PROBABILITY AND STATISTICS

ARTHUR RYMAN

1. Introduction

This article presents the basic concepts of probability and statistics from a mathematical point of view. The approach here differs from the usual approach found in most textbooks in that it raises the level of abstraction by viewing the subject matter in the light of category theory. The hope is that this point of view simplifies the subject by providing a concise unifying conceptual framework.

The main concept for unifying the subject is that of a probability space. Probability spaces are the objects of a category and random variables are its morphisms.

As an example of the utility of these concepts, consider the notion of a p-value. The standard definition of a p-value is a function that assigns to a measurement the probability of observing a measurement as extreme or more extreme than the given value. This definition is somewhat hard to grasp at first. In terms of category theory, a p-value is simply a morphism from a probability space to the uniform probability space on the unit interval. Here the uniform probability space on the unit interval is a very important important object in the category of probability spaces, and p-values are simply morphisms to it.

The underlying thesis of this article is that every important concept in probability and statistics has a natural and simple interpretation in terms of category theory.

2. Measurable Spaces

A probability spaces is a type of measurable space. A measurable space $M = (X, A)$ is a set X together with a distinguished set A of subsets that are capable of being measured. The measurable sets are required to form a σ-algebra under the usual operators of set theory.

The axioms for a σ-algebra of subsets of X are as follows:

- The empty set is measurable: $\emptyset \in A$
- X is measurable: $X \in A$
- The complement of a measurable set is measurable: $\forall Y \in A, X \setminus Y \in A$

- The union of a countable number of measurable sets is measurable: $\forall f : N- > A, \cup_{i \in N} f(i) \in A$

The axiom about the union of a countable number of measurable sets being measurable is related to the definition of the sum of infinite series.

A consequence of these axioms is that the intersection of a countable number of measurable sets is also measurable.

Clearly, if X is any set then its power set 2^X is a σ-algebra.

A measure space is a measurable space together with a measure. A measure if a function $m : A- > R^+$ that assigns to any measurable set, a non-negative real number, where we have extended the real numbers with positive infinity and defined addition in the obvious way.

A measure space satisifies the following axioms:

- The empty set has measure 0: $m(\emptyset) = 0$
- The measure of the union of a countable sequence of disjoint measurable sets is the sum of the measures of the sets in the sequence: $\forall f : N- > A | \forall i, j : N | i \neq j => A_i \cap A_j = \emptyset, m(\cup_i A_i) = \Sigma_i m(A_i)$