Section 1.3. The Probability Set Function

Note. For an experiment we denote by $\mathcal{C}$ the sample space of all possible outcomes. We need to define a set function that assigns a probability to subsets of $\mathcal{C}$; the subsets of $\mathcal{C}$ for which the probability is defined is called an *event*. We denote the collection of events as $\mathcal{B}$. If $\mathcal{C}$ is a finite set, then we hope to assign a probability to all subsets of $\mathcal{C}$ (that is, to define a probability set function on the power set of $\mathcal{C}$). More generally, we require that the collection of subsets of $\mathcal{C}$ for which a probability (i.e., the events in $\mathcal{B}$) is defined to satisfy: (1) the sample space $\mathcal{C}$ itself is an event, (2) the complement of every event is again an event, and (3) every countable union of events is again an event. Symbolically, this means (1) $\mathcal{B} \in \mathcal{B}$, (2) if $A \in \mathcal{B}$ then $A^c \in \mathcal{B}$, and (3) if $A_1, A_2, \ldots \in \mathcal{B}$ then $\cup_{n=1}^{\infty} A_n \in \mathcal{B}$. Combining (2) and (3), we see by DeMorgan's Law (for countable unions) that if $A_1, A_2, \ldots \in \mathcal{B}$ then $\cap_{n=1}^{\infty} A_n \in \mathcal{B}$ (since $\cap_{n=1}^{\infty} A_n = (\cup_{n=1}^{\infty} A_n^c)^c$). So the collection of events $\mathcal{B}$ is closed under complements, countable unions, and countable intersections. Such a collection of sets form a σ-*field* (or σ-*algebra*).

Note. For more information on σ-fields (or "σ-algebras) we shoudl consider the introductory material on measure theory. See my online notes for Real Analysis 1 (MATH 5210) on Section 1.4. Borel Sets. In Real Analysis 1 we introduce Lebesgue measure on a collection of subsets of the real numbers which form a σ-algebra (the σ-algebra of Lebesgue measurable sets). Lebesgue measure is a generalization of the length of an interval and is used to set up Lebesgue integration, which is a generalization of Riemann integration. A class on modern probability theory requires a knowledge of Lebesgue measure and Lebesgue integration (given in Real Analysis 1, MATH 5210), abstract measure and integration (given in Real Analysis 2, MATH 5220), and Hilbert space, and linear operators on normed linear spaces (given in Fundamentals of Functional Analysis, MATH 5740). You can find my class notes for these classes, as well as notes on probability theory as follows:

- Real Analysis 1
- Real Analysis 2
- Fundamentals of Functional Analysis
- Measure Theory Based Probability

The book *Probability and Statistics*, 4th Edition, by M. H. DeGroot and M. J. Schervish (Pearson, 2012), a text for a slightly less advanced course than Mathematical Statistics 1 (MATH 4047/5047), does a nice job of addressing some of these ideas. See the online notes from this source on 1.4. Set Theory.

Note. Based on the intuitive relative frequency approach to probability (see Section 1.1), we are motivated to state the following definition.

Definition 1.3.1. Let $\mathcal{C}$ be a sample space and let $\mathcal{B}$ be the set of all events (thus, $\mathcal{B}$ is a σ-field). Let R be a real-valued function defined on $\mathcal{B}$. Then P is a *probability set function* if P satisfies the following three conditions:

1. $P(A) \geq 0$ for all $A \in \mathcal{B}$.

2. $P(\mathcal{C}) = 1$.

3. If $\{A_n\}$ is a sequence of events in $\mathcal{B}$ and $A_m \cap A_n = \varnothing$ for all $m \neq n$, then $P\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ (in measure theory, this is called *countable additivity*).

Definition. A collection of events $\{A_n \mid n \in I\}$ (where I is some indexing set) such that $A_i \cap A_j = \varnothing$ is a *mutually exclusive collection* of events. A countable collection of mutually exclusive events $\{A_n \mid n \in \mathbb{N}\}$ such that $\cup_{n=1}^{\infty} A_n = \mathcal{C}$ is called *exhaustive* (and notice $\sum_{n=1}^{\infty} P(A_n) = 1$) and the sets $\{A_n \mid n \in \mathbb{N}\}$ is said to *partition* sample space $\mathcal{C}$.

Note. For mutually exclusive events A and B, we denote their union as $A \mathbin{\dot\cup} B$ (and similarly for a countable collection of mutually exclusive events).

Note. In what follows, it is understood that the probability set function (or simply "probability function") is defined on some σ-field $\mathcal{B}$ of events. For example, if A is an event, then A^c is an event.

Theorem 1.3.1. For each event $A \in \mathcal{B}$, $P(A) = 1 - P(A^c)$.

Theorem 1.3.2. The probability of the null set is zero; that is, $P(\varnothing) = 0$.

Note. We now have, by Definition 1.3.1(3) (countable additivity), for finite number of mutually exclusive events $A_1, A_2, \ldots, A_N$ that with $A_{N+1} = A_{N+2} = \cdots = \varnothing$,

$$P(A_1 \cup A_2 \cup \cdots \cup A_N) = P\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{N} P(A_n).$$

This property is called *finite additivity*, so we see that this is an implication of Definition 1.3.1 and the fact that the events form a σ-field.

Theorem 1.3.3. If A and B are events such that $A \subset B$, then $P(A) \leq P(B)$ (in measure theory, this is called *monotonicity*).

Theorem 1.3.4. For each event $A \in \mathcal{B}$ we have $0 \leq P(A) \leq 1$.

Theorem 1.3.5. If A and B are events in $\mathcal{C}$, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Examples. Exercises 1.3.4 and 1.3.6.

Definition 1.3.2. Let $\mathcal{C} = \{x_1, x_2, \ldots, x_m\}$ be a finite sample space. Let $p_i = 1/m$ for $i = 1, 2, \ldots, m$ and for all subsets A of $\mathcal{C}$ define

$$P(A) = \sum_{x_i \in A} \frac{1}{m} = \frac{|A|}{m} = \frac{\#(A)}{m},$$

where $|A| = \#(A)$ denotes the number of elements in set A. Then P is a probability on $\mathcal{C}$ and is called the *equilikely case*.

Note. The equilikely case arises in several common examples, such as rolling dice or dealing cards.

Note. We now address three counting techniques or "rules."

Note/Rule 1. Let $A_1 = \{x_1, x_2, \ldots, x_m\}$ and $B - \{y_1, y_2, \ldots,_m\}$ so that $|A| = m$ and $|B| = n$. Then there are mn ordered pairs (x_i, y_j) where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$; that is,

$$|\{x_i, y_i) \mid x_i \in A, y_j \in B\}| = mn.$$

This is called the *mn-rule* or the *multiplicative rule*.

Definition. Let A be a set where $|A| = n$. An ordered k-tuple of distinct elements of A is called a *permutation* (of size k) of elements of A.

Note. We can use the multiplicative rule to count the number of permutations of size k on a set of size n. There are n choices for the first element of the k-tuple, then $n - 1$ choices for the second element (since the elements of the k-tuples are disjoint), then $n - 2$ choices for the third, and so forth up to the kth (last) element of the k-tuple for which there are $n - k + 1$ choices. The multiplication rule tells us that the total number of permutations is the product of the number of choices.

Note/Rule 2. The number of permutations of size k on a set of size n is

$$P_n^n = n(n-1)(n-2)\cdots(n-(k-2))(n-(k-1)) = \frac{n!}{(n-k)!}$$

where n-factorial is $n! = n(n-1)(n-2)\cdots(2)(1)$ and $0! = 1$.

Example 1.3.3. The Birthday Problem.

Suppose there are k people in a room where $k < 365$ and the people are "chosen at random." We want to find the probability that at least two people have the same birthday. First, number the people $1, 2, \ldots, k$. Create the k-tuples consisting of the birthdays (in order) of the k people. Since there are 365 possible birthdays then there are 365^k possible k-tuples of birthdays by the multiplication rule and these make up the sample space. We have that each birthday is equilikley (this is where "chosen at random" comes in) and so the probability of each k-tuple is $1/365^k$. The complement event of "at least two people have the same birthday" is the event "all birthdays are different." The number of k-tuples representing k different birthdays is the number of permutations of size k from a set of size of $n = 365$, of which there are $P_k^{365} = \dfrac{365!}{(365-k)!}$. So the probability that at least two people share a birthday is

$$1 - \frac{P_k^{365}}{265^n} = 1 = \frac{365!}{(365-k)!365^n}.$$

Given the large numbers, these are awkward computations, Some values of the probability for different values of k are:

k	Probability	k	Probability
2	0.003	21	0.444
3	0.008	22	0.476
4	0.016	23	0.507
5	0.027	24	0.538
6	0.040	25	0.569
7	0.056	26	0.598
8	0.074	27	0.627
9	0.095	28	0.654
10	0.117	29	0.681
11	0.141	30	0.706
12	0.167	31	0.730
13	0.194	32	0.753
14	0.223	33	0.775
15	0.253	34	0.795
16	0.283	35	0.814
17	0.315	40	0.891
18	0.347	50	0.970
19	0.379	60	0.994
20	0.411	100	0.9999997

Notice the surprising fact that at least two people in a random crowd have hte same birthday first tops 50% when there are only 23 people in the crowd.

Definition. Let A be a set where $|A| = n$. A subset of A of size k is a *combination* of k things taken from a set of n things.

Note/Rule 3. First, the number of permutations of size k of elements of a set of size n is $P_k^n = \dfrac{n!}{(n-k)!}$. Now any subset of size k of the given set can be arranged in $P_k^k = k!/0! = k!$ ways, so the number of combinations of k things taken from a set of n things, denoted $\binom{n}{k}$ or C_k^n, is

$$\binom{n}{k} = C_k^n = \frac{P_k^n}{k!} = \frac{n!}{k!(n-k)!}.$$

Note. The number of combinations $\binom{n}{k}$ arises in the Binomial Theorem as coefficients (so $\binom{n}{k}$ is also called a binomial coefficient):

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Example 1.3.4. Poker Hands. A standard deck of playing cards contain 52 cards. There are 4 suits, spade, club, diamond, heart, and 13 cards in each suit. If you are dealt 5 cards (at random) then the number of possible outcomes for this (since order doesn't matter) is $C_5^{52} = \dfrac{52!}{5!47!} = \dfrac{(52)(51)(50)(49)(48)}{(5)(4)(3)(2)(1)} = 2{,}598{,}960.$ Let E be the event that you are dealt three-of-a-kind (the other two cards are distinct and are of different kids). We apply the multiplication rule to count the possible number of hands containing three-of-a-kind. First, choose the "kind"; there are 13 different kinds of cards (namely, $2, 3, \ldots, 9, 10, J, Q, K, A$) so there is $\binom{13}{1} = 13$ ways to choose the kind. Given the kind, there are 4 such cards (one of each suit) so there are $\binom{4}{3} = 4$ ways to choose the cards of a given kind. Now the other two cards must be one of the 12 remaining kinds and must themselves be different kinds. So there are $\binom{12}{2} = 66$ ways to choose the other two kinds and then $\binom{4}{1} = 4$ ways to choose each the fast two given kinds of cards. So, the possible number of hands of cards containing three-of-a-kind is $\binom{13}{1}\binom{4}{3}\binom{12}{2}\binom{4}{1}^2 = 54{,}912$ and the probability of being dealt such a hand is

$$P(E) = \frac{\binom{13}{1}\binom{4}{3}\binom{12}{2}\binom{4}{1}^2}{\binom{52}{5}} \approx 0.0211.$$

Example. Exercise 1.3.19.

Note. Recall that a function f defined in an open interval containing point a is continuous at $x = a$ when $\lim_{x \to a} f(x) = f(a)$. This terminology motivates the name of the following result. Also recall that the σ-field of events is closed under countable unions and countable intersections.

Theorem 1.3.6. Continuity of the Probability Functions.
Let $\{C_n\}$ be a nondecreasing sequence of events. Then

$$\lim_{n\to\infty} P(C_n) = P\left(\lim_{n\to\infty} C_n\right) = P\left(\cup_{n=1}^{\infty} C_n\right).$$

Let $\{C_n\}$ be a nonincreasing sequence of sets. Then

$$\lim_{n\to\infty} P(C_n) = P\left(\lim_{n\to\infty} C_n\right) = P\left(\cap_{n=1}^{\infty} C_n\right).$$

Note. The next theorem is called "countable subadditivity" in measure theory.

Theorem 1.3.7. Boole's Inequality/Countable Subadditivity.

Let $\{C_n\}$ be an arbitrary sequence of events. Then

$$P\left(\cup_{n=1}^{\infty} C_n\right) \leq \sum_{n=1}^{\infty} P(C_n).$$

Note. The following is established in Exercise 1.3.9.

Theorem 1.3.A. Inclusion Exclusion Formula.

For events C_1, C_2, C_3 we have

$$P(C_1 \cup C_2 \cup C_3) = p_1 - p_2 + p_3$$

where

$$\begin{aligned} p_1 &= P(C_1) + P(C_2) + P(C_3) \\ p_2 &= P(C_1 \cap C_2) + P(C_1 \cap C_3) + P(C_2 \cap C_3) \\ p_3 &= P(C_1 \cap C_2 \cap C_3). \end{aligned}$$

Note. By induction, we can shoe the following more general version of the Inclusion Exclusion Formula; for a proof, see my online notes on Intermediate Probability and Statistics, Theorem 1.10.2 in 1.10. The Probability of a Union of Events.

Theorem 1.3.B. Inclusion Exclusion Formula.

For events $C_1, C_2, \ldots, C_k$ we have

$$P(C_1 \cup C_2 \cup \cdots \cup C_k) = p_1 - p_2 + p_3 - \cdots + (-1)^k p_{k-1} + (-1)^{k+1} p_k$$

where p_i equals the sum of the probabilities of all possible intersections involving i sets.

Note. When $k = 2$ in Theorem 1.3.B, we have Theorem 1.3.5:

$$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2).$$

Since $P(C_1 \cup C_2) \leq 1$, this implies $P(C_1 \cap C_2) \geq P(C_1) + P(C_2) - 1$. This is called *Bonferroni's Inequality.*

Revised: 9/14/2019