PROBABILITY

SECOLINSKY

Our development begins with definitions of a sample space, a σ-algebra, a probability measure, and an event.

Definition 1. *A sample space Ω is a non-empty set whose elements ω are called outcomes.*

The section of positive integers less than n will be denoted as S_n so that the experiment of rolling a six-sided dice will have S_7 as the sample space.

Definition 2. *Let Ω be a non-empty set. A σ-algebra $\mathcal{H}$ is a non-empty family of subsets of Ω which is closed under countable unions and complementation.*

Using Definition 2, we make a few observations. Since $\mathcal{H} \neq \emptyset$, there exists an $A \in \mathcal{H}$ so that $\Omega = A \cup A^c \in \mathcal{H}$. Also, $\emptyset = \Omega^c \in \mathcal{H}$. Now consider the countable sequence $A_n \in \mathcal{H}$ and note that $\cap A_n = (\cup A_n^c)^c \in \mathcal{H}$. To summarize, a σ-algebra $\mathcal{H}$ of Ω has the additional properties of being closed under countable intersections and of $\Omega, \emptyset \in \mathcal{H}$.

The σ-algebra we'll consider will always be the power set of Ω when the sample space is countable. As an example, the power set of S_7 is a σ-algebra of S_7. When Ω is not countable, we'll consider a less trivial σ-algebra . To understand what kind of σ-algebra it is, we'll need to show that an arbitrary intersection of σ-algebras is a σ-algebra.

Statement 1. *Let $\mathcal{A}$ be a family of σ-algebras of Ω. Then $\cap_{A \in \mathcal{A}} A$ is a σ-algebra.*

Proof. Assuming that $B = \cap_{A \in \mathcal{A}} A$ is nonempty, let $X \in B$ so that $X \in A$ for all $A \in \mathcal{A}$. Hence $X^c \in A$ for all $A \in \mathcal{A}$ and thus $X^c \in B$. Now let $X_n \in B$. Since $X_n \in A$, $\cup X_n \in A$ for all $A \in \mathcal{A}$. Hence $\cup X_n \in B$. □

It's clear from the statement that given a subset $\mathcal{T}$ of the power set of Ω and the collection $\mathcal{A}$ of all σ-algebras of Ω which contain $\mathcal{T}$, the smallest sigma algebra containing $\mathcal{T}$ is $\cap_{A \in \mathcal{A}} A$ and say it is the sigma algebra generated by $\mathcal{T}$ and denote it as $\langle \mathcal{T} \rangle$. For the case when Ω is not countable, we'll consider the σ-algebra generated by a topology $\mathcal{T}$ and name it a Borel σ-algebra where $E \in \langle \mathcal{T} \rangle$ is called a Borel set.

To give an example of a Borel σ-algebra, consider the standard topology $\mathcal{T}$ on the real line and let $\mathcal{L} = \{(a, b] \subset \mathbb{R} : a < b \text{ or } a = -\infty\}$ so that we may prove the equality $\langle \mathcal{L} \rangle = \langle \mathcal{T} \rangle$. Given $(a, b] \in \mathcal{L}$ and a strictly decreasing convergent sequence $b_n \downarrow b$, we'll first show that $\langle \mathcal{L} \rangle \subset \langle \mathcal{T} \rangle$. Because $(a, b_n) \in \mathcal{T}$ for all natural n, $(a, b] = \cap_{n \in \mathbb{Z}_+} (a, b_n) \in \langle \mathcal{T} \rangle$. So we have shown that $\mathcal{L} \subset \langle \mathcal{T} \rangle$ which implies that $\langle \mathcal{L} \rangle \subset \langle \mathcal{T} \rangle$. Now let $b_n \uparrow b$ be a strictly increasing sequence so that $(a, b) = \cup_{n \in \mathbb{Z}_+} (a, b_n] \in \langle \mathcal{L} \rangle$. It's clear now $\mathcal{T} \subset \langle \mathcal{L} \rangle$ and hence $\langle \mathcal{T} \rangle \subset \langle \mathcal{L} \rangle$.

The equality has been established so that we may now understand what are the Borel sets belonging to $\langle \mathcal{T} \rangle$. The sets $(a, b), (-\infty, b), (a, \infty), (-\infty, \infty)$ are Borel sets because they are open. The set $[a, b]$ is a Borel set because it is the complement of some open set. Also, $(a, b]$ and $(-\infty, b]$ are Borel sets since they belong to $\langle \mathcal{L} \rangle$ and $\langle \mathcal{L} \rangle = \langle \mathcal{T} \rangle$. Finally, it's not difficult to notice that, for a strictly increasing sequence $a_n \uparrow a$, $[a, \infty) = \cap_{n \in \mathbb{Z}_+} (a_n, \infty)$ and $[a, b) = \cap_{n \in \mathbb{Z}_+} (a_n, b)$ are Borel sets.

Date: July 25,07.

Probability provides the theoretical basis for statistics, which is the art of using data obtained randomly to draw conclusions about social and scientific phenomena. With that in mind, the building blocks of probability theory are introduced.

Axioms of Probability. *Consider a sample space Ω and let $\mathcal{H}$ be a σ-algebra of Ω where any element of $\mathcal{H}$ is called an event. A probability measure $P : \mathcal{H} \to \mathbb{R}$ satisfies the following:*

(i) $P(\Omega) = 1$,
(ii) $\forall A \in \mathcal{H}$, $P(A) \geq 0$, *and*
(iii) *for any countable sequence of events $A_n \in \mathcal{H}$ for which the A_ns are pairwise disjoint,* $P(\cup A_n) = \sum P(A_n)$.

It's important to notice that a probability measure is a function on events which assigns a likelihood to them, and that the nature of those events depend on the σ-algebra being considered. If the sample space is discrete, the power set will be considered as its σ-algebra. Otherwise, the Borel σ-algebra will be considered.

There are properties of a probability measure that are derived from using solely the Axioms of Probability. Some of the properties are below and left as exercises.

$$P(A) = 1 - P(A^c) \tag{1}$$

$$P(\emptyset) = 0 \tag{2}$$

$$A \subset B \Rightarrow P(A) \leq P(B) \tag{3}$$

$$P(A) \leq 1 \tag{4}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{5}$$

We'll now consider two examples of using the properties. The first one is known as the birthday problem. We'll use property (1) and the multiplication principle for counting to find the likelihood that at least two of k people share the same birthday. It'll be assumed that all k people were born in a common year of 365 days. So let $k \in \mathbb{Z}_+$ where $k \leq 365$, $U = S_{366}$, $\Omega = U^k$, and $A = \{x \in \Omega : (\forall i \neq j)(x_i \neq x_j)\}$. Since

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\prod_{i=0}^{k-1}(365 - i)}{365^k},$$

we obtain the answer of the birthday problem to be

$$1 - \frac{\prod_{i=0}^{k-1}(365 - i)}{365^k}.$$

As a second example, we use property (5) to derive the formula for the union of three events.

$$\begin{aligned}
P(A \cup B \cup C) &= P[(A \cup B) \cup C] \\
&= P(A \cup B) + P(C) - P[(A \cup B) \cap C] \\
&= P(A \cup B) + P(C) - P[(A \cap C) \cup (B \cap C)] \\
&= P(A \cup B) + P(C) - [P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)] \\
&= P(A) + P(B) + P(C) - [P(A \cap B) + P(A \cap C) + P(B \cap C)] + P[A \cap B \cap C].
\end{aligned}$$

A generalization of the aforementioned formula and (5) is known as the Inclusion-Exclusion Property for probability.

Theorem 1 (Inclusion-Exclusion Property)**.** *For any finite collection of events $\mathcal{A} = \{A_1, ..., A_n\}$, let S_k be the sum of intersection probabilities of the events in $\mathcal{A}$ taken $k > 2$ at a time. For $k = 1$, S_k is the sum of the probabilities $A \in \mathcal{A}$. Then $P(\cup_{A\in\mathcal{A}} A) = \sum_{k=1}^{n} S_k(-1)^{k+1}$.*

Proof. Suppose that ω is an outcome in $m > 0$ of the n events of $\mathcal{A}$. Obviously, the probability of the simple event $\{\omega\}$ is added once on the left side of the equation. For the right side of the equation, the probability of $\{\omega\}$ is added $\sum_{j=1}^{m} \binom{m}{j}(-1)^{j+1}$ times since S_j has $\binom{m}{j}$ terms with the probability of $\{\omega\}$. Because $0 = 0^m = [1 + (-1)]^m = \sum_{j=0}^{m} \binom{m}{j} 1^{m-j}(-1)^j = \binom{m}{0} - \sum_{j=1}^{m} \binom{m}{j}(-1)^{j+1}$, $\sum_{j=1}^{m} \binom{m}{j}(-1)^{j+1} = 1$. Thus the right side of the equation has the probability of $\{\omega\}$ added once. □

An example that uses Theorem 1 is now given. Let Ω be the set of a thousand sitcom fans that a TV newtork had surveyed. The survey had asked whether they were married, working, or graduated from college. Letting M, W, and G denote the sets of 470 married, 312 working, and 525 graduated people, respectively, it was reported that $|M \cap W| = 86$, $|M \cap G| = 147$, $|W \cap G| = 42$ and $|M \cap W \cap G| = 25$. The network reported incorrect numbers since

$$\begin{aligned}
&P(M \cup W \cup G) \\
&= P(M) + P(W) + P(G) - [P(M \cap W) + P(M \cap G) + P(W \cap G)] + P(M \cap W \cap G) \\
&= \frac{470 + 312 + 525}{1000} - \frac{86 + 147 + 42}{1000} + \frac{25}{1000} = 1.057 > 1,
\end{aligned}$$

which is not possible for the probability measure.

Another example, which makes heavy use of the multiplication principle for counting, shows that the computations involved easily become cumbersome when using Theorem 1. The example begins with telling about someone who is responsible for taking care of four hats from four people at a party, but becomes drunk along with everybody else. When the four people leave the party, they are only aware that a hat was returned. Making the situation worse, the one responsible for taking care of the hats wasn't aware of which hat was returned to whom. To measure the likelihood of the event that at least one of the four received their appropriate hat, let $\Omega = \{x \in S_5^4 : (\forall i \neq j)(x_i \neq x_j)\}$. Also, for $i \in S_5$, consider the events $A_i = \{x \in \Omega : x_i = i\}$ that person i receives their hat i and observe that $|\Omega| = 4! = 24$. Since the A_is are not mutually exclusive, we'll use Theorem 1.

$$\begin{aligned}
&P(A_1 \cup A_2 \cup A_3 \cup A_4) \\
&= P(A_1) + P(A_2) + P(A_3) + P(A_4) \\
&- [P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_1 \cap A_4) + P(A_2 \cap A_3) + P(A_2 \cap A_4) + P(A_3 \cap A_4)] \\
&+ P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_4) + P(A_1 \cap A_3 \cap A_4) + P(A_2 \cap A_3 \cap A_4) \\
&- P(A_1 \cap A_2 \cap A_3 \cap A_4) \\
&= 4(\frac{6}{24}) - 6(\frac{2}{24}) + 4(\frac{1}{24}) - \frac{1}{24} \\
&= \frac{15}{24}.
\end{aligned}$$

It's about 63 percent probable that someone will have their hat returned, but more interesting are the observations $\frac{15}{24}, \frac{12}{24} \leq P(A_1 \cup A_2 \cup A_3 \cup A_4) \leq \frac{24}{24}, \frac{16}{24}$ which are collectively known as Bonferroni's inequalities. For any even and odd numbers m and n less than the size of a finite family of events $\mathcal{A}$, the generalization of the observations is expressed as

$$\sum_{k=1}^{m} S_k(-1)^{k+1} \leq P(\bigcup_{A\in\mathcal{A}} A) \leq \sum_{k=1}^{n} S_k(-1)^{k+1}$$

The proof of the general result is left as an exercise.

It's not always best to apply Theorem 1 to compute the probability of a union of events that are not mutually exclusive. The limitation is apparent when asking about the probability of having some element in the sample of k objects taken in order and with replacement from a universe of n.

Example 1. *Let $k, n \in \mathbb{Z}_+$, $t \in S_{n+1}$ be fixed, $\Omega = S_{n+1}^k$ and $B_i = \{x \in \Omega : x_i = t\}$. Then*

$$P(\bigcup_{i\in S_{k+1}} B_i) = 1 - \left(\frac{n-1}{n}\right)^k.$$

First note that $\left|\left(\cup_{i\in S_{k+1}} B_i\right)^c\right| = \left|\cap_{i\in S_{k+1}} B_i^c\right| = |\{x \in \Omega : x_1, ..., x_k \neq t\}| = (n-1)^k$. Assuming all $x \in \Omega$ to be equally likely, $P\left(\cap_{i\in S_{k+1}} B_i^c\right) = \left(\frac{n-1}{n}\right)^k$. The result of the example is clearly obtained from using property (1).

After introducing the definition of mutually independent events and statements of equivalence, a simpler way of obtaining the result of the example can be done.

Definition 3. *The events in the finite collection $\mathcal{A} \subset \mathcal{H}$ are mutually independent means that for any sub-collection $\mathcal{A}' \subset \mathcal{A}$,*

$$P(\bigcap_{A\in\mathcal{A}'} A) = \prod_{A\in\mathcal{A}'} P(A),$$

where it is assumed that both $\mathcal{A}$ and $\mathcal{A}'$ are nonempty.

Statement 2. *Let A and B be events. Then the following are equivalent.*

(i) *A and B are independent;*
(ii) *A and B^c are independent;*
(iii) *A^c and B are independent; and*
(iv) *A^c and B^c are independent.*

Proof.
(i) $\Rightarrow$ (ii). Note that because $A = (A \cap B^c) \cup (A \cap B)$ is the union of mutually exclusive events, $P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)[1 - P(B)] = P(A)P(B^c)$.

(ii) $\Rightarrow$ (iii). $P(A^c \cap B) = 1 - P(A \cup B^c) = 1 - [P(A) + P(B^c) - P(A \cap B^c)]$
$= 1 - P(A) - P(B^c) + P(A)P(B^c) = [1 - P(A)][1 - P(B^c)] = P(A^c)P(B)$.

(iii) $\Rightarrow$ (iv). Because $A^c = (A^c \cap B) \cup (A^c \cap B^c)$ is the union of disjoint events, $P(A^c \cap B^c)$
$= P(A^c) - P(A^c)P(B) = P(A^c)[1 - P(B)] = P(A^c)P(B^c)$.

(iv) $\Rightarrow$ (i). $P(A \cap B) = 1 - P(A^c \cup B^c) = 1 - [P(A^c) + P(B^c) - P(A^c \cap B^c)]$
$= [1 - P(A^c)][1 - P(B^c)]$. □

The generalization of Statement 2 is expressed below and its validity and truth is left for the reader to explore.

Theorem 2. *For some finite collection $\mathcal{A} \subset \mathcal{H}$ and any sub-collection $\mathcal{A}' \subset \mathcal{A}$ where $\mathcal{A}$ and $\mathcal{A}'$ are non-empty, the following are equivalent:*

(i) The events in $\mathcal{A}$ are mutually independent; and
(ii) The events in $\mathcal{A} - \mathcal{A}'$ and the complements of those in $\mathcal{A}'$ are mutually independent.

It's clear from the theorem that for the family of independent events $\{A_1, ..., A_n\}$,

$$P(\cup_{i=1}^{n} A_i) = 1 - \prod_{i=1}^{n}[1 - P(A_i)]. \tag{6}$$

The formula above can be used as a shortcut to using the Inclusion-Exclusion formula when the events are mutually independent. As an illustration, the result of Example 1 is easily obtained since the B_is are mutually independent.

Before a discussion of applying combinatorics to probability theory, two more statements of independence are introduced.

Statement 3. *If A and B are disjoint events with positive probability, then they are not independent.*

Proof. To prove the statement, assume to the contrary that $0 = P(\emptyset) = P(A \cap B) = P(A)P(B)$. Thus either $P(A) = 0$ or $P(B) = 0$. Assuming $P(B) = 0$, we obtain a contradiction. Therefore, if $P(A), P(B) > 0$ and $A \cap B = \emptyset$, then A and B are dependent. □

Statement 4. $(\forall A \in \mathcal{H})(\Omega$ *and* $\emptyset$ *are independent of* $A)$.

Proof. Let $A \in \mathcal{H}$ and note that $P(A \cap \Omega) = P(A)1 = P(A)P(\Omega)$ and that $P(A \cap \emptyset) = P(\emptyset) = 0 = P(A)0 = P(A)P(\emptyset)$. □

The experiment of randomly sampling from a population is one of a statistician's best tools for obtaining information about the population inexpensively and reliably. A sampling experiment done in order and with replacement has been introduced in Example 1. Other sampling situations are permutations and combinations.

The number of ways for selecting k objects in sequence and without replacement from a population of n objects is denoted as $P_{n,k}$. A permutation is one of those k-tuples. Using the multiplication principle for counting, the formula for $P_{n,k}$ is $n(n-1)\cdots(n-k+1)$. The following result shows a simpler way of calculating the number of permutations.

Statement 5. *Let $k, n \in \mathbb{Z}_+$ such that $k \leq n$. Then*

$$n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

Proof. Let T be the set of integers for which the statement holds true and note that $1 \in T$ since $1(1-k+1) = (1)(1) = 1 = \frac{1!}{(1-1)!}$. Now assume $n \in T$ so that when multiplying both sides by $(n+1)$ and then dividing by $(n-k+1)$ yields $(n+1)\cdots(n-k+2) = (n+1)\cdots((n+1)-k+1) = \frac{(n+1)!}{(n+1-k)!}$. Thus $n+1 \in T$. □

To find the likelihood that an individual m from a population S_{n+1} appears in a sample of $P_{n,k}$ permutations, let B_i be the event that m appears at the i^{th} coordinate for $i \in S_{k+1}$. Showing the probability of events B_i to be $\frac{1}{n}$ will test the understanding of Statement 5 and thus is left as an exercise. Now because the B_is are mutually exclusive, $\frac{k}{n}$ is the measurement for the likelihood. Therefore, we've shown that every individual in the population has an equal chance of $\frac{k}{n}$ of appearing in the sample.

It's intuitive to believe that it is more likely to have a member from a population in a sample of permutations than it is to have it in a sequenced sample done with replacement as in Example 1. The intuition that $\frac{k}{n} > 1 - (\frac{n-1}{n})^k$ for $k \leq n$ will be left as an exercise.

The number of subsets of size k from a universe of n objects is denoted as $C_{n,k}$, where a sample is said to be a combination of k chosen from n if it's one of those subsets. The relationship of the numbers $P_{n,k}$ and $C_{n,k}$ is described in the formula $P_{n,k} = k!C_{n,k}$. Note that $C_{n,k}$ is the binomial coefficient $\binom{n}{k}$ so that the next statement can be interpreted to give equality between the size of the power set of a universe of n objects and 2^n.

Statement 6. *Let $n \in \overline{\mathbb{Z}}_+$. Then*

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

Proof. The statement is true for $n = 0$ since $1 = \binom{0}{0} = 2^0$. Now assume the statement is true for $n-1$, i.e. $\sum_{k=0}^{n-1} \binom{n-1}{k} = 2^{n-1}$. Multiplying both sides by 2 yields the desired result that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$. $\square$

The next example is concerned with the experiment of sampling from a set of $C_{n,k}$ combinations and asking for the likelihood of an element from the universe of n objects to be in a sample.

Example 2. *Let $k, n \in \mathbb{Z}_+$ such that $k \leq n$ and $\Omega = \{B \subset S_{n+1} : |B| = k\}$ where $A = \{M \in \Omega : t \in M$ for some fixed $t \in S_{n+1}\}$. Then $P(A) = \frac{k}{n}$.*

Since $|A| = \binom{n-1}{k-1} = \frac{(n-1)!}{(k-1)!(n-k)!}$ and $|\Omega| = \binom{n}{k} = \frac{n!}{k!(n-k)!}$, the result of the example is obtained after simplifying the expression of $|A|$ divided by $|\Omega|$.

To introduce another counting scenario, consider k different types of objects where there are n_i of type i for $i \in S_{k+1}$ and let $n = \sum_{i=1}^{k} n_i$. A sample of the n objects in sequence and without replacement is taken such that no distinction is made among the n_i objects of type i. The total number N of such samples can be expressed in two formulae.

The derivation for one of the formulae requires the observation of a function being surjective. The f needed is a map from the set of $P_{n,n}$ permutations to the set of N samples which takes x to a sample with type i objects at positions corresponding to where the type is in x. Since f is surjective, the N pre-images are nonempty with a size of $n_1! \cdot \cdots \cdot n_k!$ and from the definition of a function, all N pre-images are disjoint. Thus the sum of the sizes of all N pre-images is $P_{n,n}$.

The second formula is derived from using the multiplication principle for counting. Again thinking of the n positions of a sequence, n_i of the $n - \sum_{j=1}^{i-1} n_j$ positions are chosen to place the objects of type i for $i = 1$ to $i = k$. This k-step procedure will uniquely yield one of the N samples.

From the two derivations for N we have

$$\frac{n!}{\prod_{i=1}^{k} n_i!} = \prod_{i=1}^{k} \binom{n - \sum_{j=1}^{i-1} n_j}{n_i}. \tag{7}$$

It should be verified using formula (7) that there are $\binom{n}{k}$ sequences which have exactly k ones and $n-k$ zeros. Therefore, the sum of $\binom{n}{k}$ over all possible k from 0 to n will give the number of all possible n-tuples of ones and zeros. This sum equals 2^n since for each coordinate of a n-tuple there are two choices, a one or a zero. We've established an alternative interpretation of Statement 6 for which we'll now give two examples.

Example 3. *Let* $U = \{0, ..., 9\}$, $\Omega = U^4$, *and* $A = \{x \in \Omega : (\exists i \in S_5)(x_i \geq 6)\}$. *Then the probability of event* A *is approximately* 87%.

Consider the surjective function f which maps elements x from Ω to elements y of the set of 2^4 sequences of ones and zeros such that $y_i = 1$ if $x_i \geq 6$ and $y_i = 0$ otherwise. The k ones of y determine the cardinality of its pre-image to be $6^{4-k}4^k$. Hence the size of A is a sum of the $\sum_{k=1}^{4} \binom{4}{k}$ sizes of disjoint pre-images. It follows that

$$|A| = \sum_{k=1}^{4} \binom{4}{k} 6^{4-k} 4^k = 8,704$$

and that division of $|A|$ by $|\Omega|$ gives the approximate of 87%.

Example 4. *Let* U *be a universe of* n *objects from which* m *are bad,* $0 \leq k \leq m$, *and* $0 < s \leq n$ *be the size of the sample. For* $\Omega_1 = \{x \in U^s : (\forall i \neq j)(x_i \neq x_j)\}$ *and* $\Omega_2 = \{B \subset U : |B| = s\}$, *consider the events* $A_1 = \{a \in \Omega_1 :$ *there are* k *bad in* $a\}$ *and* $A_2 = \{a \in \Omega_2 :$ *there are* k *bad in* $a\}$. *If* $s = m$, *then* $P(A_1) = P(A_2)$.

Sampling with or without order done without replacement doesn't make a difference when the number of bad elements m in the universe being sampled from is the same as the size s of the sample. The proof of Example 4 is done below.

$$\begin{aligned} P(A_1) &= \frac{\binom{s}{k} P_{m,k}\, P_{n-m,s-k}}{P_{n,s}} \\ &= \frac{\binom{s}{k} \frac{s!}{(s-k)!} \frac{(n-m)!}{[n-m-(s-k)]!}}{\frac{n!}{(n-s)!}} \\ &= \frac{\binom{s}{k} \frac{(n-m)!}{(s-k)![n-m-(s-k)]!}}{\binom{n}{s}} \\ &= \frac{\binom{m}{k}\binom{n-m}{s-k}}{\binom{n}{s}} = \\ P(A_2). \end{aligned}$$

Continuing to give alternative interpretations of what has so far been presented, consider a finite population of n objects U for which we'd like to partition into k subsets $A_1, \ldots, A_k$ of sizes $n_1, \ldots, n_k$ respectively. The N expressed in (7) can be interpreted to be the number of all possible sequences $(A_1, \ldots, A_k)$. Sampling one of the N sequences at random is done in the practice of statistics so as to reduce systematic confounding, i.e. the phenomenon that members of U who share a characteristic mostly belong to a particular set A_i.

Just as an alternative interpretation of (7) was given, an alternative proof of the equality is done using induction on k. Let T be the set of positive integers such that formula (7) holds true. When $k = 1$, then $n_1 = n$ which makes the statement true.

Assuming $k-1 \in T$ such that $k > 1$ and $n = \sum_{i=1}^{k-1} n_i$, let $n_{p_i} = n_{i-1}$ for $i \in S_{k+1}/\{1\}$, $n_{p_1} = n_k$ and $n^{'} = \sum_{i=1}^{k} n_{p_i}$. It follows that

$$\begin{aligned}
&\frac{n^{'}!}{\prod_{i=1}^{k} n_{p_i}!} \\
=&\frac{(n+n_k)!}{n_k!n!}\ \frac{n!}{\prod_{i=1}^{k-1} n_i!} \\
=&\binom{n+n_k}{n_k}\prod_{i=1}^{k-1}\binom{n-\sum_{j=1}^{i-1} n_j}{n_i} \\
=&\prod_{i=1}^{k}\binom{n^{'}-\sum_{j=1}^{i-1} n_{p_j}}{n_{p_i}}.
\end{aligned}$$

Hence $k \in T$ and so the proof is now finished.

The previous definition was a statement about the probability of the intersection of events and the next definition will help to give another formula for the intersection of events.

Definition 4 (Conditional Probability)**.** *Let event B have positive probability. The probability of event A given that B has occurred is*

$$\frac{P(A \cap B)}{P(B)}$$

and is denoted as $P(A|B)$.

Statement 7 (Multiplication Rule of Probability)**.** *Let $\mathcal{B}$ be a nonempty finite collection of n events all with positive probability and for which any intersection of the B_is are too with positive probability. Then $P[\cap_{i=1}^{n} B_i]$*

$$=\left(\prod_{i=1}^{n-1} P[B_i \mid \cap_{j=i+1}^{n} B_j]\right) P[B_n] \tag{8}$$

$$=P[B_1]\prod_{i=2}^{n} P[B_i \mid \cap_{j=1}^{i-1} B_j]. \tag{9}$$

Proof. Using induction, let A be the set of positive integers for which (8) holds true. If $|\mathcal{B}| = 2$, then

$$P(B_1 \cap B_2) = \frac{P(B_1 \cap B_2)}{P(B_2)} P(B_2) = P(B_1 \mid B_2)P(B_2)$$

and thus $2 \in A$. Now let $2 < n \in A$ so that formula (8) holds true for all $\mathcal{B}$ with size n and let $\mathcal{B}^{'}$ be of size $n+1$. Since $2 \in A$ it then follows that

$$\begin{aligned}
&P[B_{n+1} \cap (\cap_{i=1}^{n} B_i)] \\
=&P[B_{n+1}| \cap_{i=1}^{n} B_i] P[\cap_{i=1}^{n} B_i] \\
=&P[B_{n+1}| \cap_{i=1}^{n} B_i]\left(\prod_{i=1}^{n-1} P[B_i| \cap_{j=i+1}^{n} B_j]\right) P[B_n].
\end{aligned}$$

We've now shown that $n+1 \in A$. Therefore equality (8) holds for all natural $n \geq 2$. From the property of intersections being a commutative operation, equality (9) easily follows. ☐

To give an example, consider the experiment of sequentially sampling three people to speak in a panel discussion from a universe which has 4 people of a certain political party x, and 3 people of an opposing political party y. Given that a sample has at least 1 person of type x, the probability of having at least 2 people of type x in the sample is about 65%.

For the experiment, we have Ω to be the set of all $P_{7,3}$ permutations and so we consider the mutually exclusive events A_i that there are i people of type x in the sample. Observe that the size of A_i is $\binom{3}{i}P_{4,i}P_{3,3-i}$. From now on, the convention of omitting the intersection symbol will be practiced since it behaves much like the multiplication operator on $\mathbb{R}$ and with that in mind, the approximate is derived from the conditional probability below.

$$
\begin{aligned}
&P(A_2 \cup A_3 \mid A_1 \cup A_2 \cup A_3) \\
=&\frac{P[(A_2 \cup A_3)(A_1 \cup A_2 \cup A_3)]}{P(A_1 \cup A_2 \cup A_3)} \\
=&\frac{P(A_2 \cup A_3)}{P(A_1 \cup A_2 \cup A_3)} \\
=&\frac{P(A_2) + P(A_3)}{P(A_1) + P(A_2) + P(A_3)} \\
=&\frac{11}{17}.
\end{aligned}
$$

The next theorem can be exploited to better understand random walks and examples of Markov Chains concerned with machine states: If the machine is in state x today, then what's the probability of the machine being in state y after k days?

Theorem 3 (Law of Total Probability). *For all $A \in \mathcal{H}, \mathcal{B}$ is a partition of Ω made up of n events all with positive probability implies $P(A)$ is equal to*

$$\sum_{i=1}^{n} P(A \mid B_i)P(B_i).$$

The proof is very straightforward and follows from the definition of a partition to express A as a union of disjoint sets AB_i, axiom (iii) of probability, and the definition of conditional probability.

Now to show the usefulness of Theorem 3, random walks are introduced. Let $n \in \mathbb{Z}_+ \backslash S_2$ and consider the sample spaces Ω_j of all random walks which begin at $j \in \{1, \ldots, n-1\}$ and that take unit steps to the left or right until they are absorbed at either 0 or n. For events $A_j \subset \Omega_j$ of random walks absorbed at n, notice that events L and R of initially moving to the left or right from j partition Ω_j and thus $P(A_j) = P(A_j \mid L)P(L) + P(A_j \mid R)P(R) = P(A_{j-1})P(L) + P(A_{j+1})P(R)$. For sample spaces Ω_0 and Ω_n, $P(A_0) = 0$ and $P(A_n) = 1$.

We construct a $(n+1) \times (n+1)$ matrix A using the formula for the probabilities of the A_js. In the construction of A, all entries are 0 except for the entries otherwise described:

$$
\begin{array}{lcll}
a_{j+1,j+1} & = & 1 & \text{for all } j \in \{0, \ldots, n\} \\
a_{j+1,j} & = & -P(L) & \text{if } j \neq 0, n \\
a_{j+1,j+2} & = & -P(R) & \text{if } j \neq 0, n
\end{array}
$$

Letting $x = [P(A_0), \ldots, P(A_n)]$ and $y = [0, \ldots, 1]$, we have $Ax = y$. Solving for x will yield the probabilities of the A_js.

Example 5. *Consider the random walks* $A_0, \ldots, A_5$ *with right-step probability* .4. *We then solve for the system of equations*

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -.6 & 1 & -.4 & 0 & 0 & 0 \\ 0 & -.6 & 1 & -.4 & 0 & 0 \\ 0 & 0 & -.6 & 1 & -.4 & 0 \\ 0 & 0 & 0 & -.6 & 1 & -.4 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} P(A_0) \\ P(A_1) \\ P(A_2) \\ P(A_3) \\ P(A_4) \\ P(A_5) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

to yield the solution $[0, 0.0758294, 0.189573, 0.36019, 0.616114, 1]$.

To showcase another application of Theorem 3, imagine a machine to be in either of four states from one to four on any given day: from state one being in perfect condition to state four being in permanently damaged condition. The machine can't improve its condition. Furthermore, the machine can't be in perfect condition on a given day and in permanently damaged condition the next. The situation being imagined will be expressed as follows:

Letting $j, m, t \in S_5$, and $k \in \mathbb{Z}_+ \backslash S_2$, consider the sample space Ω_j^k

$$= \{x \in S_5^k : x_1 = j \wedge (\forall i < p)(x_i \leq x_p) \wedge (x_i = 1 \Rightarrow x_{i+1} \neq 4)\},$$

the set $A_{jm}^k \subset \Omega_j^k$ of elements x such that $x_k = m$ and $B_{jt}^k = \{x \in \Omega_j^k : x_{k-1} = t\}$ where $j \leq m, t$ by definition of Ω_j^k. Using the Law of Total Probability and letting $k > 3$ yields

$$\begin{aligned} P(A_{jm}^k) =& P[\bigcup_{q=j}^{m} (A_{jm}^k \cap B_{jq}^k)] \\ =& \sum_{q=j}^{m} P(A_{jm}^k \cap B_{jq}^k) \\ =& \sum_{q=j}^{m} P(A_{jm}^k \mid B_{jq}^k) P(B_{jq}^k) \\ =& \sum_{q=j}^{m} P(A_{qm}^2) P(A_{jq}^{k-1}). \end{aligned}$$

Observe that the Law of Total Probability doesn't guarantee that $P(A_{14}^3) = \sum_{q=1}^4 P(A_{q4}^2) P(A_{1q}^2)$ because $P(B_{14}^3) = 0$ as a result of B_{14}^3 being empty. Nevertheless, the formula holds true in the case when $k = 3$ and so we have from the recursive relation that for any $k \geq 3$,

$$\overrightarrow{a_j^k} = A^{k-2} \overrightarrow{a_j^2}$$

where

$$\overrightarrow{a_j^k} = [P(A_{jj}^k), \ldots, P(A_{j4}^k)]^T$$

and

$$A = \begin{pmatrix} P(A_{jj}^2) & \cdots & P(A_{4j}^2) \\ \vdots & \ddots & \vdots \\ P(A_{j4}^2) & \cdots & P(A_{44}^2) \end{pmatrix}.$$

To give a concrete example, consider the sample spaces

$$\Omega_1^3 = \{111, 112, 113, 122, 123, 124, 133, 134\},$$
$$\Omega_2^3 = \{222, 223, 224, 233, 234, 244\},$$
$$\Omega_3^3 = \{333, 334, 344\},$$
$$\Omega_4^3 = \{444\},$$

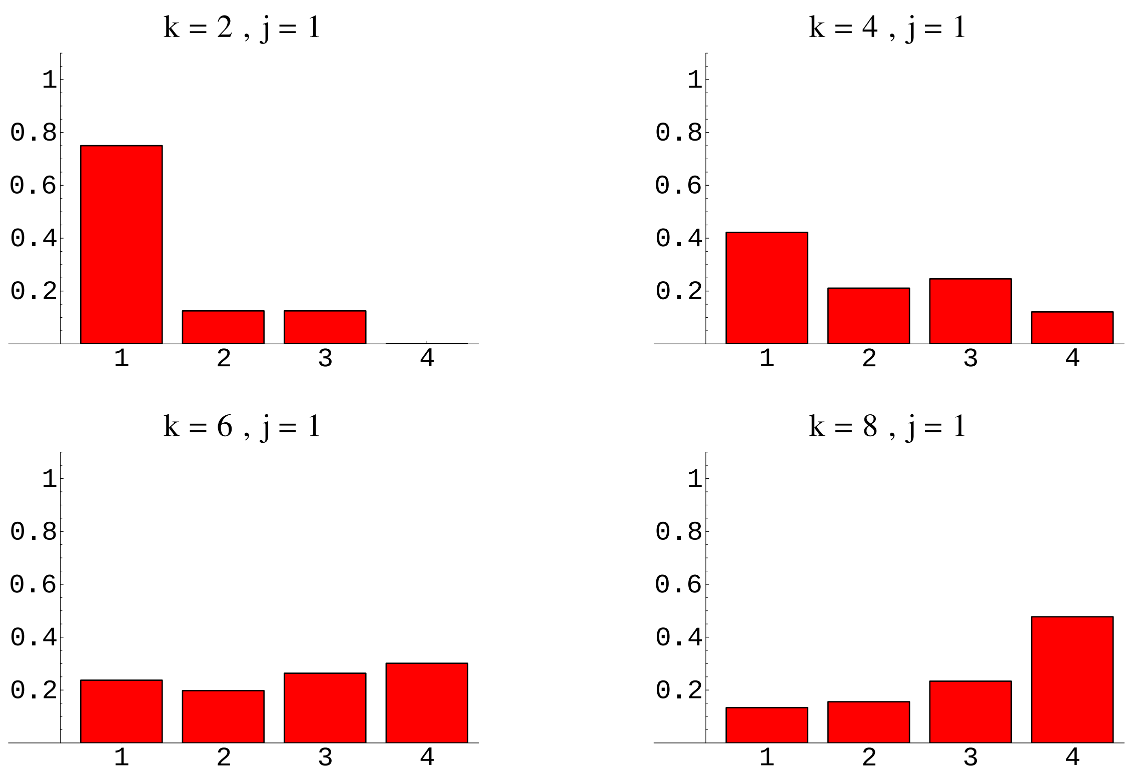
where the conditional probabilities of the events A_{jm}^2 are described in the table

$j \backslash m$	1	2	3	4
1	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	0
2	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
3	0	0	$\frac{3}{4}$	$\frac{1}{4}$
4	0	0	0	1

from which we obtain that

$$\overrightarrow{a_1^3} = A\overrightarrow{a_1^2} = (\frac{36}{64}, \frac{12}{64}, \frac{13}{64}, \frac{3}{64}),$$
$$\overrightarrow{a_2^3} = A\overrightarrow{a_2^2} = (0, \frac{9}{16}, \frac{3}{16}, \frac{4}{16}),$$
$$\overrightarrow{a_3^3} = A\overrightarrow{a_3^2} = (0, 0, \frac{9}{16}, \frac{7}{16}),$$
$$\overrightarrow{a_4^3} = A\overrightarrow{a_4^2} = (0, 0, 0, 1).$$

The distributions below give an affirmation to the validity of the model just constructed since they show what is to be expected; namely that as k becomes large, the more probable our machine will malfunction.



If it is desired to know the likelihood of the machine being in a certain state on the $(k-1)^{th}$ day knowing it was on state 4 on the k^{th} day, it'll help to be familiar with a certain formula.

Theorem 4 (Bayes' Formula)**.** *For all $A \in \mathcal{H}, \mathcal{B}$ is a partition of Ω with n events all with positive probability implies that $P(B_i \mid A)$ is equal to*

$$\frac{P(A \mid B_i)P(B_i)}{P(A)}$$

where $B_i \in \mathcal{B}$.

Applying Bayes' Formula to the study of machine states gives the result of $P(B^k_{jt} \mid A^k_{jm})$

$$\begin{aligned} &= \frac{P(A^k_{jm} \mid B^k_{jt})P(B^k_{jt})}{P(A^k_{jm})} \\ &= \frac{P(A^2_{tm})P(A^{k-1}_{jt})}{P(A^k_{jm})}. \end{aligned}$$

Continuing with the study of Ω^3_1, it can then be said that the likelihood of being on state $i \in S_5$ on the second day given that the machine malfunctioned on third day is 0, $\frac{1}{3}$, $\frac{2}{3}$, and 0 respectively.