STAT 230

Notes: Syed Mustafa Raza Rizvi

June 17, 2020

These are my notes for the Probability course at the University of Waterloo (STAT 230). They are pretty similar to the online lecture content
You will find that these aren't very useful as notes, in the sense that they are not significantly shorter than the content in the lecture content, they're really just a way for me to type down the content I am learning and absorb it.
Thanks to Prof. Dan Wolczuk for providing me with the macros to typeset this LaTeX document.
If the university or staff feel that I should take down this document, please feel free to contact me on github

Week 1

1 Introduction to Probability

Probability

- A strong likely hood or chance of something
- The relative possibility an event will occur
- The ratio of the number of actual occurrences to the total number of possible occurences

In this course we will consider so-called "random" experiments that have several possible outcomes and are repeatable

Definitions of Probability

Let $\mathcal{S}$ be the set of all possible distinct outcomes of a random experiment. Then the probability of an event, provided that all outcomes are equally likely, is

$$\frac{\text{Number of ways the event can occur}}{\text{Total number of outcomes in } \mathcal{S}}$$

Definition
Classical Definition of Probability

The probability of an event in an experiement is the (limiting) proportion or fraction of times the event occurs in a very long (theoretically infinite) series of (independent) repetitions of the experiement

Definition
Relative Frequency Definition

The probability of an event is a 'best guess' by a person making the statement of the chances that the event will happen (e.g., 30% chance of rain)

Definition
Subjective Probability

The classical definition and the relative frequency definition are **consistent** with one another if we are careful in constructing our model.

Characteristics of a Random Experiment

- It should have more than one possible outcome
- We should be able to repeat the experiment under similar/identical conditions
- It *may* have equally likely outcomes

Probability Models

- A sample space of all possible outcomes of a random experiement must be defined
- A set of events is defined. An event is a subset of the sample space, to which we can assign a probability
- A way of assigning probabilites, which are numbers between 0 and 1, to events is specified.

2 Random Experiments and Sample spaces

When we repeat the experiement under *controlled conditions*, (repetitions are called **trials** of the experiment) different outcomes may occur

Definition
Random Experiment

Properties of a Random experiment

- We should be able to repeat it
- Different outcomes may occur on different trials even if the conditions are the same
- Outcomes have probabilities associated with them

A sample space, $\mathcal{S}$, is the set of distinct outcomes for an experiement or process, with the property that in a single trial, one and only one of these outcomes occurs

Definition
Sample Space

The sample space is a set and the outcomes in a sample space are called **sample points** or **points**

A **discrete sample space**, $\mathcal{S}$, is one with a **finite** number of sample points or **countably many** sample points

Definition
Discrete Sample Space

A set, $\mathcal{S}$, is countable if the elements can be put in a 1-1 correspondency with the positive integers

Definition
Countable

E.g. A **countably infinite** set is the sample space of a coin tossing experiment that ends on the event that a Tails occurs. The set

$$\mathcal{S} = \{T, HT, HHT, HHHT, ...\}$$

is countably infinite as we can form a 1-1 correspondence of each sample point with the positive integers, and since it is countable, it is discrete

Example of an **uncountably infinite** set is the sample space

$$\mathcal{S} = [0, \infty)$$

which has an infinite number of sample points, they **can not** be put into a 1-1 correspondence with the positive integers. The sample space $\mathcal{S}$ is not discrete

3 Probability Models and Events

An **event**, A, defined on a discrete sample space, $\mathcal{S}$, is a subset of $\mathcal{S}$. i.e. $A \subset \mathcal{S}$

Definition
Event

If the event $A \subset \mathcal{S}$ consists of only one sample point this A is called a **simple event**

Definition
Simple Event

If the event $A \subset \mathcal{S}$ consists of two or more sample points then A is called a **compound event**.
A is said to **occur** on a trial of the experiment if one of the simple events in A occurs

Definition
Compound Event

Let $\mathcal{S} = \{a_1, a_2, a_3, ...\}$ be a discrete sample space
Let $P(a_1), P(a_2), P(a_3), ...$ be a set of numbers associated with the sample points $a_1, a_2, a_3, ...$ such that:
1) $0 \leq P(a_i) \leq 1, i = 1, 2, ...$
2) $\sum_{i=1} P(a_i) = 1$
Then $P(a_i)$ is called a probability
The set $\{P(a_i), i = 1, 2, ...\}$ is called the **probability distribution on $\mathcal{S}$**

Definition
Probability Distribution on $\mathcal{S}$

The function $P(*)$ has the sample space $\mathcal{S}$ as its domain, the condition $\sum P(a_i) = 1$ reflects the idea that when the process or experiment happens, one or other of the simple events $a_i \in \mathcal{S}$ must occur. (The sum shows that it is certain the total probability of an outcome coming from the defined events in the sample space is 1, and hence 0 for anything outside the sample space). Events consisting of a single a_i in the sample space are referred to as **simple events**

Let $\mathcal{S}$ be a discrete sample space and let A be an event defined on $\mathcal{S}$, i.e. $A \subset \mathcal{S}$
Then $P(A)$, the probability of event A, is the sum of the probabilities corresponding to the sum of all the simple events that make up A, $P(A) = \sum_{a \in A} P(a)$

Definition
Probability of an Event

According to this definition $P(A) = P(a_1) + P(a_2)$ would be the probability of a compound event $A = \{a_1, a_2\}$
Note that for any event $A, 0 \leq P(A) \leq 1$

A discrete sample space $\mathcal{S} = \{a_1, a_2, ...\}$ together with a probability distribution $\{P(a_i), i = 1, 2...\}$ is referred to as a **discrete probability model**

Definition
Discrete Probability Model

The odds in favour of an event A is the probability the event occurs divided by the probability it does not occur or $\frac{P(A)}{1-P(A)}$. The odds against the vent is the reciprocal, $\frac{1-P(A)}{P(A)}$

Definition
Odds

4 Counting Techniques

When all the simple events have the same probability, for any even $A \subset \mathcal{S}$,

$$P(A) = \frac{\text{number of points in } A}{N(\text{number of points in } \mathcal{S})}$$

Definition
$P(A)$ for equi-probable outcomes

Suppose we can do job 1 in p ways and job 2 in q ways. Then we can do either job 1 **OR** job 2 (but not both), in a total of $p+q$ ways

Definition
Addition Rule

Suppose we can do job 1 in p ways and, for each of these ways, we can do job 2 in q ways. Then we can do both job 1 **AND** job 2, in a total of pq distinct ways

Definition
Multiplication Rule

Some useful combinatorial symbols

• $n^{(k)} = n(n-1)(n-2)...(n-k+1)$ (also written as ${}^nP_r$)
= *number of arrangements of n different elementes taken k at a time*

• $n! = n(n-1)...(2)(1)$
= *number of arrangements (permutations) of n different elements taken n at a time*

• $n^k = n...n$
= *number of arrangements of n elements taken k at a time allowing repeats*

• By definition $0^0 = 1,$ and $0! = 1,$ therefore $0^{(0)} = 1$

• Note $n^{(k)} = \frac{n!}{(n-k)!}$ when $k \geq 0$ is an integer

• Note that $n^{(k)}$ is also defined for a real number n, with k as a non-negative integer (e.g. $e^{(3)} = e(e-1)(e-2), 3^{(4)} = 3 \times 2 \times 1 \times 0$)

$$\binom{n}{k} = \frac{n^{(k)}}{k!} = \frac{n!}{(n-k)!k!} = {}^n C_k$$

Definition
n choose k

If n is a positive integer and k is a non-negative integer such that $k \leq n$ then ${}^nC_k$ is the number of subsets (combinations) of k elements which may be selected from a set containing n elements

$$\binom{n}{n_1, ..., n_k} = \frac{n!}{n_1!...n_k!}$$

Definition
Multinomial Coefficients

Week 2

Sets and Probability

Suppose $\mathcal{S}$ is a sample space for an experiment
Let $A, B, C, A_1, A_2, ...$ be events defined on $\mathcal{S}$

The union of A and B (written $A \cup B$) is the set of all points which are in either A or B or both — Definition: Union of Two Sets

The intersection of A and B (written $A \cap B$) is the set of all points which are in both A and B — Definition: Intersection of Two Sets

The complement of A (written $\bar{A}$) is the set of all points in $\mathcal{S}$ which are not in A — Definition: Complement of a Set

De Morgan's Laws

The complement of the union is the intersection of the complements:

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B}$$

The complement of the intersection is the union of the complements:

$$\overline{(A \cap B)} = \overline{A} \cup \overline{B}$$

Definition: Mutually Exclusive

Two events A and B are **mutually exclusive** if $A \cap B = \emptyset$, where $\emptyset$ is the empty set
The events $A_1, ..., A_k$ are called **(pairwise) mutually exclusive** if $A_i \cap A_j = \emptyset, \forall i, j$ with $i \neq j$
In general mutually exclusive events are such that no two events overlap

Definition: Probability Set function, Axioms of Probability

Suppose the function P associates a real value, $P(A)$, with each event A defined on a sample space $\mathcal{S}$ such that

1. $0 \leq P(A)$ for every event A
2. $P(S) = 1$
3. If $A_1, A_2, ...$ is a sequence of mutually exclusive events, then $P(A_1 \cup ... \cup A_n) = \sum_{i=1}^{n} P(A_i)$

then P is called a **probability set function** and $P(A)$ is called the probability of A

(1) - (3) are called the **Axioms of Probability**. These three axioms of probability are sufficient to allow a mathematical structure (the calculus of probability) to be developed.

Note: It does not matter if we assign the probabilities according to the classical, long-run relative frequency or subjective approach as long as the three axioms holds.

Theorem Properties of Probabilities

Suppose $\mathcal{S}$ is a sample space for an experiment
Let $A, B, C, A_1, A_2, ...$ be events defined on $\mathcal{S}$
Then:

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

$$P(A) = \sum_{a \in A} P(a) \leq \sum_{\text{all } a} P(a) = P(\mathcal{S}) = 1$$

2. **Probability of the Complement of an Event:**
$P(\overline{A}) = 1 - P(A)$ which implies $P(\emptyset) = 1 - P(\mathcal{S}) = 0$

$$\sum_{\text{all } a} P(a) = \sum_{a \in A} P(a) + \sum_{a \in \overline{A}} P(a) \Rightarrow 1 = P(A) + P(\overline{A})$$

3. **Probability of a subset of an Event:**
If $A \subset B$, then $P(A) \leq P(B)$

$$P(A) = \sum_{a \in A} P(a) \leq \sum_{a \in B} P(a) = P(B)$$

4. **Probability of the Union of Two Mutually Exclusive Events:**
Let A and B be mutually exclusive events
Then $P(A \cup B) = P(A) + P(B)$

$$P(A \cup B) = \sum_{a \in A \cup B} P(a) = \sum_{a \in A} P(a) + \sum_{a \in B} P(a) = P(A) + P(B)$$

(Note: This works since for any $a \in B, a \notin A$ and for any $a \in A, a \notin B$)

5. **Probability of the Union of Mutually Exclusive Events:**
Let $A_1, ..., A_n$ be mutually exclusive events
Then $P(A_1 \cup ... \cup A_n) = \sum_{i=1}^{n} P(A_i)$

$$A_i \cap A_j = \emptyset, i \neq j$$

6. For any two events A and B,
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$P(A) + P(B) = P(A \cup B) + P(A \cap B)$$

(This is since when we add $P(A)$ and $P(B)$ we end up double counting the points at the intersection, once when we count $P(A)$ and a second time when we count $P(B)$, so we subtract the intersection to make it count just once)

7. For any three events A, B and C,
$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Independent Events

Events A and B are **independent events** if and only if $P(A \cap B) = P(A)P(B)$. If they are not independent, we call the events **dependent**

Definition
Two Independent Events

Note: $P(A \cap B)$ can be written as $P(AB)$

Independent events are such that the ratios

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

The events $A_1, ..., A_n$ are independent if and only if $P(A_{i_1} \cap ... \cap A_{i_k}) = P(A_{i_1})...P(A_{i_k})$ for all sets $(i_1, ..., i_k)$ of distinct subscripts chosen from $(1, ..., n)$

Definition
n Independent Events

It is not sufficient for events to just be pairwise independent. All groupings have to be independent.

A and B are independent events if and only if $\overline{A}$ and B are independent events. Similarly A and $\overline{B}$ or $\overline{A}$ and $\overline{B}$

Theorem

Proof:
Note that $B = (A \cap B) \cup (\overline{A} \cap B)$, where $A \cap B$ and $\overline{A} \cap B$ are mutually exclusive events
So $P(B) = P(A \cap B) + P(\overline{A} \cap B)$. Therefore,

$$P(\overline{A} \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = P(B)(1 - P(A)) = P(\overline{A})P(B)$$

Week 3

Conditional Probability

The conditional probability of event A, given that event B occurs, is defined by

Definition

Conditional Probability

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

provided $P(B) \neq 0$

Suppose A and B are two ecents such that $P(A) > 0$ and $P(B) > 0$. A and B are independent if and only iff $P(A|B) = P(A)$ or, equivalently, if $P(B|A) = P(B)$

Theorem

For any two events A anf B,

Theorem

Mulitplication rule

1. $P(A \cap B) = P(A|B)P(B)$
2. $P(A \cap B) = P(B|A)P(A)$

Let the sample space, $\mathcal{S}$, be partitioned into k mutually exclusive sets $B_1, B_2, \ldots, B_k$ such that $P(B_i) > 0$, for $i = 1, 2, \ldots, k$ and $\mathcal{S} = B_1 \cup \cdots \cup B_k$.
Then, for any event A, $P(A) = \sum_{i=1}^{k} P(A \cap B_i)$

Theorem

Law of Total Probability

Let the sample space, $\mathcal{S}$, be partitioned into k mutually exclusive sets $B_1, \ldots, B_k$ such that $P(B_i) > 0$, for $i = 1, \ldots, k$ and $\mathcal{S} = B_1 \cup \cdots \cup B_k$.
Then, for any events A, and for $j = 1, \ldots, k$, we have

Theorem

Bayes' Theorem

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}$$

Proof:

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A \cap B_i)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}$$

Week 4

Random Variables and Distributions

A random variable (e.g. X) is a function that assigns a real number to each point in a sample space $\mathcal{S}$

Definition
Random Variable

The event $[X = 2]$ means $\{s \in \mathcal{S} | X(s) = 2\}$ or the set of all points, s, in the sample space, $\mathcal{S}$, such that $X(s) = 2$.
We write the probability of this event as $P(X = 2)$

The random variable, X, takes a finite or countably infinite number of distinct values (the range of X is countable)

Definition
Discrete Random Variable X

X can take any value in a non-degenerate interval (range of X is not countable)

Definition
Continuous Random Variable X

The **funtion** $f(x) = P(X = x)$, for all x in the set of all possible values of X

Definition
Probability Function

1. $f(x) \geq 0$ for all values of x
2. $\sum_{\text{all } x} f(x) = 1$ (sum of all probabilities is 1)

A cumulative distribution function (c.d.f) is defined as a function $F(x) = P(X \leq x)$, for all real numbers x

Definition
Cumulative Distribution Function c.d.f

Properties of a c.d.f $F(x)$

1. $F(x)$ is a non-decreasing function of x for all $x \in \mathbb{R}$
2. $0 \leq F(x) \leq 1$ for all $x \in \mathbb{R}$
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

Relation between $F(x)$ and $f(x)$

1. If a random variable, X, takes only non-negative integer values, then $F(x)$ is the probability of the values less than or equal to x (or up to and including)
2. $f(x) = F(x) - F(x-1)$ is the size of the **jump** in F at the point x, and
3. $F(x) = \sum_{z \le x} f(z)$

Discrete Uniform Distribution

Suppose X takes values $a, a+1, \ldots, b$ with all values being equally likely. Then X has a discrete uniform distribution, on $a, a+1, \ldots, b$. The probability function of the discrete uniform distribution is

Theorem

$$f(x) = \begin{cases} \frac{1}{b-a+1} & \text{for } x = a, a+1, \ldots, b \\ 0 & \text{otherwise} \end{cases}$$

Hypergeometric Distribution

We have a collection of N objects that can be classified into two distinct types, r of type 1 and $N-r$ of type 2.

Theorem

Pick a sample of $n < N$ ojects at random **without replacement**.
Let X be the number of type 1 in the sample.
Then X has a **hypergeometric distribution** with **probability function**

$$f(x) = \frac{(\text{ways to choose } x \text{ type 1}) \times (\text{ways to choose } n-x \text{ type 2})}{(\text{ways to choose sample of } n)}$$

$$= \frac{\binom{r}{x} \times \binom{N-r}{n-x}}{\binom{N}{n}}$$

The range of X is $\max(0, n-N+r) \le x \le \min(r, n)$

$$\sum_{\text{all } x} \binom{a}{x}\binom{b}{n-x} = \binom{a+b}{n}$$

Hypergeometric Identity

Acceptance sampling is a statistical method which is used for deciding whether a batch of items produced by a company is acceptable for distribution or not

Definition
Acceptance sampling

Binomial Distribution

Supposing we have an experiment with 2 possible outcomes which, for convenience, we call Success (S) and Failure (F). Suppose also that $P(S) = p$. Repeat the experiment (called a **trial**). Such a sequence of independent trials are called **Bernoulli trials**. Let the random variable X be the number of successes in n Bernoulli trials

Definition

Bernoulli trials

Key assumptions:

1. The probability of success p must be **constant** over the n trials

2. The n trials must be **independent**

The **probability function** (p.f.) of X is

Bernoulli's Formula

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n$$

A **binomial distribution** of n trials with probability of success p (and probability of failure $1-p$) is written as $BIN(n, p)$. As these two parameters are sufficient to describe the entire distribution

To check that $\sum f(x) = 1$

Binomial theorem

$$\sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} = (a+b)^n$$

Comparison of Binomial and Hypergeometric distribution

For both, there are 2 types of outcome on each trial, Success or Failure. Both count the number of **Successes**

The difference is that the Binomial requires **independent** trials (e.g. draws with replacement) and the Hypergeometric requires that the draws be made from a collection of objects **without replacement**.

For large N and relatively small n we can use the Binomial to approximate Hypergeometric

Week 5

Negative Binomial Distribution

Suppose we have a sequence of Bernoulli trials with $P(S) = p$.
Let the random variable X be the number of failures (F's) observed *before* obtaining the k^{th} success (S).

The p.f. of X is $f(x) = \binom{x+k-1}{x} p^k (1-p)^x$ for $x = 0, 1, 2, \ldots$ Theorem

We write $X \sim NB(k, p)$
(The negative binomial distribution can be defined to count the number of trials $(x + k)$ needed to get the k^{th} success instead)

Binomial vs Negative Binomial
For the binomial distribution, the number of trials is fixed, and the unkown is the number of successes.
For the negative binomial distribution, number of trials is not fixed, the number of successes is specified.

Geometric Distribution

Special case of the Negative Binomial distribution with $k = 1$
Then the random variable X is the waiting time or the number of failures until the **first** success occurs

The p.f. of X is $f(x) = P(X = x) = p(1-p)^x$ for $x = 0, 1, \ldots$ Theorem

We write $X \sim Geo(p)$
(c.d.f of $X \sim Geo(p)$ is $P(X \leq x) = 1 - (1-p)^{y+1}$, $y = 0, 1, \ldots$)

Poisson Distribution

Suppose that X represents the number of events of some type, occurring at the rate of $\mu > 0$. Then a random variable X has a **Poisson distribution** if the probability of X is: Definition Poisson Distribution

$$f(x) = \frac{\mu^x}{x!} e^{-\mu} \text{ for } x = 0, 1, 2, \ldots$$

We write $X \sim Poisson(\mu)$
The Poisson distribution is obtained when for a binomial distribution, n is large and p is small.

Hence, we can use a poisson distribution to approximate binomial distributions with large n's and small p's, where $\mu = np$

As $n \to \infty$, μ is constant and $p = \mu/n \to 0$, if $X \sim Bin(n, p)$ then

$$P(X = x) \to \frac{\mu^x}{x!}e^{-\mu} \text{ for } x = 0, 1, \ldots$$

Theorem
Poisson approximation to Binomial

Poisson Process

A poisson process is a model that describes events that occur randomly over time or space. There are three properties that need that are required

1. **Independence**: the number of occurrences in non-overlapping time intervals are independent

2. **Individuality**: The probability of 2 or more events in a sufficiently short period of time is approximately zero. i.e. $P(2 \text{ or more events in } (t, t + \Delta t)) = o(\Delta t)$ as $\Delta t \to 0$
   Note: $o(\Delta t)$ is called the 'order' notation. When a function $g(\Delta t) = o(\Delta t)$ as $\Delta t \to 0$, it means that $\frac{g(\Delta t)}{\Delta t} \to 0$ as $\Delta t \to 0$

3. **Homogeneity**: events occur at a uniform rate, λ, over time.
   i.e. $P(\text{one event in } (t, t + \Delta t)) = \lambda \Delta t + o(\Delta t)$

Suppose a process satisfies the three conditions above (independence, individuality, homogeneity).
Assume events occur at the average rate of λ per unit time
Let X be the number of events ina time interval of length t units. Then $X \sim Poisson(\mu = \lambda t)$

Theorem

Info from external readings

◦ A particular set S is well defined if it is possible to tell wether a particular point belongs to it or not.