encodes "did A happen" as a real number.

Remark. Given a probability mass function, we can always construct a probability space $(\Omega, \mathcal{F}, \mathbb{P}())$ and a RV defined on it with this pmf.