Probability and Statistics

Ash Bellett

Contents

1 Probability Theory

1.1 Probabilistic Experiments

1.1.1 Experiments

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

1.1.2 Trials

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

1.1.3 Outcomes

An outcome is an observed output of a trial.

1.1.4 Sample space

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

1.1.5 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 axioms of set theory 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.

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

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

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

$$A \cap B = \varnothing$$

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

1.1.6 Probability space

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

1.2 Axioms

1.2.1 First axiom: non-negative, real

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

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

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.

1.2.2 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.

1.2.3 Third axiom: countable additivity

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

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

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

1.3 Theorems

1.3.1 Probability of an empty set

The probability of the null event is 0.

$$P(\varnothing) = 0$$

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

$$P(\bigcup_{i=1}^{\infty} \varnothing) = \sum_{i=1}^{\infty} P(\varnothing)$$

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

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

1.3.2 Additivity

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

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

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

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

1.3.3 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, $P(A) \leq P(B)$.

1.3.4 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 $S = A \cup \overline{A}$ and A and $\overline{A}$ are disjoint events. From the first axiom:

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

From the third axiom:

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

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

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

1.3.5 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:

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

1.3.6 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 $P(B \setminus A) = P(B) - P(A \cap B)$:

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

1.3.7 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 $\{E_i\}$.

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

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:

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

2 Counting

2.1 Product Rule

If an experiment E consists of k experiments $E_1, E_2, ..., E_k$ 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 a 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:

$$n(n-1)...(n-k+1)$$

Alternatively, this expression can be written as:

$$\frac{1(2)...(n)}{1(2)...(n-k)}$$

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

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

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 $\frac{n!}{n_1!n_2!...n_m!}$.

2.4 Combinations

A combination is an unordered subset.

2.4.1 Combinations without repetition

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

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

2.4.2 Combinations with repetition

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

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

3 Conditional Probability and Independence

3.1 Conditional Probability

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

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

The conditional probability $P(A \mid B)$ 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.

$$P(E \mid B) = \alpha P(E) \ \forall E \in B$$

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

Substituting into the third axiom:

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

Using the probability from elementary events theorem:

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

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

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

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

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:

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

Using the probability from elementary events theorem:

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

3.2 Independent Events

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

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

3.3 Bayes' Theorem

4 Univariate Random Variables

4.1 Probability Distributions

4.2 Random Variables

4.3 Discrete Random Variables

4.4 Continuous Random Variables

4.5 Distribution Parameters

5 Moments of Univariate Random Variables

5.1 Measures of Centrality

5.2 Measures of Variability

5.3 Transformations of Centrality and Variability

6 Bivariate Random Variables

7 Product Moments of Bivariate Random Variables

7.1 Covariance

7.2 Independence

7.3 Correlation

7.4 Periodicity

7.5 Moment Generating Functions

8 Functions of Random Variables

9 Sequences of Random Variables

10 Sampling

11 Estimation

11.1 Point Estimators

11.2 Interval Estimators

12 Hypothesis Testing

13 Variance Analysis

14 Goodness-of-Fit Testing