Beyond Equally Likely Outcomes. Last time we showed that the axioms of probability are satisfied with respect to a finite set $S$, such that for $A \subset S$, $P(A) = \frac{|A|}{|S|}$. See also Example 2.5a. It can be also shown that the axioms hold for general finite spaces and general countably infinite spaces.

**Theorem 1** (General Finite Spaces). The axioms of probability are satisfied with respect to a set $S = \{s_1, \ldots, s_n\}$, such that for $i = 1, \ldots, n$, $$P(\{s_i\}) = p_i$$ where each $p_i \in [0, 1]$ and $\sum_{i=1}^{n} p_i = 1$. For more general events $A \subset S$, $$P(A) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} p_i$$

**Theorem 2** (General Countably Infinite Spaces). The axioms of probability are satisfied with respect to a set $S = \{s_1, s_2, \ldots\}$, such that for $i = 1, 2, \ldots$, $$P(\{s_i\}) = p_i$$ where each $p_i \in [0, 1]$ and $\sum_{i=1}^{\infty} p_i = 1$. For more general events $A \subset S$, $$P(A) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} p_i$$
Inclusion-Exclusion Identity. Previously we showed that:

\[(*) \quad P(E \cup F) = P(E) + P(F) - P(E \cap F).\]

Similarly,

\[
P(E \cup F \cup G) = P([E \cup F] \cup G)
= P(E \cup F) + P(G) - P([E \cup F] \cap G)
= P(E) + P(F) - P(E \cap F) + P(G)
- P([E \cap G] \cup [F \cap G])
= P(E) + P(F) + P(G) - P(E \cap F)
- P(E \cap G) + P(F \cap G) + P(E \cap F \cap G)
\]

This result can be generalized to a finite union of sets.

**Theorem 3** (Exclusion-Inclusion Identity, Proposition 2.4.4).

\[(**) \quad P\left(\bigcup_{i=1}^{n} E_i \right) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 \prec i_2} P(E_{i_1} \cap E_{i_2})
+ ... + (-1)^{r+1} \sum_{i_1 \prec i_2 \prec ... \prec i_r} P(E_{i_1} \cap E_{i_2} \cap ... \cap E_{i_r})
+ ... + (-1)^{n+1} P(E_1 \cap E_2 \cap ... \cap E_n)\]

where \(\sum_{i_1 \prec i_2 \prec ... \prec i_r} P(E_{i_1} \cap E_{i_2} \cap ... \cap E_{i_r})\) is taken over all \(\binom{n}{k}\) possible subsets of size \(r\) of the set \(\{1, 2, ..., n\}\).

**Proof.** (Induction) Ross mentions an inductive proof without giving details. The base case \((n = 2)\) is given by \((*)\). For the inductive
hypothesis assume the identity holds for \( n \).

\[
P(\bigcup_{i=1}^{n+1} E_i) = P(E_{n+1} \cup \bigcup_{i=1}^{n} E_i) = P(E_{n+1}) + P(\bigcup_{i=1}^{n} E_i) - P(\bigcup_{i=1}^{n} E_i \cap E_{n+1})
\]

\[
= \sum_{i=1}^{n+1} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + \sum_{i_1 < i_2 < i_3} P(E_{i_1} \cap E_{i_2} \cap E_{i_3}) + \ldots + (-1)^{r+1} \sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}) + (-1)^{r+2} \sum_{i_1 < i_2 < \ldots < i_r < i_{r+1}} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r} \cap E_{i_{r+1}}) + \ldots + (-1)^{n+1} P(E_1 \cap E_2 \cap \ldots \cap E_n)
\]

\[
- \sum_{i=1}^{n} P(E_i \cap E_{n+1}) + \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2} \cap E_{n+1}) - (-1)^{r} \sum_{i_1 < i_2 < \ldots < i_{r-1}} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_{r-1}} \cap E_{n+1}) - (-1)^{r+1} \sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r} \cap E_{n+1}) - \ldots - (-1)^n \sum_{i_1 < i_2 < \ldots < i_{n-1}} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_{n-1}} \cap E_{n+1}) - (-1)^{n+1} P(E_1 \cap E_2 \cap \ldots \cap E_n \cap E_{n+1})
\]

which gives the desired result when the terms with the same number of intersections are collected.

(Combinatorial) As an alternative, Ross, p. 31, gives a concise combinatorial proof of the exclusion-inclusion identity using the binomial theorem.
Theorem 4 (Upper and lower bounds on the probability of a union, Ross, Ch. 2, p. 31).

\[
\begin{align*}
(4.1) \quad & P\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} P(E_i) \\
(4.2) \quad & P\left(\bigcup_{i=1}^{n} E_i\right) \geq \sum_{i=1}^{n} P(E_i) - \sum_{j<i} P(E_i \cap E_{ij}) \\
(4.3) \quad & P\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} P(E_i) - \sum_{j<i} P(E_i \cap E_j) \\
& \quad + \sum_{k<j<i} P(E_i \cap E_j \cap E_k)
\end{align*}
\]

Ross give a detailed proof of these inequalities. Essentially, every time we add or subtract a sum in the exclusion-inclusion identity, we are overcorrecting.

REFERENCES