

Probability vs. Statistics

Probability and statistics are deeply connected because all statistical statements are at bottom statements about probability. Despite this the two sometimes feel like very different subjects. Probability is logically self-contained; there are a few rules and answers all follow logically from the rules, though computations can be tricky. In statistics we apply probability to draw conclusions from data. This can be messy and usually involves as much art as science.

Probability example

You have a fair coin (equal probability of heads or tails). You will toss it 100 times. What is the probability of 60 or more heads? There is only one answer (about 0.028444) and you must know how to compute it.

Statistics example

You have a coin of unknown origin. To investigate whether it is fair you toss it 100 times and count the number of heads. Let's say you count 60 heads. Your job as a statistician is to draw a conclusion (inference) from this data. There are many ways to proceed, both in terms of the form the conclusion takes and the probability computations used to justify the conclusion. In fact, different statisticians might draw different conclusions.

Note that in the first example the random process is fully known (probability of heads = .5). The objective is to find the probability of a certain outcome (at least 60 heads) arising from the random process. In the second example, the outcome is known (60 heads) and the objective is to illuminate the unknown random process (the probability of heads).

Frequentist vs. Bayesian Interpretations

There are two prominent and sometimes conflicting schools of statistics: **Bayesian** and **frequentist**. Their approaches are rooted in differing interpretations of the meaning of probability.

Frequentists say that probability measures the frequency of various outcomes of an experiment. For example, saying a fair coin has a 50% probability of heads means that if we toss it many times then we expect about half the tosses to land heads.

Bayesians say that probability is an abstract concept that measures a state of knowledge or a degree of belief in a given proposition. In practice Bayesians do not assign a single value for the probability of a coin coming up heads. Rather they consider a range of values each with its own probability of being true.

The frequentist approach has long been dominant in fields like biology, medicine, public health and social sciences.

The Bayesian approach has enjoyed a resurgence in the era of powerful computers and big data. It is especially useful when incorporating new data into an existing statistical model, for example,

when training a speech or face recognition system. Today, statisticians are creating powerful tools by using both approaches in complementary ways.

Counting and Sets

Learning Goals

1. Know the definitions and notation for sets, intersection, union, complement.
2. Understand how counting is used computing probabilities.
3. Be able to use the rule of product, inclusion-exclusion principle, permutations and combinations to count the elements in a set.

Counting

Question 1 - A coin is fair if it comes up heads or tails with equal probability. You flip a fair coin three times. What is the probability that exactly one of the flips results in a head?

your answer here

Sets and notation

Definitions

A **set** S is a collection of elements. We use the following notation.

Element: We write $x \in S$ to mean the element x is in the set S .

Subset: We say the set A is a subset of S if all of its elements are in S . We write this as $A \subset S$.

Complement: The complement of A in S is the set of elements of S that are not in A . We write this as A^c or $S - A$.

Union: The union of A and B is the set of all elements in A or B (or both). We write this as $A \cup B$.

Intersection: The intersection of A and B is the set of all elements in both A and B . We write this as $A \cap B$.

Empty set: The empty set is the set with no elements. We denote it \emptyset .

Disjoint: A and B are disjoint if they have no common elements. That is, if $A \cap B = \emptyset$.

Difference: The difference of A and B is the set of elements in A that are not in B . We write this as $A - B$.

Let's illustrate these operations for specific example.

Question 2. Start with a set of 10 animals $S = \{ \text{Antelope, Bee, Cat, Dog, Elephant, Frog, Gnat, Hyena, Iguana, Jaguar} \}$. Consider two subsets: $M = \text{the animal is a mammal} = \{ \text{Antelope, Cat, Dog, Elephant, Hyena, Jaguar} \}$ $W = \text{the animal lives in the wild} = \{ \text{Antelope, Bee, Elephant, Frog, Gnat, Hyena, Iguana, Jaguar} \}$ **Answer 2** Our goal here is to look at different set operations. Give answers to the following:

1. Intersection: $M \cap W$
2. Union: $M \cup W$
3. Complement: M^c
4. Difference: $M - W$

Question 3 What are the Venn Diagrams?

Answer

Products of sets

The product of sets S and T is the set of ordered pairs:

$$S \times T = \{(s, t) \mid s \in S, t \in T\} \quad (1)$$

In words the right-hand side reads "the set of ordered pairs (s, t) such that s is in S and t is in T "

The following diagrams show two examples of the set product.

Counting

If S is finite, we use $|S|$ or $\#S$ to denote the number of elements of S . Two useful counting principles are the inclusion-exclusion principle and the rule of product.

Inclusion-exclusion principle

The inclusion-exclusion principle says

$$|A \cup B| = |A| + |B| - |A \cap B| \quad (2)$$

Question 4 In a band of singers and guitarists, seven people sing, four play the guitar, and two do both. How big is the band?

your answer here

Rule of Product

The Rule of Product says: If there are n ways to perform action 1 and then by m ways to perform action 2, then there are $n \cdot m$ ways to perform action 1 followed by action 2 We will also call this the multiplication rule.

Example. If you have 3 shirts and 4 pants then you can make $3 \cdot 4 = 12$ outfits.

Think: An extremely important point is that the rule of product holds even if the ways to perform action 2 depend on action 1, as long as the number of ways to perform action 2 is independent of action 1. To illustrate this answer the following

Question 5 There are 5 competitors in the 100m final at the Olympics. In how many ways can the gold, silver, and bronze medals be awarded?

your answer here:

Note that the choice of gold medalist affects who can win the silver, but the number of possible

Permutations and combinations

Permutations

A permutation of a set is a particular ordering of its elements. For example, the set $\{a, b, c\}$ has six permutations: $abc, acb, bac, bca, cab, cba$. We found the number of permutations by listing them all. We could also have found the number of permutations by using the rule of product. That is, there are 3 ways to pick the first element, then 2 ways for the second, and 1 for the first. This gives a total of $3 \cdot 2 \cdot 1 = 6$ permutations.

In general, the rule of product tells us that the number of permutations of a set of k elements is

$$k! = k \cdot (k - 1) \cdots 3 \cdot 2 \cdot 1 \quad (3)$$

We also talk about the permutations of k things out of a set of n things.

Question 6 List all the permutations of 3 elements out of the set $\{a, b, c, d\}$ in markdown language. Learn how to specify tables in markdown.

your answer here

Formulas

We'll use the following notations.

${}_nP_k$ = number of permutations (lists) of k distinct elements from a set of size n

${}_nC_k = \binom{n}{k}$ = number of combinations (subsets) of k elements from a set of size n We

emphasise that by the number of combinations of k elements we mean the number of subsets of size k .

These have the following notation and formulas:

$$\text{Permutations: } {}_nP_k = \frac{n!}{(n-k)!} = n(n-1) \cdots (n-k+1) \quad (4)$$

$$\text{Combinations: } {}_nC_k = \frac{n!}{k!(n-k)!} = \frac{{}_nP_k}{k!}$$

The notation ${}_nC_k$ is read " n choose k ". The formula for ${}_nP_k$ follows from the rule of product. It

also implies the formula for ${}_nC_k$ because a subset of size k can be ordered in $k!$ ways.

We can illustrate the relation between permutations and combinations by lining up the results of the previous two examples.

abc	acb	bac	bca	cab	cba	$\{a, b, c\}$
abd	adb	bad	bda	dab	dba	$\{a, b, d\}$
acd	adc	cad	cda	dac	dca	$\{a, c, d\}$
bcd	bdc	cbd	cdb	dbc	dcb	$\{b, c, d\}$

Permutations: $4P_3$ Combinations: $4C_3$

Notice that each row in the permutations list consists of all $3!$ permutations of the corresponding set in the combinations list

Questions - Examples

Question 7. Count the following:

- (i) The number of ways to choose 2 out of 4 things (order does not matter).
- (ii) The number of ways to list 2 out of 4 things.
- (iii) The number of ways to choose 3 out of 10 things.

your answers in markdown

Question 8

- (i) Count the number of ways to get 3 heads in a sequence of 10 flips of a coin
- (ii) If the coin is fair, what is the probability of exactly 3 heads in 10 flips.

your answers here