STAT131 Ch.3 Notes

William Santosa

Winter 2022 Quarter

1 Random variables and their distributions

Here, we will learn about *random variables*, a useful concept that simplifies notation and expands our ability to quantify uncertainty and summarize results of experiments. These random variables are essetnail through statistics and it's crucial to think about what they mean both intuitively and mathematically.

1.1 Random variables

To make the notion of random variable precise, we define it as a *function* mapping the sample space to the real line. In other words, given an experiment with sample space S, a *random variable* (r.v.) is a function from the sample space S to the real nu mbers $\mathbb{R}$. It's common, but not required, to denote random variables by capital letters. That means a random variable X assigns a numerical value $X(s)$ to each possible outcome s of the experiment. The randomness comes from the fact that the experiment itself it random, but the mapping itself is deterministic.

1.2 Distributions and probability mass functions

There are two main types of random variables: *discrete* random variables and *continuous* random variables.

The first, *discrete* random variable: A random variable X is said to be *discrete* if there is a finite list of values $a_1, a_2, ..., a_n$ or an infinite lst of values $a_1, a_2, ...$ such that $P(X = a_j)$ for some $j = 1$. If x is a discrete random variable, then the finite or countably infinite set of values x such that $P(X = x) > 0$ is called the *support* of x. Most applications of this varaible has the discrete random varaible as a set of integers.

The second, a *continuous* random variable, can take any real value within an interval; such random variables are defined more precisely later in Chapter 5.

The distribution of a random variab;e specifies the probabilities of all events associated with the random variable such as the probability of it equaling 3 and the probability of it being at least 110. There are multiple ways to express the distribution of a random variable. For a discrete random variable, the most natural way is to do so with a *probability mass function.*

A *probability mass function* of a discrete random varaible X is the function p_x given by $p_x(x) = P(X = x)$. This is positive if x is in the support of X, otherwise 0.

1. When writing $P(X = x)$, $X = x$ denotes an *event*, consisting of all outcomes s where X assigns the number x.
2. Defined as $X = x$ or formally, $s \in S : X(s) = x$. The former is shorter and more intuitive.

Let X be a discrete random variable with support $x_1, x_2, ...$ (assume these values are distinct and the support is countably infinite). The PMF p_x of X must satisify the following.

1. Nonnegative: $p_x(x) > 0$ if $x = x_j$ for some j and $p_x(x) = 0$ otherwise.
2. Sums to 1: $\sum_{j=1}^{\infty} p_X(x_j) = 1$

1.3 Bernouli and Binomial

There are two distributions that very prevalent in Statistics- so prevalent that these distributions have been given names. Thus, Bernoulli and Binomial distributions, both of which can only take two possible values, 0 and 1, are cornerstones of the Statistics world.

The Bernoulli distribution states that a random variable X with paper p has $P(X = 1) = p$ and $P(x = 0) = 1 - p$ where $0 < p < 1$. We write this as $X \sim Bern(p)$. The symbol $\sim$ is read "is distributed as".

That means *any* random variable whose possible variables are 0 and 1 has a Bern(p) distribution, with p the probability of the random variable equaling 1. The number p in Bern(p) is called the *parameter* of the distribution and specifies which specific Bernoulli distribution we have. If $X \sim$ Bern(1/3), you could say ””X is Bernoulli” and X is Bernoulli with parameter value 1/3.

An *indicator random variable* of event A is the random varaible which equals 1 if A occur and 0 otherwise. It is denoted by $I_A \sim$ Bern(p) with $p = P(A)$.

An experiment that results in either a success or a failure, exclusively, is called a *Bernoulli trial.* The random variable can be thought of as the indicator of success in a trial. 1 signifies a success and 0 as a failure in the trial. Thus, the parameter p is called the *success probability* of the Bern(p) distribution.

The binomial distribution has n *independent* Bernoulli trials performed, each with the same success probability p. Let X be the number of successes. The distribution of X is called the *Binomial distribution* with parameters n and p. We write $X \sim$ Bin(n, p) to mean that X has the binomial distribution with parameters n and p where n is a positive integer and $0 < p < 1$

The PMF of a binomial distribution, PMF of X, $X \sim$ Bin(n, p) is

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for k = 0, 1, ..., n (and $P(X = k) = 0$ otherwise).

Let $X \sim$ Bin(n, p) and $q = 1 - p$ (we often use q to denote the failure probability of a Bernoulli trial). Then $n - X \sim$ Bin(n, q).

1.4 Discrete Uniform

A discrete uniform distribution is one where C is a finite, nonempty set of numbers. Choose one of these numbers uniformly at random (equally likely) and call the chosen number X. Then X is said to have *Discrete Uniform distribution* with parameter C; denoted by $X \sim$ DUnif(C). The PMF of $X \sim$ DUnif(C) is

$$P(X = x) = \frac{1}{|C|}$$

for $x \in C$ (0 otherwise), since a PMF must sum to 1. That means for $X \sim$ DUnif(C) and any $A \subset C$. we have

$$P(X \in A) = \frac{|A|}{|C|}$$

1.5 Cumulative distribution functions

Another function that describes the distribution of a random variable is the *cumulative distribution function* (CDF). Unlike the PMF, which only discrete random variables possess, CDF is defined for all random variables.

The *cumulative distribution function* (CDF) of a random varaible X is the function F_X given by $F_X(x) = P(X \leq x)$. When there is no risk of ambiguity, we sometimes drop the subscript and just write F (or another letter) for a CDf.

1. Increasing: If $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$
2. Right-continuous: CDF is continuous except possibly for some jumps. With jumps, the CDF is continuous from the right. For any a, we have

$$F(a) = \lim_{x \to a^+} F(x)$$

3. Convergence to 0 and 1 in the limits:

$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to \infty} F(x) = 1$$

To convert between the two:

- PMF to CDF: Sum up all the PMF which is within the range of CDF
- CDF to PMF: Height of the jump in CDF at x is equal to value of PMF at x. Flat regions are outside the support, thus PMF is equal to 0.

1.6 Functions of random variables

The function of a random variable is a random varaible. If X is a random variable, then X^2, e^X, and $\sin(X)$, are also random varaibles, as is $g(X)$ for any function $g : \mathbb{R} \to \mathbb{R}$.

For any experiment with sample space S, a random variable X, and a function $g : \mathbb{R} \to \mathbb{R}$, g(X) is the random variable that maps s to g(X(s)) for all $s \in S$.

The PMF of g(X) is

$$P(g(X) = y) = \sum_{x : g(x) = y} P(X = x)$$

1.7 Independence of random variables

Random variables X and Y are *independent* if

$$P(X \leq x, Y \leq y) = p(X \leq x)P(Y \leq y)$$

for all $x, y \in \mathbb{R}$.

For discrete variables, it is equivalent to the condition

$$P(X = x, Y = y) = P(x = x)P(y = y)$$

for all x, y with x in support of X and y in support of Y.

Random variables $X_1, ..., X_n$ are *independent* if

$$P(X_1 \leq x_1, ..., X_n \leq x_n) = P(X_1 \leq x_1)...P(X_n \leq x_n)$$

for all $x_1, ..., x_n \in \mathbb{R}$. For infinitely many random variables, we say they are independent if every finite subset of the random variables are independent.