Probability and Statistics by DeGroot Notes

Nathan Ueda

February 24, 2024

Contents

1 Introduction to Probability

1.1 The History of Probability

The use of probability to measure uncertainty and variability dates back hundredes of years.

1.2 Interpretations of Probability

Probability Interpretations

- Frequency: If an experiment is carried out many times, the frequency with which a particular outcome occurred would define its probability.
- Classical: If an outcome of some experiment must be one of n different, equally likely outcomes, the probability of each outcome is $\frac{1}{n}$.
- Subjective: An entity assigns probabilities to each possible outcome.

Probability theory does not depend on intepretation.

1.3 Experiments and Events

Probability allows us to quantify how likely an outcome is to occur.

Experiments: Any process in which the possible outcomes can be identified ahead of time.
Events: A well defined set of possible outcomes of the experiment (such as rolling an even number on a fair dice).

Although there is controversy in regard to the proper meaning and interpration of some of the probabilities that are assigned to the outcomes of many experiements, once these probabilities are assigned, there is complete agreemenet upon the mathematical theory of probability.

Almost all work in the mathematical theory of probability is related to:

- Methods for determining probabilities of certain events from given probabilities for each possible outcome in an experiment.
- Methods for revising probabilities of events when additional relevant information is obtained.

1.4 Set Theory

Sample Space: The collection of all possible outcomes of an experiment.
Empty Set: Subset of S containing no elements, denoted $\emptyset$, representing any events that cannot occur.
Complement: For some set A, its complement, denoted A^c, is the set containing all elements of S not in A.
Union: For n sets $A_1, \ldots, A_n$, their union, denoted $A_1 \cup \ldots \cup A_n$ or $\bigcup_{i=1}^{n} A_i$, is defined as the set containing all outcomes that belong to at least one of these n sets.
Intersection: For n sets $A_1, \ldots, A_n$, their intersection, denoted $A_1 \cap \ldots \cap A_n$ or $\bigcap_{i=1}^{n} A_i$, is defined as the set containing the elements common to all these n sets.
Disjoint/Mutually Exclusive: Two sets A and B are disjoint/mutually exclusive if they have no outcomes in common, that is, if $A \cap B = \emptyset$, representing that both A and B cannot occur.

1.5 The Definition of Probability

Axioms of Probability:

1. For every event A, $P(A) \geq 0$
2. $P(S) = 1$
3. $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

Basic Theorems:

1. $P(\emptyset) = 0$
2. For every finite sequence of n disjoint events, $A_1, \ldots, A_n$, $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$
3. For every event A, $P(A^c) = 1 - P(A)$
4. If $A \subset B$, then $P(A) \leq P(B)$
5. For every event A, $0 \leq P(A) \leq 1$
6. For every two events A and B, $P(A \cap B^c) = P(A) - P(A \cap B)$
7. For every two events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
8. Bonferroni Inequality: For all events $A_1, \ldots, A_n$, $P(\bigcap_{i=1}^{n} A_i) \geq 1 - \sum_{i=1}^{n} P(A_i^c)$

1.6 Finite Sample Spaces

Simple Sample Space

- Has a finite number (n) of possible outcomes
- Each outcome has an equal probability $(\frac{1}{n})$
- If an event A has m outcomes, then $P(A) = \frac{m}{n}$

1.7 Counting Methods

Multiplication Rule: An experiment with k parts where the ith part has n_i possible outcomes (regardless of which specific outcomes have occurred in the other parts) has a sample space $S = n_1 n_2 \ldots n_k$

Permutations ($P_{n,k}$):

- Number of ways to arrange a set (order matters)
- Sampling considering n different items and making k choices from them
  - Sampling with replacement: n^k
  - Sampling without replacement: $n(n-1)\ldots(n-k+1)$
    * n options for first choice, $n-1$ options for second choice, $n-k+1$ options for kth choice

- The number of permutations of n different items is $P_{n,n} = n!$
- The number of permutations of n different items making k choices ($0 \leq k \leq n$) is

$$P_{n,k} = n(n-1)\dots(n-k+1)$$

$$P_{n,k} = n(n-1)\dots(n-k+1)\left(\frac{1}{1}\right)$$

$$P_{n,k} = n(n-1)\dots(n-k+1)\left(\frac{(n-k)(n-k-1)\dots 1}{(n-k)(n-k-1)\dots 1}\right)$$

$$P_{n,k} = \frac{n(n-1)\dots(n-k+1)(n-k)(n-k-1)\dots 1}{(n-k)(n-k-1)\dots 1}$$

$$P_{n,k} = \frac{n!}{(n-k)!}$$

1.8 Combinatorial Methods

Combinations ($C_{n,k}$):

- Number of subsets (order does not matter)
- Permutations may be thought of as combinations of size k chosen out of n, multiplied by the number of ways to arrange the size k subsets, $k!$. More formally, this says

$$P_{n,k} = C_{n,k}k!$$

- Combinations (binomial coefficient) are the number of distinct subsets of size k that can be chosen from a set of size n (this is the same formula as for permutations, except we are dividing out the number of ways we can rearrange the subsets, $k!$, since order does not matter):

$$C_{n,k} = \binom{n}{k} = \frac{P_{n,k}}{k!} = \frac{\frac{n!}{(n-k)!}}{k!} = \frac{n!}{(n-k)!k!}$$

- Combinations without replacement: $\binom{n+k-1}{k}$

1.9 Multinomial Coefficients

The total number of different ways of dividing n elements into k groups is

$$\binom{n}{n_1, n_2, \dots, n_k} = \binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-\cdots-n_{k-2}}{n_{k-1}} = \frac{n!}{n_1!n_2!\cdots n_k!}$$

1.10 The Probability of a Union of Events

- Union of two events A_1 and A_2:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

- Union of three events A_1, A_2, and A_3:

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - [P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3)]$$
$$+P(A_1 \cap A_2 \cap A_3)$$

- Union of n events $A_1, \ldots, A_n$:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) -$$
$$\sum_{i<j<k<l} P(A_i \cap A_j \cap A_k \cap A_l) + \cdots + (-1)^{n+1} P(A_1 \cap A_2 \cap \cdots \cap A_n)$$

2 Conditional Probability

2.1 The Definition of Conditional Probability

- Conditional probability is the updating of probabilities when certain events are observed
- The updated probability of event A after we learn that event B has occured is the conditional probability of A given B, denoted $P(A|B)$
- When we go from $P(A)$ to $P(A|B)$, we say we are conditioning on B
- For $P(B) > 0$
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
- Intuitively, this is saying the probability of A occurring, given B has occurred is equal to the outcomes where both A and B occurred (which makes sense since we know B has occurred and we want to find the probability of A also occurring) divided by the probability of B occurring (which makes sense since we know B occurred, we can renormalize the sample space to only contain outcomes where B occurred).

Multiplication Rule for Conditional Probabilities

- For 2 events A, B
  - If $P(B) > 0$
$$P(A \cap B) = P(B)P(A|B)$$
  - If $P(A) > 0$
$$P(A \cap B) = P(A)P(B|A)$$
- For n events $A_1, \ldots, A_n$ such that $P(A_1 \cap \ldots \cap A_{n-1}) > 0$
$$P(A_1 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap \ldots A_{n-1})$$
- For n events $A_1, \ldots, A_n, B$ such that $P(B) > 0$ and $P(A_1 \cap \ldots \cap A_{n-1}) > 0$
$$P(A_1 \cap \ldots \cap A_n|B) = P(A_1|B)P(A_2|A_1 \cap B)P(A_3|A_1 \cap A_2 \cap B) \ldots P(A_n|A_1 \cap \ldots A_{n-1} \cap B)$$

Law of Total Probability

- Tells us that to get the unconditional probability of A, we can divide the sample space into disjoint slices B_j, find the conditional probability of A within each of these slices, then take a weighted sum of the conditional probabilities, where the weights are the probabilities $P(B_j)$
- Often used in tandum with Bayes' Rule
- Relates conditional probability to unconditional probability
- Partition: Let S denote the sample space and consider k events $B_1, \ldots, B_k$ in S such that $B_1, \ldots, B_k$ are disjoint and $\bigcup_{i=1}^{k} B_i = S$. Then events $B_1, \ldots, B_k$ form a partition in S. In other words, only one of these events can occur and combined they fill the entire sample space
- Suppose events $B_1, \ldots, B_k$ form a partition of the space S and $P(B_j) > 0$ for $j = 1, \ldots, k$. Then
$$P(A) = \sum_{j=1}^{k} P(B_j \cap A) = \sum_{j=1}^{k} P(B_j)P(A|B_j)$$
- Conditional LOTP
$$P(A|C) = \sum_{j=1}^{k} P(B_j|C)P(A|B_j \cap C)$$

2.2 Independent Events

Independence of Two Events:

- Two events are independent if learning that one occurred does not change the probability of the other event.
- Two events A and B with positive probabilities are independent if
$$P(A \cap B) = P(A)P(B)$$
- Similarly, two events A and B with positive probabilities are independent if
$$P(A|B) = P(A) \text{ and } P(B|A) = P(B)$$
- If two events A and B are independent, then A and B^C are independent

Independence of k Events:

- For k events, if knowing what happened with any particular subset of the events gives us no information about what happened with the events not in the subset, the events are independent
- k events are (mutually) independent if
  - Any pair $P(A_i \cap A_j) = P(A_i)P(A_j)$ for $i \neq j$
  - Any triplet $P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k)$ for $i \neq j \neq k$
  - ...

- The n-tuplet $P(A_1 \cap \ldots \cap A_n) = P(A_1) \ldots P(A_n)$

When deciding whether or not to model events as independent, try to answer the following question: "If I were to learn that some of these events occurred, would I change the probabilities of any of the others?"

Conditional Independence:

- Two events A and B are conditionally independent given E if $P(A \cap B|E) = P(A|E)P(B|E)$
- It is defined as independence but with respect to the conditional probabilities

2.3 Bayes' Theorem

- Bayes' Theorem relates $P(A|B)$ to $P(B|A)$
- This is important as often it is easier to solve either $P(A|B)$ or $P(B|A)$
- Suppose that we are interested in which of several disjoint events $A_1, \ldots, A_k$ will occur and that we will get to observe some other event B. If $P(B|A_i)$ is available for each i, then Bayes' theorem is a useful formula for computing the conditional probabilities of the A_i events given B, that is, $P(A_i|B)$ for each i

Bayes' Theorem for 2 Events

- $$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$
- It is often common to use LOTP in the denominator
$$P(A|B) = \frac{P(B|A)P(A)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Bayes' Theorem for n Events

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$$

Prior vs. Posterior Probabilities

- Prior probabilities represent probabilities *before* new evidence is introduced.
  - $P(A)$
- Posteror probabilities represent probabilities *after* the new evidence in taken into consideration.
  - $P(A|B)$

Conditional Version of Bayes' Theorem

$$P(A_i|B \cap C) = \frac{P(B|A_i \cap C)P(A_i|C)}{\sum_{j=1}^{n} P(B|A_j \cap C)P(A_j|C)}$$