A concise course in Probability Theory

Jacob Vartuli-Schonberg

January 2019

This is a recounting of the contents I learned through the reading of Y.A. Rozanov's Probability Theory: a concise course.

To purchase this book see http://store.doverpublications.com/0486635449.html

1 Chapter 2: Combination of Events

In this chapter of our concise course in Probability Theory, we will be covering the following concepts:

1. Elementary Events. The Sample Space.

2. The Addition Law for Probabilities

1.1 Elementary Events. The Sample Space.

A collection of elementary events, e_i, belonging to Ω are to subscribe to the following conditions:

$$\forall e_i \in (E \subseteq \Omega), e_j \cap e_k = \emptyset | i \neq j$$

Where the set Ω will be called the *sample space* or *the space of elementary events*. It should be noted that when one makes the statement 'the event A', we should always keep in mind that this could be taken in the sense of the union of elementary events which compose it. ie.

$$A = \bigcup_{\forall e \subseteq A} e$$

In some sense, we've just constructed the prime factorization of the event A. Just as factorization is a means to understand the natural numbers, a consideration for a sets factorization by union or intersection will undoubtedly be a useful mode of thought.

The equality of events. Suppose you're given events A_1 and A_2 such that the elementary event factorization of these sets are identical, then those two

events are also identical.

$$\forall e_i \in A_i \land \forall e_j \in A_j$$

1. $\exists f(e_i) = e_j \land \exists f^{-1}(e_j) = e_i$

The mutual exclusivity of events. Suppose you're given events A_1 and A_2 such that $A_1 \cap A_2 = \emptyset$, then these events are mutually exclusive. The way I'm thinking about this construction currently is as follows. Suppose you are able to capture the events of A_i into a vector v where the ith element of v is the ith elementary event of A_i. Let us next construct an outer product of two events where:

$$v_{A_n} v_{A_m}^T = V$$

$$V_{ij} = A_{m,e_i} \cap A_{n,e_j} = \emptyset(\forall(i,j))$$

Notice that if $A_m = A_n$ as defined in the previous section, this product resolves to something of an orthogonal square matrix. If the events are ordered particularly, it'll look very similar to an 'identity matrix'

$$v_{A_m} v_{A_m}^T = V$$

$$V_{ij} = A_{m,e_i} \cap A_{n,e_j} = \begin{cases} e_i & i = j \\ \emptyset & i \neq j \end{cases}$$

The difference of events. Suppose you're given events A_1 and A_2 arbitrarily. The difference of event A_1 and A_2 will be defined as follows:

$$A_1 - A_2 = W | \forall w \in W, w \in A_1 \land w \notin A_2$$

Notice that by this definition the difference is a non-commutative binary operation. In this moment, let's stop to consider a few different differences of events. Firstly, lets consider the function $f(\Omega, A_1) = \Omega - A_1$. The resolution of this expression are elements which belong to the sample space, Ω, and not to A_1. This, of course is the definition of the complement, A_1^c.

We might also consider the expression $f(A_1, \Omega) = A_1 - \Omega$. Notice these are the elements which belong to A_1 and not to Ω. Therefore, this difference must resolve to the empty set, $\emptyset$.

We should also notice that under the condition that A_1 and A_2 are mutually exclusive and $A = A_1 \cup A_2$ we have:

$$A_1 = A - A_2, A_2 = A - A_1$$

Additionally, again, provided that A_1 and A_2 are mutually exclusive, let us consider the difference $A_1^c - A^c$. In plain english, this expression resolves to elements which are not in A_1 and not not in A. That is elements in A and not in A_1. This of course, by our current definition of A, must resolve to A_2. More generally, if $A = \bigcup_{i=1}^n A_i$ and $V_{ij} = A_i \cap A_j = \emptyset, \forall i \neq j$ then

$$A_i^c - A^c = \bigcup_{j \neq i} A_j$$

1.2 The Addition Law for Probabilities

Let us perform an arbitrary experiment whose outcomes are mutually exclusive *(see section on mutual exclusivity above)*. Let those events be named B_k and the event which contains those $B = \bigcup B_k$. Given this configuration, for every experiment performed where our event B occurs, it must be the case that an event B_i, where $i \in k$, must have occurred. Put more specifically, given a function $w(*) = $ count of *, we can the state the previous as follows:

$$w(B) = w(\bigcup B_k) = \sum_{i \in k} w(B_i)$$

with M as a symbolic representation of the experiment, we can construct the relative frequency as:

$$\frac{w(B)}{w(M)} = \frac{w(\bigcup B_k)}{w(M)} = \frac{1}{w(M)} \sum_{i \in k} w(B_i)$$

If we're to consider the frequentist perspective as $w(M) \to \infty$:

$$P(B) = \lim_{w(M) \to \infty} \frac{w(B)}{w(M)} = \lim_{w(M) \to \infty} \sum_{i \in k} \frac{w(B_i)}{w(M)} = \sum_{i \in k} P(B_i)$$

This is called the *addition law for probabilities*

We will now discuss some important relations of probability that hold for arbitrarily constructed events A_i, A_j and $A \in \Omega$

1. $0 \leq P(A) < 1$

2. $P(A_i - A_j) = P(A_i) - P(A_i \cap A_j)$

3. $P(A_i \cup A_j) = P(A_i) + P(A_j) - P(A_i \cap A_j)$

4. for $A_i \subset A_j$, $P(A_i) < P(A_j)$

Pertaining to $0 \leq P(A) < 1$: Recall the frequentist perspective of probability:

$$P(A) = \lim_{w(M) \to \infty} \frac{w(A)}{w(M)}$$

:where M is the event that the experiment took place. Because A is defined as some proper or improper subset of M, it is necessary that $w(A) \leq w(M)$. If we're to guarantee that $w(*)$ is a mapping into the reals larger or equal to zero

and $w(M) \neq 0$ then the previous inequality can be expressed equivalently as $0 \leq \frac{w(A)}{w(M)} \leq 1$. In the limit, our expression will become

$$0 \leq \lim_{w(M) \to \infty} \frac{w(A)}{w(M)} = P(A) \leq 1$$

.

Pertaining to $P(A_i - A_j) = P(A_i) - P(A_i \cap A_j)$ Recall the definition of the difference of sets.

$$A_i - A_j = D | \forall d \in D, d \in A_i \wedge d \notin A_j$$

Again, this is a construction which forms an event which consists of elementary events belonging to A_i and excluding those which belong to A_j. In the case that $A_i \cap A_j = \emptyset$ as before, our difference $A_i - A_j$ would simply resolve to A_i and the statement above could be written as $P(A_i - A_j) = P(A_i)$. However, this expression only captures events which are non-intersecting.

We now justify the case when $A_i \cap A_j \neq \emptyset$. Recall that if the intersection of two events is non-empty, then there exists elementary events which suggest the occurrence of both the included sets. That is, when we construct the probability $P(A_i) = \Sigma_{e_i \in A_i} P(e_i)$ will capture elementary events which belong only to A_i and elements which belong to $A_i \cap A_j$. Upon recognizing this, it is a simple matter of removing from our computation elementary events which belong to that set $A_i \cap A_j$

$$P(A_i - A_j) = \Sigma_{e_i \in A_i} P(e_i) - \Sigma_{e_j \in A_i \cap A_j} P(e_j) = P(A_i) - P(A_i \cap A_j)$$

Pertaining to $P(A_i \cup A_j) = P(A_i) + P(A_j) - P(A_i \cap A_j)$. Our approach again will be from the perspective of elementary events. Recall that the event $A_i \cap A_j$ may be expressed as

$$A_i \cup A_j = \bigcup_{e_i \in A_i} e_i \cup \bigcup_{e_j \in A_j} e_j$$

provided that $A_i \cap A_j = \emptyset$ and so the probability is reasonably expressed as

$$P(A_i \cap A_j) = P(A_i) + P(A_j)$$

Of course, if the case is $A_i \cap A_j \neq \emptyset$ the previous equation does not hold due to the over-counting of elementary events which suggest both A_i and A_j. Upon closer inspection, the previous equation can be viewed in the following way:

$$= P(A_i \cap A_j)$$
$$= P(A_i) + P(A_j)$$
$$= \sum_{e_i \in A_i} P(e_i) + \sum_{e_j \in A_j} P(e_j)$$

4

$$= (\sum_{e_i \in A_i - A_j} P(e_i) + \sum_{e_i \in A_i \cap A_j} P(e_i)) + (\sum_{e_j \in A_j - A_i} P(e_j) + \sum_{e_j \in A_i \cap A_j} P(e_j))$$

$$= \sum_{e_i \in A_i - A_j} P(e_i) + \sum_{e_k \in A_i \cap A_j} 2P(e_k) + \sum_{e_j \in A_j - A_i} P(e_j)$$

You'll notice that when $A_i \cap A_j = \emptyset$ the interior term above will resolve to 0. However, when $A_i \cap A_j \neq \emptyset$ that term will resolve to $2P(A_i \cap A_j)$. To avoid this over counting, we will reform $P(A_i \cup A_j) = P(A_i) + P(A_j)$ into $P(A_i \cup A_j) = P(A_i) + P(A_j) - P(A_i \cap A_j)$

Pertaining to *for* $A_i \subset A_j$**,** $P(A_i) < P(A_j)$. We will again take the perspective of the elementary event. Suppose, as stated in the statement of the problem, that $A_i \subset A_j$. This suggests that upon deconstructing A_j into its elementary events, one might use the commutivity and associtivity of the union in the following way:

$$A_j = \bigcup_{e_j \in A_j} e_j = \bigcup_{e_i \in A_i \cap A_j} e_i \cup \bigcup_{e_j \in A_j - A_j} e_j = A_i \cup \bigcup_{e_j \in A_j - A_j} e_j$$

By the addition law of probabilities, we can transform the above equation into

$$P(A_j) = P(A_i) + \sum_{e_j \in A_j - A_i} P(e_j)$$

Which suggests:

$$P(A_j) > P(A_i)$$

In our next article, we will discuss the principle of inclusion/exclusion with respect to probabilities, the continuity property of probability and a set of problems will be given pertaining to our current article and the one which follows.

2 Practice Problems

Come back next week where for a collection of practice problems on this weeks topic, which will be further explored in the following article.