

Probability and Statistics

Ash Bellett

Contents

1	Set Theory	1
1.1	Sets	1
1.1.1	Notation	1
1.1.2	Membership	1
1.1.3	Empty Set	1
1.1.4	Subsets	1
1.1.5	Partitions	1
1.1.6	Power Sets	1
1.1.7	Cardinality	1
1.2	Operations	2
1.2.1	Absolute Complement	2
1.2.2	Union	2
1.2.3	Intersection	2
1.2.4	Relative Complement	2
1.2.5	Symmetric Difference	2
1.2.6	Cartesian Product	2
1.3	Properties	2
1.3.1	Commutativity	2
1.3.2	Associativity	3
1.3.3	Distributivity	3
1.3.4	De Morgan's Laws	3
2	Counting Techniques	4
2.1	Product Rule of Counting	4
2.2	Permutations	4
2.2.1	Permutations with repetition	4
2.2.2	Permutations without repetition	4
2.2.3	Permutations of non-distinct objects	5
2.3	Combinations	5
2.3.1	Combinations without repetition	5
2.3.2	Combinations with repetition	5
2.4	Binomial Theorem	5
2.5	Multinomial Theorem	5
3	Probability Theory	6
3.1	Probabilistic Experiments	6
3.1.1	Experiments	6
3.1.2	Trials	6
3.1.3	Outcomes	6
3.1.4	Sample space	6
3.1.5	Events	6
3.1.6	Probability space	7
3.2	Axioms	7
3.2.1	First axiom: non-negative, real	7

3.2.2	Second axiom: unitarity	8
3.2.3	Third axiom: countable additivity	8
3.3	Theorems	8
3.3.1	Probability of an empty set	8
3.3.2	Additivity	8
3.3.3	Monotonicity	9
3.3.4	Complement rule	9
3.3.5	Numeric bounds	10
3.3.6	Sum rule	10
3.3.7	Probability from elementary events	11
3.3.8	Equally-Likely Outcomes	11
3.4	Limits	11
4	Conditional Probability and Independence	12
4.1	Conditional Probability	12
4.2	Independent Events	13
4.3	Law of Total Probability	13
4.4	Bayes' Theorem	13
5	Univariate Random Variables	14
5.1	Probability Distributions	14
5.2	Random Variables	14
5.3	Discrete Random Variables	14
5.4	Continuous Random Variables	14
5.5	Distribution Parameters	14
6	Moments of Univariate Random Variables	15
6.1	Measures of Centrality	15
6.2	Measures of Variability	15
6.3	Transformations of Centrality and Variability	15
7	Bivariate Random Variables	16
8	Product Moments of Bivariate Random Variables	16
8.1	Covariance	16
8.2	Independence	16
8.3	Correlation	16
8.4	Periodicity	16
8.5	Moment Generating Functions	16
9	Vectors of Random Variables	16
10	Functions of Random Variables	17
11	Sequences of Random Variables	18
12	Sampling	19

13 Estimation	20
13.1 Point Estimators	20
13.2 Interval Estimators	20
14 Hypothesis Testing	21
15 Variance Analysis	22
16 Goodness-of-Fit Testing	23

1 Set Theory

1.1 Sets

A set is an unordered collection of distinct objects. Objects that belong to a set are called the elements of that set.

1.1.1 Notation

Sets can be represented using enumeration where the collection of objects is listed within curly brackets: $A = \{0, 1, \dots\}$. Sets can also be represented using conditions with set-builder notation: $A = \{x \mid x \geq 0\}$. The vertical bar in set-builder notation means "such that" and the right-hand side is a condition to define the elements in the set.

1.1.2 Membership

If an object a is an element of the set A then a is in A : $a \in A$. If an object a is not an element of the set A then a is not in A : $a \notin A$.

1.1.3 Empty Set

If a set has no elements, it is an empty set \emptyset .

1.1.4 Subsets

If every element of a set A is an element of another set B then A is a subset of B : $A \subseteq B$. If every element of a set A is an element of another set B , and every element of B is an element of A , then A and B are equal: $A = B$. If every element of a set A is an element of another set B but not equal to B , then A is a proper subset of B : $A \subset B$.

1.1.5 Partitions

A partition of a set is a set of non-empty subsets of the set such that every element of the set is in exactly one of the subsets. The subsets are pairwise disjoint which means that no two subsets contain the same element.

1.1.6 Power Sets

The power set of a set is the set of all subsets of the set. The power set contains both the set itself and the empty set since both of these sets are subsets of the set. The power set of a finite set with n element has 2^n elements.

1.1.7 Cardinality

The cardinality of a set A is the number of elements in the set: $|A|$. The cardinality of an empty set is zero.

1.2 Operations

1.2.1 Absolute Complement

The absolute complement \overline{A} of a set A is the set of elements that are not in A .

$$\overline{A} = \{x \mid x \notin A\}$$

1.2.2 Union

The union of sets $A \cup B$ is the set of elements in A , B or both.

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$$

1.2.3 Intersection

The intersection of sets $A \cap B$ is the set of elements in both A and B .

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$$

1.2.4 Relative Complement

The relative complement of set A in set B is the set of elements in B that are not in A .

$$B \setminus A = \{x \mid x \in B \text{ and } x \notin A\}$$

1.2.5 Symmetric Difference

The symmetric difference of sets A and B is the set of elements in A or B and not in A and B .

$$A \oplus B = \{x \mid x \in A \text{ and } x \in B \text{ and } x \notin A \cup B\}$$

1.2.6 Cartesian Product

The Cartesian product of sets $A \times B$ is the set of all ordered pairs (a, b) where a is in A and b is in B .

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}$$

1.3 Properties

1.3.1 Commutativity

The union and intersection of sets is commutative.

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

1.3.2 Associativity

The union and intersection of sets is associative.

$$(A \cup B) \cup C = A \cup (B \cup C)$$

$$(A \cap B) \cap C = A \cap (B \cap C)$$

1.3.3 Distributivity

The union and intersection of sets is distributive.

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

1.3.4 De Morgan's Laws

The union and intersection of sets follow De Morgan's laws.

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

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

2 Counting Techniques

2.1 Product Rule of Counting

If an experiment E consists of k experiments E_1, E_2, \dots, E_k performed sequentially where each experiment E_i has n_i possible outcomes, then experiment E will have $\prod_{i=1}^k n_i$ possible outcomes.

2.2 Permutations

A permutation is an ordered subset.

2.2.1 Permutations with repetition

The number of permutations containing k of n distinct objects with repetition is n^k .

Derivation: Given a set N of n distinct objects, a new set K of $k \leq n$ ordered objects is to be constructed. The new set K is constructed by selecting an object from N . The object that is selected remains in N for subsequent selections. The selection process is repeated k times, where each selection is taken from n possible objects. The number of permutations of K is n^k . When the size of K is equal to N , $k = n$, the number of permutations is n^n .

2.2.2 Permutations without repetition

The number of permutations containing k of n distinct objects without repetition is:

$${}_n P_r = \frac{n!}{(n-k)!}$$

Derivation: Given a set N of n distinct objects, a new set K of $k \leq n$ ordered objects is to be constructed. The new set K is constructed by selecting an object from N . The object that is selected is removed from N for subsequent selections. The selection process is repeated k times. On the first selection, there are n possible objects to select from. On the second selection, there are $n - 1$ possible objects to select from. On the k^{th} selection, there are $n - k + 1$ possible objects to select from. The number of permutations of K is:

$$n(n-1)\dots(n-k+1)$$

Alternatively, this expression can be written as:

$$\begin{aligned} & \frac{1(2)\dots(n)}{1(2)\dots(n-k)} \\ &= \frac{n!}{(n-k)!} \end{aligned}$$

When the size of K is equal to N , $k = n$, the number of permutations is $n!$.

2.2.3 Permutations of non-distinct objects

Given a set N of n non-distinct objects, where there are m groups of distinct objects and each group i has n_i objects.

$$n = \sum_{i=1}^m n_i$$

A new set K of n ordered objects is to be constructed. The new set K is constructed by selecting an object from N . The object that is selected is removed from N for subsequent selections. The selection process is repeated n times. The number of permutations of K is $\frac{n!}{n_1!n_2!\dots n_m!}$.

2.3 Combinations

A combination is an unordered subset.

2.3.1 Combinations without repetition

The number of combinations containing k of n distinct objects without repetition is:

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

2.3.2 Combinations with repetition

The number of combinations containing k of n distinct objects with repetition is:

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{(n-1)!k!}$$

2.4 Binomial Theorem

2.5 Multinomial Theorem

3 Probability Theory

3.1 Probabilistic Experiments

3.1.1 Experiments

An experiment is a process of observation where the output cannot be predicted with certainty due to random effects.

3.1.2 Trials

A trial is a single occurrence of an experiment. Multiple trials of an experiment can form a new experiment.

3.1.3 Outcomes

An outcome is an observed output of a trial.

3.1.4 Sample space

The sample space Ω is the set of all possible outcomes of an experiment.

3.1.5 Events

An event A is a subset of outcomes in a sample space Ω .

$$A \subseteq \Omega$$

The event space \mathcal{F} is the set of all possible events.

$$A \in \mathcal{F}$$

The operations of sets apply to events:

- The absolute complement \overline{A} of an event A is the set of outcomes in the sample space Ω that are not in A .

$$\overline{A} = \{x \mid x \in \Omega \text{ and } x \notin A\}$$

- The union of events $A \cup B$ is the set of outcomes in A , B or both.

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$$

- The intersection of events $A \cap B$ is the set of outcomes in both A and B .

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$$

- The relative complement of event A in event B is the set of outcomes in B that are not in A .

$$B \setminus A = \{x \mid x \in B \text{ and } x \notin A\}$$

The properties of sets apply to events:

- The union and intersection of events is commutative.

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

- The union and intersection of events is associative.

$$(A \cup B) \cup C = A \cup (B \cup C)$$

$$(A \cap B) \cap C = A \cap (B \cap C)$$

- The union and intersection of events is distributive.

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

- The union and intersection of events follow De Morgan's laws.

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

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

The event consisting of no outcomes is called the null event \emptyset . If events A and B have no outcomes in common then A and B are disjoint events (mutually exclusive).

$$A \cap B = \emptyset$$

The event consisting of a single outcome is called an elementary event.

3.1.6 Probability space

A probability space (Ω, \mathcal{F}, P) describes the characteristics of an experiment: the sample space Ω , the event space \mathcal{F} and the probability measure P . A probability measure $P : \mathcal{F} \rightarrow \mathbb{R}$ is a function that assigns each event in the event space to a real number. A probability measure must follow the axioms of probability.

3.2 Axioms

3.2.1 First axiom: non-negative, real

The probability of an event is a non-negative real number.

$$P(A) \geq 0 \quad \forall A \in \mathcal{F}$$

This axiom means that the smallest probability of an event is 0 (impossible events). It does not specify an upper bound, however a theorem does.

3.2.2 Second axiom: unitarity

The probability that at least one outcome in the sample space will occur is 1.

$$P(\Omega) = 1$$

This axiom means that it is certain that an outcome will occur from observing an experiment.

3.2.3 Third axiom: countable additivity

If A_1, A_2, \dots is an infinite set of disjoint events in a sample space Ω :

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

This axiom forms a relationship between a set of disjoint events in a sample space and the individual probabilities of each event.

3.3 Theorems

3.3.1 Probability of an empty set

The probability of the null event is 0.

$$P(\emptyset) = 0$$

Proof: Let an infinite set of events be $\{A_i = \emptyset\}_{i=1}^{\infty}$. Since $\bigcup_{i=1}^{\infty} \emptyset = \emptyset$ and substituting A_i into the third axiom:

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} \emptyset\right) &= \sum_{i=1}^{\infty} P(\emptyset) \\ P(\emptyset) &= \sum_{i=1}^{\infty} P(\emptyset) \end{aligned}$$

The only solution to this equation is $P(\emptyset) = 0$.

3.3.2 Additivity

Given a finite set of n disjoint events A_1, A_2, \dots, A_n in a sample space Ω :

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

Proof: Let an infinite set of disjoint events be $E_1 = A_1, E_2 = A_2, \dots, E_n = A_n, E_{n+1} = \emptyset, E_{n+2} = \emptyset, \dots$. Substituting this into the third axiom:

$$\begin{aligned}
 P\left(\bigcup_{i=1}^{\infty} E_i\right) &= \sum_{i=1}^{\infty} P(E_i) \\
 &= \sum_{i=1}^n P(E_i) + \sum_{i=n+1}^{\infty} P(E_i) \\
 &= \sum_{i=1}^n P(E_i) + \sum_{i=n+1}^{\infty} P(\emptyset) \\
 &= \sum_{i=1}^n P(E_i) + \sum_{i=n+1}^{\infty} 0 \\
 &= \sum_{i=1}^n P(E_i) \\
 P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i)
 \end{aligned}$$

3.3.3 Monotonicity

If an event A is a subset of or equal to another event B which is a subset of or equal to a sample space Ω , then the probability of A occurring is less than or equal to the probability of B occurring.

$$\text{If } A \subseteq B \subseteq \Omega \text{ then } P(A) \leq P(B)$$

Proof: Let event A be a subset of or equal to event B . The event $B \setminus A$ is the set difference of B and A and is the set of outcomes in B that are not in A . The union of A and $B \setminus A$ is equal to B .

$$B = A \cup (B \setminus A)$$

Taking probabilities of both sides:

$$P(B) = P(A \cup (B \setminus A))$$

From the probability theorem of additivity where $n = 2$:

$$P(B) = P(A) + P(B \setminus A)$$

From the first axiom, $P(B \setminus A) \geq 0$. Therefore, $P(A) \leq P(B)$.

3.3.4 Complement rule

If A is an event in a sample space Ω , then the probability of the complement of A is given by:

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

Proof: Let event A be a subset of or equal to a sample space Ω . Then $S = A \cup \bar{A}$ and A and \bar{A} are disjoint events. From the first axiom:

$$P(\Omega) = P(A \cup \bar{A}) = 1$$

From the third axiom:

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}) = 1$$

Rearranging to make $P(\bar{A})$ the subject:

$$P(\bar{A}) = 1 - P(A)$$

3.3.5 Numeric bounds

If A is an event in a sample space Ω , then the probability of A is bounded between 0 and 1.

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

Proof: From the first axiom and the complement rule:

$$1 - P(A) \geq 0$$

Rearranging to make $P(A)$ the subject:

$$P(A) \leq 1$$

From the first axiom:

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

3.3.6 Sum rule

If A and B are events in a sample space Ω , the probability that either A or B will occur is the sum of the probabilities that A will occur and that B will occur minus the probability that both A and B will occur.

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

Proof: Let events A and B be a subset of or equal to a sample space Ω . The probability of A or B can be expressed as:

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

Making the substitution $P(B \setminus A) = P(B) - P(A \cap B)$:

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

3.3.7 Probability from elementary events

The probability of an event A in a sample space Ω is equal to the sum of the probabilities of its elementary events $\{E_i\}$.

$$P(A) = \sum_{i=1}^{\infty} P(E_i)$$

Proof: Any event A in a sample space Ω can be expressed as the union of its elementary events $\{E_i\}$.

$$A = \bigcup_{i=1}^{\infty} E_i$$

Substituting into the third axiom:

$$\begin{aligned} P(A) &= P\left(\bigcup_{i=1}^{\infty} E_i\right) \\ &= \sum_{i=1}^{\infty} P(E_i) \end{aligned}$$

3.3.8 Equally-Likely Outcomes

3.4 Limits

4 Conditional Probability and Independence

4.1 Conditional Probability

The conditional probability of an event A occurring given that event B has occurred is:

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

The conditional probability $P(A | B)$ is a new probability function on the sample space Ω such that outcomes not in B have zero probability and the probability of outcomes in B are scaled such that their relative magnitudes are preserved and the probability measure is consistent with the axioms of probability.

Derivation: Let Ω be a sample space with elementary events $\{E\}$. The event $B \subseteq \Omega$ has occurred. New probabilities are to be assigned to the set of elementary events $\{E\}$. The elementary events $E \in \overline{B}$ will have zero probability as B has occurred. The probability of elementary events $E \in B$ will preserve their relative magnitudes, represented by a scaling factor α .

$$P(E | B) = \alpha P(E) \quad \forall E \in B$$

$$P(E | B) = 0 \quad \forall E \in \overline{B}$$

Substituting into the third axiom:

$$\begin{aligned} 1 &= \sum_{E \in \Omega} P(E | B) \\ 1 &= \sum_{E \in B} P(E | B) + \sum_{E \in \overline{B}} P(E | B) \\ 1 &= \alpha \sum_{E \in B} P(E) \end{aligned}$$

Using the probability from elementary events theorem:

$$\begin{aligned} 1 &= \alpha P(B) \\ \Rightarrow \alpha &= \frac{1}{P(B)} \end{aligned}$$

The conditional probability of an elementary event given that event B has occurred $P(E | B)$ is given by:

$$P(E | B) = \frac{P(E)}{P(B)} \quad \forall E \in B$$

$$P(E | B) = 0 \quad \forall E \in \overline{B}$$

An event $A \subseteq \Omega$ is comprised of $A \cap B$ and $A \cap \overline{B}$.

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

The conditional probability of A given that B has occurred can be expressed as:

$$\begin{aligned} P(A | B) &= \sum_{E \in A \cap B} P(E | B) + \sum_{E \in A \cap \overline{B}} P(E | B) \\ &= \sum_{E \in A \cap B} \frac{P(E)}{P(B)} \end{aligned}$$

Using the probability from elementary events theorem:

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

4.2 Independent Events

Events A and B in a sample space Ω are independent if:

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

4.3 Law of Total Probability

4.4 Bayes' Theorem

5 Univariate Random Variables

5.1 Probability Distributions

5.2 Random Variables

5.3 Discrete Random Variables

5.4 Continuous Random Variables

5.5 Distribution Parameters

6 Moments of Univariate Random Variables

6.1 Measures of Centrality

6.2 Measures of Variability

6.3 Transformations of Centrality and Variability

7 Bivariate Random Variables

8 Product Moments of Bivariate Random Variables

8.1 Covariance

8.2 Independence

8.3 Correlation

8.4 Periodicity

8.5 Moment Generating Functions

9 Vectors of Random Variables

10 Functions of Random Variables

11 Sequences of Random Variables

12 Sampling

13 Estimation

13.1 Point Estimators

13.2 Interval Estimators

14 Hypothesis Testing

15 Variance Analysis

16 Goodness-of-Fit Testing