Section 1.10. The Probability of a Union of Events

Note. In this section we return to a set theoretic approach to probability and consider the probability of the union of an arbitrary (finite) number of events.

Theorem 1.10.1. For any events A_1, A_2, A_3 we have

$$\Pr(A_1 \cup A_2 \cup A_3) = \Pr(A_1) + \Pr(A_2) + \Pr(A_3) - \Pr(A_1 \cap A_2)$$

$$-\Pr(A_2 \cap A_3) - \Pr(A_1 \cap A_3) + \Pr(A_1 \cap A_2 \cap A_3).$$

Note. Theorem 1.10.1 can be generalized from three events to n events as follows.

Theorem 1.10.2. For any n events $A_1, A_2, \ldots, A_n$ we have

$$\Pr\left(\cup_{i=1}^n A_i\right) = \sum_{i=1}^n \Pr(A_i) - \sum_{i<j} \Pr(A_i \cap A_j) + \sum_{i<j<k} \Pr(A_i \cap A_j \cap A_k)$$

$$- \sum_{i<j<k<\ell} \Pr(A_i \cap A_j \cap A_k \cap A_\ell) + \cdots + (-1)^{n+1} \Pr(A_1 \cap A_2 \cap \cdots \cap A_n).$$

Note. We now use Theorem 1.10.2 to address **The Matching Problem**. DeGroot and Schervish describe a few different examples to illustrate the matching problem (see page 49). One is as follows. Suppose someone type n letters of correspondence to n different individuals and then types out the n different addresses for the individuals. The letters are randomly placed in envelopes. We want the probability p_n that at least one letter is in the correct envelope. Let A_i be the event that letter

i is places in the correct envelope (so $i = 1, 2, \ldots, n$). Then we want to calculate $\Pr(\cup_{i=1}^n A_i)$ using Theorem 1.10.2. Since there are n envelops then the probability that any given A_i is placed in the correct envelope is $1/n$. So $\sum_{i=1}^n \Pr(A_i) = n(1/n) = 1$. Next, $A_i \cap A_j$ for $i \neq j$ requires that A_i occurs, which happens with probability $1/n$, and then A_j occurs, which then happens with probability $1/(n-1)$. So $\Pr A_i \cap A_j) = \dfrac{1}{n(n-1)}$ for $i \neq j$. Now there are $\binom{n}{2}$ choices for i and j so

$$\sum_{i<j} \Pr(A_i \cap A_j) = \binom{n}{2} \frac{1}{n(n-1)} = \frac{1}{2!}.$$

Similarly for distinct i, j, k, $\Pr(A_i \cap A_j \cap A_k) = \dfrac{1}{n(n-1)(n-2)}$ and

$$\sum_{i<j<k} \Pr(A_i \cap A_j \cap A_k) = \binom{n}{3} \frac{1}{n(n-1)(n-2)} = \frac{1}{3!}.$$

Also, for m distinct A_i's, $\sum \Pr(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m}) = \dfrac{1}{m!}$. So, by Theorem 1.10.2,

$$p_n = \Pr(A_1 \cap A_2 \cap \cdots \cap A_n) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \frac{1}{5!} - \cdots + (-1)^{n+1} \frac{1}{n!}.$$

Note. Parameter p_n is the Matching Problem is a partial sum for the series

$$1 - e^{-1} = 1 - \sum_{i=0}^{\infty} \frac{(-1)^i}{i!}.$$

So for n "large," $p_n \approx 1 - e^{-1} \approx 0.63212$. In fact, since the series is an alternating series then by the Alternating Series Test (see my online Calculus 2 [MATH 1920] notes for 10.6. Alternating Series, Absolute and Conditional Convergence) we see that the value of p_n is within $1/(n+1)!$ of $1 - e^{-1}$. For example, with $n = 7$ we have $1/(7+1)! = 1/40,320 < 1/10,000$ and so p_n agrees with $1 - e^{-1}$ to four decimal places.

Note. In set theory, for a sequence of sets $A_1 \subset A_2 \subset \cdots$ we define $\lim_{n\to\infty} A_n = \cup_{i=1}^{\infty} A_n$. For a sequence of sets $A_1 \supset A_2 \supset \cdots$ we define $\lim_{n\to\infty} A_n = \cap_{i=1}^{\infty} A_n$. In Exercise 1.10.12 it is to be shown that for events $A_1, A_2, \ldots$ such that $A_1 \subset A_2 \subset \cdots$ we have

$$\Pr\left(\cup_{i=1}^{\infty} A_i\right) = \Pr\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \Pr(A_n).$$

In Exercise 1.10.13 it is to be shown that for events $A_1, A_2, \ldots$ such that $A_1 \supset A_2 \supset \cdots$ we have

$$\Pr\left(\cup_{i=1}^{\infty} A_i\right) = \Pr\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \Pr(A_n).$$

Since this involves passing a limit in and out of the Pr function, this is called *continuity of the probability function.* This idea is similarly encountered in continuity of Lebesgue measure; see my online note for Section 2.5. Countable Additivity, Continuity, and the Borel-Cantelli Lemma (see Theorem 2.15. Measure is Continuous).

Revised: 7/18/2019