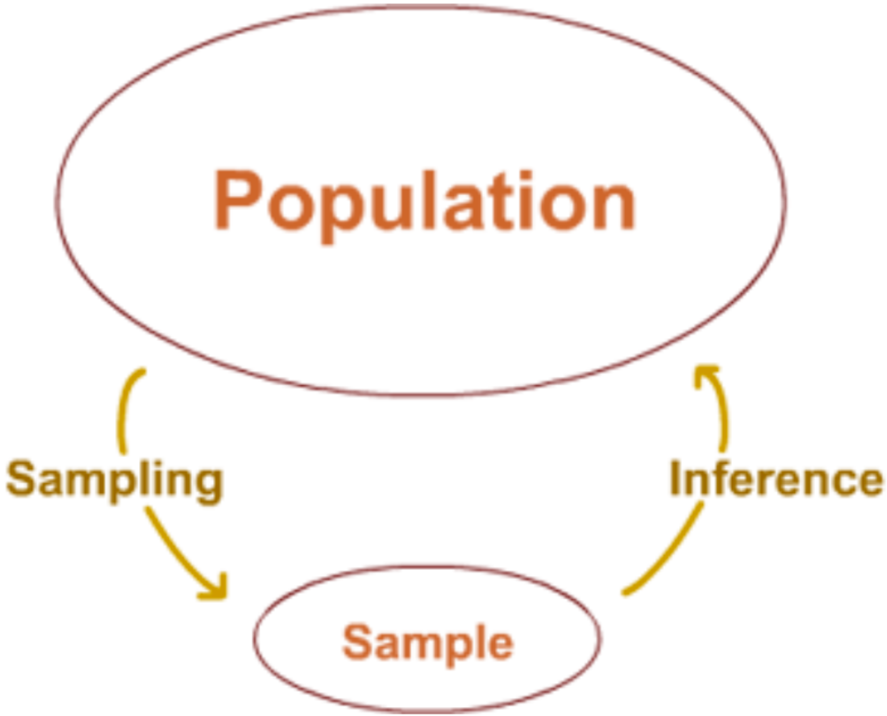
Introduction to probability models

Some questions

1. What percentage of college students skip breakfast?
2. What is the probability that a randomly selected PNU student plays sports more than five hours per week?
3. What is the typical number of blind dates that PNU students have per year?

Populations and Random samples



Sample Spaces

Definition. The sample space (outcome space), denoted S, is the collection of all possible outcomes of an experiment

1. What percentage of college students skip breakfast?
   - S = {yes, no}
2. What is the probability that a randomly selected PNU student plays sports more than five hours per week?
   - $S = \{h: h \geq 0$ hours$\}$
3. What is the typical number of blind dates that PNU students have per year?
   - $S = \{0, 1, 2, \ldots\}$

Events

For sample space S, an event - denoted with capital letters A, B, C, … - is a subset of S. Note a subset is denoted with " $\subset$ ".

- How many blind dates does a randomly selected PNU student have per year?
  - $S = \{0, 1, 2, \ldots\}$
  - If A is the event that a randomly selected student has no more than five times:

    $A = \{0,1,2,3,4,5\}$

Probability

- Probability is a number between 0 and 1, where:
  - a number close to 0 means "not likely"
  - a number close to 1 means "quite likely"
- The mathematical definition of probability
  - For each event *A* of the sample space *S*, P(*A*), called the probability of the event *A* satisfies the following three conditions:
    1) $0 \leq \mathrm{P}(A) \leq 1$
    2) $\mathrm{P}(S) = 1$
    3) Given mutually exclusive events $A_1, A_2, A_3, \ldots$ that is, where $A_i \cap A_j = \emptyset$, for $i \neq j$, $\mathrm{P}(A_1 \cup A_2 \cup \cdots) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \cdots$

Die tossing

- Suppose that all six numbers were equally likely to appear, then
  - $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = 1/6$
  - What is the probability of getting an even number?

Five theorems

1. $\mathrm{P}(A) = 1 - \mathrm{P}(A')$: complement
2. $\mathrm{P}(\emptyset) = 0$
3. $\mathrm{P}(S) = 1$
4. If events A and B are such that $A \subset B$, then $\mathrm{P}(A) < \mathrm{P}(B)$.
5. $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$

Some examples

A company has bid on two large construction projects. The company president believes that the probability of winning the first contract is 0.6, the probability of winning the second contract is 0.4, and the probability of winning both contracts is 0.2.

1. What is the probability that the company wins at least one contract?
2. What is the probability that the company wins the first contract but not the second contract?
3. What is the probability that the company wins neither contract?
4. What is the probability that the company wins exactly one contract?

The multiplication principle

When we roll a fair six-sided die and tosses a fair coin, what is the probability that we get a 6 (die) and a head (coin)?

- To count the number of all possible outcomes in the sample space, we could use the multiplication principle when replication is permitted ➔ 6×2 possible outcomes
- How many possible accounts could be created if each account were required to have exactly 2 alphabet letters and 4 numbers?
  - (e.g. ab1111)
  - What if each account were required to have 2 unique letters and 4 unique numbers? (i.e. no replication)

Permutations

- In how many ways can 7 different books be arranged on a shelf?
  - For n books, $n\times(n-1)\times(n-2)\times\cdots 1 = n!$
  - A permutation of n objects
- How many ways can 4 people fill 3 chairs?
  - ${}_nP_r = \frac{n!}{(n-r)!}$

Combinations

- How many ways can 2 people be selected from 4 to go to a concert?
  - For r people from n, the number of unordered subsets (denoted C) can be computed from the number of ordered subsets ${}_nP_r$.
  - ${}_nP_r = {}_nC_r \times r!$
  - ${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$

Distinguishable Permutations

- Suppose we toss a gold dollar coin 8 times. What is the probability that the sequence of 8 tosses yields 3 heads (H) and 5 tails (T)?

- The number of distinguishable permutations of n objects, of which:
  - n_1 are of one type
  - n_2 are of a second type
  - …
  - n_k are of the last type
  - And $n = n_1 + n_2 + \ldots + n_k$ is given by

$$\binom{n}{n_1 n_2 n_3 \ldots n_k} = \frac{n!}{n_1!\, n_2! \ldots n_k!}$$

Conditional Probabilities

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

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ as long as } P(B) > 0$$

- If a person has renal disease, what is the probability of testing positive?

True	Positive	Negative
Renal disease	44	23
Healthy	10	83

- $P(AB) = P(A \cap B) = P(A|B)P(B)$

Examples

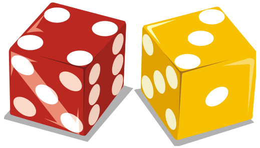
- Example 1.4
  - Cards numbered one through ten in a hat.
  - A = {number is at least five}
  - B = {number is ten}
  - $P(B|A) = ?$

- Example 1.7
  - Suppose an urn contains seven black balls and five white balls. We draw two balls from the urn without replacement. What is the probability that both drawn balls are black?
  - F = {the first ball is black}
  - E = {the second ball is black}
  - $P(E|F) =$
  - $P(F) =$
  - $P(EF) = ?$

Independent events

- [Definition] two events are statistically **independent** if and only if $\mathrm{P}(AB) = \mathrm{P}(A) \cdot \mathrm{P}(B)$ (*i.e.* $\mathrm{P}(A|B) = \mathrm{P}(A)$)
  - Two events are **dependent** if they are not independent

- Example 1.9



  - E_1 = {the sum of the dice is 6}
  - F = {the first die equals 4}
  - Are E_1 and F independent?
  - E_2 = {the sum of the dice is 7}
  - Are E_2 and F independent?

Bayes' formula

- Let E and F be events.
  - $E = EF \cup EF'$ (note EF and EF' are mutually exclusive)

$$\mathrm{P}(E) = P(EF) + P(EF') = P(E|F)P(F) + P(E|F')(1 - P(F))$$

- Example 1.12
  - First urn: two white and seven black balls
  - Second urn: five white and six black balls
  - What is the conditional probability that the outcome of the toss was heads given that a white ball was selected?
  - $P(H|W) = \frac{P(HW)}{P(W)} = \frac{P(W|H)P(H)}{P(W)}$

Example

- Example 1.14
  - D: the event that the tested person has the disease.
  - E: the event that his test result is positive.
  - $P(E|D) = 0.95, P(E|D') = 0.01$
  - $P(D|E) = ?$

Bayesian Network

- Efficient algorithms exist that perform inference and learning in Bayesian networks.
  - Bayesian networks are used for modelling beliefs in computational biology and bioinformatics (gene regulatory networks, protein structure, gene expression analysis, learning epistasis from GWAS data sets) medicine, biomonitoring, document classification, information retrieval, semantic search, image processing, data fusion, decision support systems, engineering, sports betting, gaming, law, study design and risk analysis.
  - Tool Example : http://www.bayesserver.com/BayesianNetworks.aspx