IC105: Probability and Statistics **2021-22-M**

Lecture 2: Probability Measure

Instructor: Dr. Kuldeep Kumar Kataria *Scribe:*

The algebra of set theory is applicable in probability theory. Probability is a measure of uncertainty. We are interested in quantifying uncertainty associated with various outcomes of a random experiment by assigning probability to these outcomes.

Here, we will not discuss how probabilities are assigned (which is a part of probability modelling) rather we will discuss properties of a probability as a measure.

Recall that $\mathcal{E}$ denotes a random experiment, Ω denotes the sample space of $\mathcal{E}$ and $\mathcal{F}$ denotes event space. For all practical purposes one may take $\mathcal{F} = \mathcal{P}(\Omega)$.

A set function is a function whose domain is a collection of sets (called a class of sets).

Definition 2.1 (Probability Function or Probability Measure)**.** *A probability function (or probability measure) is a real valued set function, defined on the event space $\mathcal{F}$ satisfying the following axioms:*

(a) $P(\Omega) = 1$ (certainty),

(b) $P(A) \geq 0 \ \ \forall A \in \mathcal{F}$ (positivity),

(c) If $A_1,\ A_2 \in \mathcal{F}$ be mutually exclusive/disjoint sets (i.e. $A_1 \cap A_2 = \phi$, the empty set) then

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

More generally, if $\{A_n\}_{n\geq 1}$ is a sequence of mutually exclusive (disjoint) sets in $\mathcal{F}$ i.e., $A_i \cap A_j = \phi,\ i \neq j$, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \ \ (countable \ \ additivity).$$

We call $P(A)$ the probability of event A. The triplet $(\Omega, \mathcal{F}, P)$ is called probability space.

Remark 2.2. *Axiom (b) and (c) are desirable for any measure (such as area, volume, probability etc.). Since the sample space Ω consists of all possible outcomes its occurrence is certain (100% chance of occurrence) and therefore Axiom (a) ($P(\Omega) = 1$) is also reasonable.*

Elementary Properties of Probability Function/ Measure:

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

(P1) $P(\phi) = 0$.

Proof. Let $A_1 = \Omega$ and $A_i = \phi$, $i = 2, 3, \ldots$ Also, we have $A_1 = \bigcup\limits_{i=1}^{\infty} A_i,\ A_i \cap A_j = \phi, \forall\, i \neq j$. Therefore,

$$\begin{aligned}
P(\Omega) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \\
\implies 1 &= \sum_{i=1}^{\infty} P(A_i), \quad \text{(Axioms (a) and (c))} \\
\implies 1 &= \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) \\
\implies 1 &= \lim_{n\to\infty} \left[P(\Omega) + (n-1)P(\phi)\right] \\
\implies 1 &= 1 + \lim_{n\to\infty} \left[(n-1)P(\phi)\right] \\
\implies P(\phi) &= 0.
\end{aligned}$$

This completes the proof. $\square$

(P2) For some natural number n, let $A_1, A_2, \ldots, A_n \in \mathcal{F}$ be mutually exclusive. Then, $P\left(\bigcup\limits_{i=1}^{n} A_i\right) = \sum\limits_{i=1}^{n} P(A_i)$.

Proof. Let $A_i = \phi,\ i = n+1, n+2, \ldots$ Then $A_i \cap A_j = \phi, \forall\, i \neq j$ and $\bigcup\limits_{i=1}^{n} A_i = \bigcup\limits_{i=1}^{\infty} A_i$. This implies

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \\
&= \sum_{i=1}^{\infty} P(A_i), \quad \text{(Axioms (c))} \\
&= \sum_{i=1}^{n} P(A_i), \quad (P(A_i) = P(\phi) = 0, \forall\, i = n+1, n+2, \ldots).
\end{aligned}$$

This completes the proof. $\square$

(P3) For all $A \in \mathcal{F},\ 0 \leq P(A) \leq 1$ and $P(A^c) = 1 - P(A)$.

Proof. Note that $\Omega = A \cup A^c$ and $A \cap A^c = \phi$. Therefore,

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c) \geq P(A), \quad \text{(using Axioms (a), (b) and } \textbf{(P2)}\text{)}.$$

Thus, $0 \leq P(A) \leq 1$ and $P(A^c) = 1 - P(A)$. $\square$

(P4) Let $A_1, A_2 \in \mathcal{F}$ be such that $A_1 \subseteq A_2$. Then, $P(A_2 - A_1) = P(A_2) - P(A_1)$ and $P(A_1) \leq P(A_2)$.

Proof. $A_2 = A_1 \cup (A_2 - A_1)$ and $A_1 \cap (A_2 - A_1) = \phi$. Thus,

$$P(A_2) = P(A_1) + P(A_2 - A_1) \implies P(A_2 - A_1) = P(A_2) - P(A_1).$$

By Axiom (b), we have $P(A_2 - A_1) \geq 0 \implies P(A_2) \geq P(A_1)$, that is, $P(\cdot)$ is monotone. $\square$

(P5) Let $A_1, A_2 \in \mathcal{F}$. Then,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \text{ (Inclusion-Exclusion principle for two events).}$$

Proof. Note that $A_1 \cup A_2 = A_1 \cup (A_2 - A_1)$ and $A_1 \cap (A_2 - A_1) = \phi$. This implies

$$P(A_1 \cup A_2) = P(A_1 \cup (A_2 - A_1)) = P(A_1) + P(A_2 - A_1) \quad \text{(using (\textbf{P2}))}. \tag{2.1}$$

Also, we have

$$(A_1 \cap A_2) \cap (A_2 - A_1) = \phi \text{ and } A_2 = (A_1 \cap A_2) \cup (A_2 - A_1),$$

which implies

$$\begin{aligned} P(A_2) &= P(A_1 \cap A_2) + P(A_2 - A_1) \\ \implies P(A_2 - A_1) &= P(A_2) - P(A_1 \cap A_2). \end{aligned} \tag{2.2}$$

Using (2.2) in (2.1), we get

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

This completes the proof. □

Remark 2.3. *(a) If $P(A) = 0$ and $B \subseteq A$, then $P(B) = 0$ (using (**P4**) and Axiom (b)).*

*Similarly, if $P(C) = 1$ and $C \subseteq D$, then $P(D) = 1$ (using (**P3**) and (**P4**)).*

*(b) **Exercise**: If $P(D) = 1$, then $P(A) = P(A \cap D)$, $\forall A \in \mathcal{F}$.*

Similarly, if $P(D) = 0$, then $P(A) = P(A \cap D^c)$, $\forall A \in \mathcal{F}$.

*(c) Let A_1, $A_2 \in \mathcal{F}$. Then, using (**P5**) and Axiom (b), we get*

$$P(A_1 \cup A_2) \leq P(A_1) + P(A_2) \text{ (\textit{Boole's inequality for two events})}.$$

*(d) Let A_1, $A_2 \in \mathcal{F}$. Then, using (**P3**), (**P5**) and Axiom (b), we get*

$$P(A_1 \cap A_2) \geq \max\{P(A_1) + P(A_2) - 1), 0\} \text{ (\textit{Bonferroni's inequality for two events})}.$$

Theorem 2.4 (Inclusion-Exclusion Principle)**.** *For $A_1, A_2, \ldots, A_k \in \mathcal{F}$, ($k \geq 2$ is an integer), let*

$$p_{1,k} = P(A_1) + P(A_2) + \cdots + P(A_k) = \sum_{i=1}^{k} P(A_i)$$

$$\begin{aligned} p_{2,k} &= P(A_1 \cap A_2) + P(A_1 \cap A_3) + \cdots + P(A_1 \cap A_k) + P(A_2 \cap A_3) + \cdots + P(A_2 \cap A_k) + \cdots + P(A_{k-1} \cap A_k) \\ &= \sum_{1 \leq i < j \leq k}^{\infty} \sum^{\infty} P(A_i \cap A_j) \end{aligned}$$

($\textit{sum of probabilities of all possible intersections involving 2 events out of the } k \textit{ events } A_1, \ldots, A_k$)

$\vdots$

$$p_{i,k} = \sum_{1 \leq j_1 < j_2 < \cdots < j_i \leq k}^{\infty} \sum^{\infty} \cdots \sum^{\infty} P(A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_i})$$

($\textit{sum of probabilities of all possible intersections involving } i \textit{ events out of } k \textit{ events } A_1, \ldots, A_k, i = 1, \ldots, k$).

Then,

$$P\left(\bigcup_{i=1}^{k} A_i\right) = p_{1,k} - p_{2,k} + p_{3,k} - p_{4,k} + \cdots + (-1)^{k-1} p_{k,k}.$$

Proof. Note that, for $k = 2$, $p_{1,2} = P(A_1) + P(A_2)$, $p_{2,2} = P(A_1 \cap A_2)$ and

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = p_{1,2} - p_{2,2}.$$

Thus the result is true for $k = 2$. Now suppose that the result is true for $k = 2, 3, \ldots, m$, that is,

$$P\left(\bigcup_{i=1}^{k} A_i\right) = p_{1,k} - p_{2,k} + p_{3,k} - p_{4,k} + \cdots + (-1)^{k-1} p_{k,k}, \;\; \forall\, k = 2, 3, \ldots, m.$$

Then,

$$\begin{aligned}
P\left(\bigcup_{i=1}^{m+1} A_i\right) &= P\left(\left(\bigcup_{i=1}^{m} A_i\right) \bigcup A_{m+1}\right) \\
&= P\left(\bigcup_{i=1}^{m} A_i\right) + P(A_{m+1}) - P\left(\left(\bigcup_{i=1}^{m} A_i\right) \bigcap A_{m+1}\right), \;\; \text{(using result for } k = 2) \\
&= \sum_{j=1}^{m} (-1)^{j-1} p_{j,m} + P(A_{m+1}) - P\left(\bigcup_{i=1}^{m} (A_i \cap A_{m+1})\right), \;\; \text{(using the result for } k = m \text{ on } \bigcup_{i=1}^{m} A_i) \\
&= \sum_{j=1}^{m} (-1)^{j-1} p_{j,m} + P(A_{m+1}) - \sum_{j=1}^{m} (-1)^{j-1} t_{j,m}, \;\; \text{(using the result for } k = m \text{ on } \bigcup_{i=1}^{m} (A_i \cap A_{m+1})),
\end{aligned}$$

where

$$\begin{aligned}
t_{1,m} &= \sum_{i=1}^{m} P(A_i \cap A_{m+1}) \\
t_{2,m} &= \sum\sum_{1 \le i < j \le m} P(A_i \cap A_j \cap A_{m+1}) \\
t_{j,k} &= \sum\sum\cdots\sum_{1 \le i_1 < i_2 < \cdots < i_j \le m} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_j} \cap A_{m+1}), \;\; j = 1, 2, \ldots, m \\
t_{m,m} &= P(A_1 \cap A_2 \cap \cdots \cap A_m \cap A_{m+1})
\end{aligned}$$

Therefore,

$$\begin{aligned}
P\left(\bigcup_{i=1}^{m+1} A_i\right) &= (p_{1,m} + P(A_{m+1})) - (p_{2,m} + t_{1,m}) + (p_{3,m} + t_{2,m}) + \cdots + (-1)^{m-1}(p_{m,m} + t_{m-1,m}) + (-1)^m t_{m,m} \\
&= p_{1,m+1} - p_{2,m+1} + p_{3,m+1} + \cdots + (-1)^{m-1} p_{m,m+1} + (-1)^m p_{m+1,m+1},
\end{aligned}$$

as

$$\begin{aligned}
p_{1,m} + P(A_{m+1}) &= \sum_{j=1}^{m} P(A_j) + P(A_{m+1}) = p_{1,m+1}, \\
p_{2,m} + t_{1,m} &= \sum\sum_{1 \le i < j \le m} P(A_i \cap A_j) + \sum_{i=1}^{m} P(A_i \cap A_{m+1}) \\
&= \sum\sum_{1 \le i < j \le m+1} P(A_i \cap A_j) = p_{2,m+1}, \\
&\;\;\vdots \\
p_{m,m} + t_{m-1,m} &= P(A_1 \cap A_2 \cap \cdots \cap A_m) + \sum\sum\cdots\sum_{1 \le i_1 < i_2 < \cdots < i_{m-1} \le m} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_{m-1}} \cap A_{m+1}) \\
&= p_{m,m+1}
\end{aligned}$$

and $t_{m,m} = P(A_1 \cap A_2 \cap \cdots \cap A_m \cap A_{m+1}) = p_{m+1,m+1}$. The result now follows by induction. $\square$

Remark 2.5. *Let $A_1, A_2, A_3 \in \mathcal{F}$. Then*

$$\begin{aligned}
P(A_1 \cup A_2 \cup A_3) &= p_{1,3} - p_{2,3} + p_{3,3} \\
&= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3).
\end{aligned}$$

Theorem 2.6. *For some positive integer $k \geq 2$, let $A_1, A_2, \ldots, A_k \in \mathcal{F}$. Then*

$$p_{1,k} - p_{2,k} \leq P\left(\bigcup_{i=1}^{k} A_i\right) \leq p_{1,k}.$$

Proof. Note that for $k = 2$, $p_{1,2} = P(A_1) + P(A_2)$, $p_{2,2} = P(A_1 \cap A_2)$ and

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2).$$

This implies $p_{1,2} - p_{2,2} = P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$. Thus the result is true for $k = 2$. Now suppose that for some positive integer $m (\geq 2)$

$$p_{1,k} - p_{2,k} \leq P\left(\bigcup_{i=1}^{k} A_i\right) \leq p_{1,k}, \ \ \forall\, k = 1, 2, \ldots, m.$$

Then,

$$\begin{aligned}
P\left(\bigcup_{i=1}^{m+1} A_i\right) &= P\left(\left(\bigcup_{i=1}^{m} A_i\right) \bigcup A_{m+1}\right) \\
&\leq P\left(\bigcup_{i=1}^{m} A_i\right) + P(A_{m+1}), \quad \Big(\text{using result for } k = 2, A = \cup_{i=1}^{m} A_i \text{ and } B = A_{m+1}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{then } P(A \cup B) \leq P(A) + P(B)\Big) \\
&\leq p_{1,m} + P(A_{m+1}) \\
&= p_{1,m+1}. \qquad (2.3)
\end{aligned}$$

Also using the result for $k = m$, we get

$$P\left(\bigcup_{i=1}^{m} A_i\right) \geq p_{1,m} - p_{2,m}$$

and

$$P\left(\bigcup_{i=1}^{m} (A_i \cap A_{m+1})\right) \leq \sum_{i=1}^{m} P(A_i \cap A_{m+1})$$

Thus,

$$\begin{aligned}
P\left(\bigcup_{i=1}^{m+1} A_i\right) &= P\left(\left(\bigcup_{i=1}^{m} A_i\right) \bigcup A_{m+1}\right) \\
&= P\left(\bigcup_{i=1}^{m} A_i\right) + P(A_{m+1}) - P\left(\bigcup_{i=1}^{m} (A_i \cap A_{m+1})\right) \\
&\geq p_{1,m} - p_{2,m} + P(A_{m+1}) - \sum_{i=1}^{m} P(A_i \cap A_{m+1}) \\
&= (p_{1,m} + P(A_{m+1})) - \left(p_{2,m} + \sum_{i=1}^{m} P(A_i \cap A_{m+1})\right) \qquad (2.4)
\end{aligned}$$

Using (2.3) and (2.4), we get

$$p_{1,m+1} - p_{2,m+1} \leq P\left(\bigcup_{i=1}^{m+1} A_i\right) \leq p_{1,m+1}$$

and the result follows using principle of mathematical induction. $\square$

Remark 2.7. *It can also be shown that*

$$\begin{aligned} p_{1,k} - p_{2,k} + p_{3,k} - p_{4,k} \leq & P\left(\bigcup_{i=1}^{k} A_i\right) \leq p_{1,k} - p_{2,k} + p_{3,k} \\ & \vdots \\ p_{1,k} - p_{2,k} + \cdots + p_{2m-1,k} - p_{2m,k} \leq & P\left(\bigcup_{i=1}^{k} A_i\right) \leq p_{1,k} - p_{2,k} + \cdots + p_{2m-1,k}, \end{aligned}$$

for $m = 1, 2, \ldots, [\frac{k}{2}]$.

Theorem 2.8 (Bonferroni's Inequality)**.** *Let* $A_1, A_2, \ldots, A_k \in \mathcal{F}$*. Then*

$$P\left(\bigcap_{i=1}^{k} A_i\right) \geq \max\left\{\sum_{i=1}^{k} P(A_i) - (k-1), 0\right\}.$$

Proof. We have

$$\begin{aligned} P\left(\bigcap_{i=1}^{k} A_i\right) &= P\left(\left(\bigcup_{i=1}^{k} A_i^c\right)^c\right), \text{ (De-Morgan's law)} \\ &= 1 - P\left(\bigcup_{i=1}^{k} A_i^c\right) \\ &\geq 1 - \sum_{i=1}^{k} P(A_i^c), \quad \text{(Boole's inequality)} \\ &= 1 - \sum_{i=1}^{k} (1 - P(A_i)) \\ &= \sum_{i=1}^{k} P(A_i) - (k-1). \end{aligned} \tag{2.5}$$

Also,

$$P\left(\bigcap_{i=1}^{k} A_i\right) \geq 0. \tag{2.6}$$

Combining (2.5) and (2.6), we get

$$P\left(\bigcap_{i=1}^{k} A_i\right) \geq \max\left\{\sum_{i=1}^{k} P(A_i) - (k-1), 0\right\}.$$

This completes the proof. $\square$

Example 2.9. *Random experiment: casting a red and white die.*

Sample space: $\Omega = \{(i, j) : i \in \{1, 2, \ldots, 6\},\ j \in \{1, 2, \ldots, 6\}\}$.

For $(i, j) \in \Omega$, i*: number of spots up on the red die;* j*: number of spots up on the white die.*

Event space $\mathcal{F} \equiv$ *power set of* Ω. *For* $A \in \mathcal{F}$, *define* $Q : \mathcal{F} \to \mathbb{R}$ *as*

$$Q(A) = \frac{|A|}{36}, \text{ where } |A| = \text{number of elements in } A.$$

Then

(a) $Q(\Omega) = \frac{|\Omega|}{36} = \frac{36}{36} = 1$.

(b) $Q(A) = \frac{|A|}{36} \geq 0, \ \forall\, A \in \mathcal{F}$.

(c) For mutually exclusive events $A_1, A_2, \ldots$

$$Q\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{|\bigcup_{i=1}^{\infty} A_i|}{36} = \frac{\sum_{i=1}^{\infty} |A_i|}{36} = \sum_{i=1}^{\infty} \frac{|A_i|}{36} = \sum_{i=1}^{\infty} Q(A_i).$$

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