Carathéodory condition and its implication: Inclusion-Exclusion Identity

stma

March 16, 2024

Given a (pre-) measure space $(\mathfrak{X}, \mathscr{A}, \mu)$ (i.e. think of $\mathscr{A}$ a *ring of sets* when we see μ a pre-measure), consider the following condition about $A \subset \mathfrak{X}$.

$$\forall B \in \mathscr{A},\ B = (B \cap A) \cup (B \cap A^c). \tag{1}$$

The condition (1) gives the sufficient condition for A to satisfy the Carathéodory condition (i.e. equivalently A being μ-measurable), namely,

$$\forall B \in \mathscr{A},\ \mu(B) = \mu(B \cap A) + \mu(B \cap A^c). \tag{2}$$

The following implication is shown to be true if $A, B \subset \mathfrak{X}$ both are μ-measurable.

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B). \tag{3}$$

This identity (3) is, as seen in many literatures written in relatively mathematical point of view, amongst all such measure theoretic introduction, often to be paid less attention despite its practical importance arise in such fields of statistics and applied probability theory.

This fundamental identity provides with a key ingredient to formulate and/or solve some real problems, where the probability of either union or intersection of events are difficult to determine.

In what follows, we'll prove a generalized form of the equation (3) known as *inclusion-exclusion identity*, in a relatively elementary approach.

Theorem 1. Let $(\mathfrak{X}, \mathscr{A}, \mu)$ be a measure space. For each family of sets $\{A_j\}_{j \le n} \subset \mathscr{A}$, the following identity holds:

$$\mu(\cup_j^n A_j) = \sum_{k=1}^{n} \sum_{\sigma \in sh(k,n-k)} (-1)^{k-1} \mu(A_{\sigma(1)} \cap \ldots \cap A_{\sigma(k)}). \tag{4}$$

The element $\sigma \in sh(k, n-k)$ iterates through $(k, n-k)$-shuffles so that each σ is thought of as a monotone function from $\{1 < \ldots < k\}$ with $1 \le \sigma(1) < \sigma(2) < \ldots < \sigma(k) \le n$.

For convenience, we express (4) by the linear combination of

$$S_k = \sum_{\sigma \in sh(k,n-k)} \mu(A_{\sigma(1)} \cap \ldots \cap A_{\sigma(k)}).$$

Proof. Let E_k be the set of all sample points that are contained in exactly k of the events $A_1, \ldots, A_n$, and for each $\sigma_k \in sh(k, n-k)$, we define:

$$A^{\sigma_k} = \bigcup_{j \notin \sigma_k} A_j,$$
$$A_{\sigma_k} = \bigcap_{j=1}^{k} A_{\sigma_k(j)},$$

where $j \notin \sigma_k$ stands for $j \notin \sigma_k(\{1, \ldots, k\})$.

By definition, it is immediate to see

$$\bigcup_j^n A_j = \bigsqcup_{k=1}^{n} E_k. \tag{5}$$

For each $\sigma_k \neq \sigma'_k$, $(A_{\sigma_k} - A^{\sigma_k}) \cap (A_{\sigma'_k} - A^{\sigma'_k}) = \varnothing$ since if $A_{\sigma_k} \cap A_{\sigma'_k} \neq \varnothing$, there is some $l \in \{1, \ldots, k\}$ such that $A_{\sigma_k(l)} \subset A^{\sigma'_k}$, then we have

$$E_k = \bigsqcup_{\sigma_k \in sh(k,n-k)} (A_{\sigma_k} - A^{\sigma_k}).$$

Now the following calculation is justified.

$$\begin{aligned}
\mu(\cup_{k=1}^n A_k) &= \sum_{k=1}^n \mu(E_k) \\
&= \sum_{k=1}^n \sum_{\sigma_k \in sh(k,n-k)} \mu(A_{\sigma_k} - A^{\sigma_k}) \qquad (6) \\
&= \sum_{k=1}^n \big(S_k - \sum_{\sigma_k \in sh(k,n-k)} \mu(A_{\sigma_k} \cap A^{\sigma_k})\big).
\end{aligned}$$

Assume (4) holds till $(n-1)$. By induction, $\mu(A_{\sigma_k} \cap A^{\sigma_k})$ has a form

$$\begin{aligned}
\mu(A_{\sigma_k} \cap A^{\sigma_k}) = &\sum_{j \notin \sigma_k} \mu(A_{\sigma_k} \cap A_j) \\
&- \sum_{j_1, j_2 \notin \sigma_k; j_1 < j_2} \mu(A_{\sigma_k} \cap A_{j_1} \cap A_{j_2}) \\
&+ \ldots \\
&+ (-1)^{n-k} \mu(A_{\sigma_k} \cap A_{j_1} \cap \ldots \cap A_{j_{n-k}}).
\end{aligned}$$

Taking a close look at the all terms of the form $\mu(A_{\sigma_k} \cap A_{j_1} \cap \ldots \cap A_{j_t})$ for each $1 \leq t \leq n-k$, it is shown that these terms amount to $(-1)^t \binom{k+t}{k}$ of S_{k+t}'s in $\sum_{\sigma_k} \mu(A_{\sigma_k} \cap A^{\sigma_k})$.

Together with (6), we have

$$\mu(E_k) = \sum_{j=0}^{n-k} (-1)^j \binom{k+j}{k} S_{k+j}. \qquad (7)$$

Summing (7) over $1 \leq k \leq n$, we can find c_k such that $\sum_{k=1}^n \mu(E_k) = \sum_{k=1}^n c_k S_k$. These c_k is calculated as

$$\begin{aligned}
c_k &= \sum_{j=0}^{k-1} (-1)^j \binom{k}{k-j} \\
&= \sum_{j=0}^{k-1} (-1)^j \binom{k}{j} \\
&= (-1)^{k+1} + \sum_{j=0}^{k} (-1)^j \binom{k}{j} \\
&= (-1)^{k+1},
\end{aligned}$$

the claim follows. □