**Probability & Applied Statistics**

**Formula Sheet**

Hope Lepeshko

Definitions

Experiment: the process by which an observation is made

Simple event: an event that cannot be decomposed. Each simple event corresponds to one and ONLY one sample point. The letter *E* with a subscript will be used to denote a simple event

Sample space: is associated with an experiment and is the set consisting of all possible sample points, A sample space will be denoted by *S*

Discrete sample space: one that contains either a finite or a countable number of distinct sample points

Event: in a discrete sample space *S* is a collection of sample pints - that is, any subset of *S*

Permutation: an ordered arrangement of *r* distinct objects is called a premutation. The number of ways of ordering n distinct objects taken *r* at a time will be designated by the symbol

Definition of probability

Suppose S is a sample space associated with an experiment. To every event A in *S* (*A* is a subset of *S*), we assign a number, *P*(*A*), called the *probability* of A so that the following axioms hold:

Axiom 1:

Axiom 2:

Axiom 3: If from a sequence of pairwise mutually exclusive events in S

(That is ) then

The sample-point method

The following steps are used to find the probability of an event

1. Define the experiment and clearly determine how to describe one simple event.
2. List the sample event associated with an experiment and test each to make certain that it cannot be decomposed. This defines the sample space *S*
3. Assign reasonable probabilities to the sample pints in *S*, making certain that
4. Define the event of interest, *A*, as a specific collection of sample points. (A sample point is *A* if *A* occurs when the sample point occurs. Test **all** sample points in *S* to identify this in *A*
5. Find *P(A)* by summing the pourabilities of the sample points in *A*

Combinations: the number of combinations of n objects taken *r* at a time is the number of subsets, each of size *r*, that can be formed from the *n* objects. This number will be denoted by or

Conditional Probability of an event: the conditional probability of an event *A*, given that an event *B* has occurred is equal to

Provided [The symbol is read “probability of *A* given *B*”]

Independent: Two event *A* and *B* are said to be independent if any one of the following holds:

Otherwise, the events are said to be **dependent**

Partition: For some positive integer *k*, let the sets be such that

Then the collection of sets { is said to be a partition of *S*

Ransom variable: a random variable is a real-valued function for which the domain is a sample space

Random Sample: Let *N* and *n* represent the numbers of element in the population and sample, respectively. If the sampling is conducted in such a way that each of the samples have an equal probability of being selected, the sampling is said to be random, and the result is said to be a *random sample*.

Discrete: a random variable *Y* is said to be *discrete* if it can assume only a finite or countably infinite number of distinct values

Sum of the probabilities of all sample points: The probability that *Y* takes on the value y, , is defined as the *sum of the probabilities* of all sample points in *S* that are assigned the value *y*. We will sometimes denote by

.

Probability distribution: The *probability distribution* for a discrete variable *Y* can be represented by a formula, a table, or a graph the provides for all *y*

Expected Value: Let *Y* be a discrete random variable with the probability function . Then the *expected value* of *Y, E(Y)*, is defined to be

Standard Deviation: If *Y* is a random variable with mean , the variance of a random variable *Y* is defined to be the expected value of . That is

The standard deviation of *Y* is the positive square root of *V(Y)*

Binomial Experiment: A binomial experiment possesses the following properties:

1. The experiment consists of a fixed number, n, of identical trials.

2. Each trial results in one of two outcomes: success, *S*, or failure, *F.*

3. The probability of success on a single trial is equal to some value p and

remains the same from trial to trial. The probability of a failure is equal to

4. The trials are independent.

5. The random variable of interest is *Y*, the number of successes observed

during the *n* trials.

Binomial Distribution: A random variable Y is said to have a binomial distribution based on n trials with success probability p if and only if

Geometric Probability Distribution: A random variable Y is said to have a geometric probability distribution if and only if

Negative Binomial Probability Distribution: A random variable Y is said to have a negative binomial probability distribution if and only if

Hypergeometric probability distribution: A random variable Y is said to have a hypergeometric probability distribution if and only if

where y is an integer *0, 1, 2, . . . , n,* subject to the restrictions *y ≤ r and n − y ≤ N − r .*

Poisson Probability Distribution: A random variable Y is said to have a Poisson probability distribution if and only if

Moment-generating Function: The moment-generating function *m(t)* for a random variable *Y* is defined to be *m(t) = E(etY ).*We say that a moment-generating function for *Y* exists if there

exists a positive constant b such that *m(t)* is finite for *|t| ≤ b.*

Probability-generating function: Let Y be an integer-valued random variable for which P(Y = i ) = pi , where i = 0, 1, 2, . . . . The probability-generating function P(t) for Y is defined to

be

Kth factorial moment: The kth factorial moment for a random variable Y is defined to be

where k is a positive integer.

Theorems

mn Rule

With *m* elements and *n* elements , it is possible to form *mn = m x n* pairs containing one element from each group

Proof:

Verification of the theorem can be seen by observing the rectangular table in figure 2.9/ there is one square in the table for each p pair and hence a total of *m* x *n* squares

*A graph of a graph of a number

Description automatically generated with medium confidence*

Theorem 2.2

We are concerned with the number of ways of filling r positions with *n* distinct objects. Applying the extension of the *mn* rule, we see that the first object can be chosen in one of *n* ways. After the first id chosen, the second can be chosen in and the in

ways. Hence the total number of distinct arrangements is

Expressed in terms if factorials

Where and

Theorem 2.3

The number of ways of partitioning *n distinct objects into k* distinct groups containing objects, respectively, where each object appears in exactly one group and is

Proof:

*N* is the number of distinct arrangements of *n* objects in a row for a case in which rearrangement of the objects withing a group does not count. For example, the letters *a* to *l* are arranged in three groups, where

Is one such argument

The number of distinct arrangements if the *n* objects, assuming all objects are distinct, (from the theorem 2.2) Then equals the number of ways of partitioning the *n* objects to *k* groups (ignoring order within groups) multiplied by the number of ways of ordering the elemensts within each group. This application of the extended *mn* rule gives

Where Si the number of distinct arrangements of the objects in group *i* . Soling for *N*, we have,

Theorem 2.4

The number of unordered subsets of size *r* chose (without replacement) from *n* available objects is

Proof:

The selection or r objects form a total of n is equivalent to partitioning the n objects into k = 2 groups, the r selected, and the (n-r) remaining. This is a special case f the general partitioning problem dealt with in Theorem 2.3. In the present case and therefore,

Theorem 2.5

**The Multiplicative Law of Probability** is the probability of the intersection of two events *A* and *B* is

If A and B are independent, then

Proof:

The multiplicative law follows directly from the definition of conditional probability

Theorem 2.6

**The Additive Law of Probability** The probability of the union of two events *A* and *B* is

If A and B are mutually exclusive events. and

Proof:

The proof of the additive law can be followed by inspecting the Venn diagram in figure 2.10

A diagram of a diagram of a diagram

Description automatically generated with medium confidence

Notice that , where A and are mutually exclusive events. Further, , where ) and are mutually exclusive events. Then by Axiom 3,

and

The equality on the right implies that . Substituting this expression for into the expression for given in the left-hand equation of the preceding air, we obtain the desired result:

Theorem 2.7

If A is an event, then

Proof:

Observe that Because *A* and are mutually exclusive events, it follows that Therefore, and the result follows

Theorem 2.8

Assume that { is a partition of *S* such that , for Then for any event *A*

Proof:

Any subset *A* of *S* can be written as

Notice that, because is a partition of *S*, if

And that and are mutually exclusive events. Thus,

Theorem 2.9

**Bayes’ Rule** Assume that is a partition of *S* such that , for . Then,

Proof:

The proof follows directly form the definition of conditional probability and the law of total probability. Note that

Theorem 3.1

For any discrete probability distribution, the following must be true:

1. for all *y*
2. , where the summation is over all values if overall values of *y* with nonzero probability

Theorem 3.2

Let *Y* be a discrete random variable with probability function *p(y)* and *g(Y)* be a real-valued function of *Y*. Then the expected value of *g(Y)* is given by

Proof:

We prove the result in the case where the random variable Y take on the finite number of values . Because the function g(y) may not be one-to-one, suppose that g(Y) takes on values . (Where ). It follows g(Y) is a random variable such that for

Thus, by definition of *expected value*

Theorem 3.3

Let Y be a discrete random variable with probability function p(y) and c be a

constant. Then E(c) = c.

Proof:

Consider the function g(Y ) ≡ c. By Theorem 3.2,

But

Theorem 3.4

Let Y be a discrete random variable with probability function p(y), g(Y ) be a

function of Y , and c be a constant. Then

Proof:

By Theorem 3.2

Theorem 3.5

Let Y be a discrete random variable with probability function *p(y)* and *g1(Y ),*

*g2(Y ), . . . , gk(Y )* be *k* functions of *Y* . Then

Proof:

We will demonstrate the proof only for the case *k = 2*, but analogous steps will

hold for any finite k. By Theorem 3.2,

Theorem 3.6

Let *Y* be a discrete random variable with probability function *p(y)* and mean

*E(Y ) = μ;* then

Proof:

Noting that μ is a constant and applying Theorems 3.4 and 3.3 to the second

and third terms, respectively, we have

But *μ = E(Y )* and, therefore

Theorem 3.7

Let *Y* be a binomial random variable based on n trials and success probability *p*. Then

Proof:

By Definitions 3.4 and 3.7,

Notice that the first term in the sum is 0 and hence that

The summands in this last expression bear a striking resemblance to binomial probabilities. In fact, if we factor np out of each term in the sum and let *z = y−1,*

From Theorem 3.6, we know that . Thus, *σ2* can be

calculated if we find *E(Y 2)*. Finding *E(Y 2)* directly is difficult because

and the quantity *y2* does not appear as a factor of *y!*. Where do we go from here? Notice that

and, therefore,

Theorem 3.8

If Y is a random variable with a geometric distribution,

Proof:

This series might seem to be difficult to sum directly. Actually, it can be summed

easily if we take into account that, for *y ≥ 1*,

And hence,

(The interchanging of derivative and sum here can be justified.) Substituting, we obtain

The latter sum is the geometric series*, q + q2 + q3 + . . .,* which is equal to *q/(1 − q)* (see Appendix A1.11).Therefore,

To summarize, our approach is to express a series that cannot be summed

directly as the derivative of a series for which the sum can be readily obtained.

Once we evaluate the more easily handled series, we differentiate to complete

the process.

The derivation of the variance is left as Exercise 3.85.

Theorem 3.9

If *Y* is a random variable with a negative binomial distribution,

Theorem 3.10

If Y is a random variable with a hypergeometric distribution,

Theorem 3.12

If m(t) exists, then for any positive integer k,

Theorem 3.13

If P(t) is the probability-generating function for an integer-valued random variable, Y , then the kth factorial moment of Y is given by

Theorem 3.14

Tchebysheff’s Theorem Let Y be a random variable with mean μ and finite variance σ2. Then, for any constant k > 0,