Probability

Module 3

Statistics 251: Statistical Methods

Updated 2020

Definitions

probability: likelihood of an event in an experiment ocurring; like the chance of precipitation tomorrow. There are personal probabilities (subjective) and empirical probabilitites (objective)

outcome: result of experiment

event: any combination of outcomes, can be a collection of one or more outcomes; most often mathematically represented with capital (English) letters, like A, B, etc.

probability of event A: denoted as $P(A)$, meaning the chance that event A occurs

sample space (S): the entire collection of events from an experiment; the set of all possible outcomes of an experiment

trial: every instance of an experiment; the total number of trials $= n$

Events logistics II

sampling with replacement (swr): when a member of a population is chosen, it is then replaced back into the population and has another chance to be chosen for the sample; probabilities will not change for the second pick, in other words, events are considered to be independent

sampling without replacement (swor): when a member of a population is chosen, it is not replaced back into the population and no longer has another chance to be chosen for the sample; probabilities will change for the second pick since there are 1 fewer elements to choose from each time one is chosen, the events are considered to be dependent (not indpendent)

Probability Rules

The first ones

(1) $0 \leq P(A) \leq 1$, probabilities must be between 0 and 1 (0% and 100%)

(2) $\sum P(A_i) = 1 = S$, the sum of the probabilities for an experiment must sum to 1 (the sample space)

(3) $P(A') = 1 - P(A)$, the complement of the probability of event A is 1 minus the probability of A; it is whatever is not in A.
   - The symbol is "A-prime", is stated as "A-complement" or "A-not" (complement, a number or quantity of something required to make a group complete, not to be confused with compliment, to give praise)

More to add to the list…

Intersections and Unions

union: the union of two events, A and B, is all of the outcomes from A *or* B. The symbol is a $\cup$ and the key word to watch for is *or*; the probability of the union of two events A and B is $P(A \cup B)$, or $P(A \text{ or } B)$

intersection: the intersection of two events, A and B, is the set of events that A and B have in common (where they overlap). The symbol is a $\cap$ and the key word to watch for is *and*; the probability of the intersection of two events A and B is $P(A \cap B)$, or $P(A\ \textit{and}\ B)$

Example

Let $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $A = \{1, 2, 3, 4, 5\}$, and $B = \{4, 5, 6, 7, 8\}$.

union:

$$A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

with {4,5} not listed twice (no duplicates)

intersection:

$$A \cap B = \{4, 5\}$$

Events logistics I

mutually exclusive/disjoint: when there is no intersection between events, $P(A \cap B) = \varnothing$ (the intersection is an empty set, i.e. it does not exist). Events A and B cannot happen at the same time if they are disjoint (mutually exclusive). It is like being in two places at the same time (it cannot happen)

independent: If two events, A and B are independent, then the outcome of one event has no impact on the outcome of the other event; independence of events *cannot be assumed at your convenience*

Disjoint events $\neq$ independent events

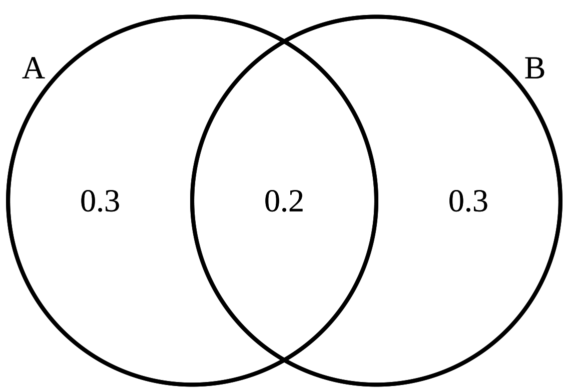
Venn diagram I

A Venn diagram is a graph (picture) that represents the events of an experiment. It usually is a box that represents the sample space S and has circles that represent the events (A, B, etc.)

Venn diagram II

This one shows two events, A and B with an intersection.

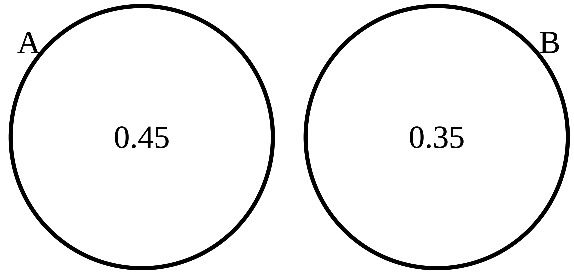
$$P(A) = 0.5,\ P(B) = 0.5,\ P(A \cap B) = 0.2$$



Venn diagram III

This one shows two events, A and B for disjoint/mutually exclusive events.

$$P(A) = 0.45,\ P(B) = 0.35,\ P(A \cap B) = \varnothing$$



Addition rule

(4) Addition rule
The probability of the *union* of A and B is:

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

This formula can be modified if the union of one or both complements is required; if the union is given, then the intersection can be solved for. If two events are mutually exclusive, then the intersection is an empty set (it does not exist), the intersection is the sum of the probabilities of the two events. $P(A \cup B) = P(A) + P(B)$

Multiplication Rule

(5) Multiplication rule (**THIS RULE IS FOR INDEPENDENT EVENTS ONLY!!!**)

If two events, A and B are independent, the *intersection* of A and B can be calculated as:

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

This equation can also be used to prove or disprove independence

Again, it cannot be assumed, you have to either be told they are independent or prove it mathematically with the above formula.

Rules recap

(1) Valid probability rule: $0 \leq P(A) \leq 1$
(2) Total probability rule: $\sum P(A_i) = 1 = S$
(3) Complement rule: $P(A') = 1 - P(A)$
(4) Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(5) Multiplication rule: $P(A \cap B) = P(A)P(B)$ (independent events only)

Pay attention to the formulas. They can be modified for finding complements, as well as solving for unknown values with a bit of algebra

Intersections and unions

Use addition rule for intersections (modification of the formulas is necessary to solve for values concerning complements)

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

When intersections are needed and you cannot assume independence, then use of the addition rule can be used and solved for the intersection.

$P(A) = 0.5,\ P(B) = 0.5\ P(A \cap B) = 0.2$

$P(A') = 1 - P(A) = 1 - 0.5 = 0.5$ or 50%

$P(B') = 1 - P(B) = 1 - 0.5 = 0.5$ or 50%

$P(A \cup B) = 0.5 + 0.5 - 0.2 = 0.8$

Note that $P(A \cup B) = P(A \cap B)' = 1 - P(A \cap B) = 1 - 0.2 = 0.8$

Confusion Matrix

One way to help calculating probabilities of intersections, many time for use in other calculations

(1) Create a 2X2 table, including row and column headers and totals

(2) Label the rows with $P(A), P(A')$ and the columns with $P(B), P(B')$

(3) The row totals are the values of $P(A), P(A')$ and the columns totals are $P(B), P(B')$

(4) The grand total (bottom right corner cell) is always 1 ($\sum p(x_i) = 1$)

(5) The upper left cell of the matrix (table) is $P(A \cap B)$; it has to be given or calculated

(6) Solve for the other 3 cells of the matrix by way of row and column totals

Matrix setup

Matrix example

Suppose that $P(A) = 0.5,\ P(B) = 0.3,\ P(A \cap B) = 0.2$

$P(A \cap B') = 0.5 - 0.2 = 0.3$
$P(A' \cap B) = 0.3 - 0.2 = 0.1$
$P(A' \cap B') = 0.5 - 0.1 = 0.4$ or $= 0.7 - 0.3 = 0.4$

More matrix

Are A and B independent? Use $P(A \cap B) = P(A)P(B)$ to prove it.

$$P(A \cap B)? =? P(A)P(B) \Rightarrow 0.2 \neq (0.5)(0.5)$$

Since the statement becomes false, events A and B are not independent (they are dependent).

Example of flipping two fair coins

When you flip a fair coin (one that is not weighted so that one side is more likely to come up than the other), the chance that it is a head is 50% and the chance that it is a tail is 50%; either side is equally likely.

List out all possible combinations

2 heads: (H,H); 1 head: (H,T), (T,H); 0 heads: (T,T)
$P(H) = 0.5, P(T) = 0.5$: flipping coins are independent events

2 heads: $P(H,H) = P(H1^{st} \cap H2^{nd}) = P(H1)P(H2) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$

1 head: $P(H,T) = P[(H1^{st} \cap T2^{nd}) \cup (T1^{st} \cap H2^{nd})] = P(H1)P(T2) + P(T1)P(H2) = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \left(\frac{1}{2}\right)$

0 heads: $P(T,T) = P(T1^{st} \cap T2^{nd}) = P(T1)P(T2) = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$

Probability distribution of number of heads

Example of the sum of two 6-sided dice

When you roll a fair 6-sided die (one that is not weighted so that one side is more likely to come up than the others), the chance that it is a 1 is one-sixth, 2 is one-sixth, etc.; all sides are equally likely.

List out all possible combinations

sum=2: (1,1)
sum=3: (1,2), (2,1)
sum=4: (1,3), (3,1), (2,2)
sum=5: (1,4), (4,1), (2,3), (3,2)
sum=6: (1,5), (5,1), (2,4), (4,2), (3,3)
sum=7: (1,6), (6,1), (2,5), (5,2), (3,4), (4,3)
sum=8: (2,6), (6,2), (3,5), (5,3), (4,4)
sum=9: (3,6), (6,3), (4,5), (5,4)
sum=10: (4,6), (6,4), (5,5)
sum=11: (5,6), (6,5)
sum=12: (6,6)

Probability distribution of sum of two 6-sided dice

Contingency tables

Table displays sample values in relation to two different variables that may be dependent (contingent) on one another. You can display with counts or relative requencies (probabilities)

Cell counts or probabilities are intersections, row and column totals are the individual counts or probabilities

Refer to the confusion matrix, it is the same thing. The inside cells of the table are intersections between rows and columns, with the row and column totals being the single events (A, B, etc.)

A basic probability, where all events have an equal chance of happening (equal probability of occurrence), is

$$P(X) = \frac{events\ of\ interest}{n} = \frac{X}{n}$$

Table of counts

The following table shows a random sample of 100 hikers and the areas of hiking they prefer.

Counts table example

$$P(male) = \frac{55}{100} = 0.55\ ,\ P(female) = \frac{45}{100} = 0.45$$

$$P(lakes') = \frac{34+25}{100} = 0.59$$

$$P(male \cap coastine) = \frac{16}{100} = 0.16$$