Probability (Part A)

Today's lecture covers basic probability theory (including appendix 5a) up to the concept of a random variable.

Next lecture will review selected discrete and continuous distributions, the concepts of expectation and variance, and a brief overview over bivariate distributions

Basic Definitions: Elementary Outcomes, Events and Set Theory

*Note: **Set theory** takes a distinct role in GISciences. E.g., areas can be conceived as a set of points. The professional literature expresses many spatial operations in set theoretic terms. Furthermore, SQL database concepts and operations are framed as set theoretic operations.*

Definitions: Elementary Outcomes and Sample Space

- Each individual outcome ω_i (small omega) of an experiment is known as an elementary outcome, and the set of all possible elementary outcomes denotes the sample space $\Omega = \{\omega_1, \ldots, \omega_n\}$ (capital omega).

Definition: Event

- An event is a subset of the sample space, i.e., $A \equiv \{\omega_2, \omega_5\}$, with elementary outcomes ω_2 and ω_5 constituting the event A.

Definition: Empty set

- An empty set does not contain any elementary outcomes and is denoted by $\varnothing = \{\ \}$.

Definition: Complementary Event $\bar{A}$:

- Let $\Omega \equiv \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$ and event $A \equiv \{\omega_2, \omega_5\}$ then $\bar{A} = \Omega - A = \{\omega_1, \omega_3, \omega_4\}$

Definition: Subset:

- Definition: a **subset** is a set whose members are *also* elements of another set.
  - Let $A = \{\omega_1, \omega_2\}$ and $B = \{\omega_1, \omega_2, \omega_3\}$ then $A \subset B$
  - Any event A is a subset of the sample space: $A \subset \Omega$
  - A set A and its complement $\bar{A}$ can never be subsets $A \not\subset \bar{A}$

Definition: Event Intersection

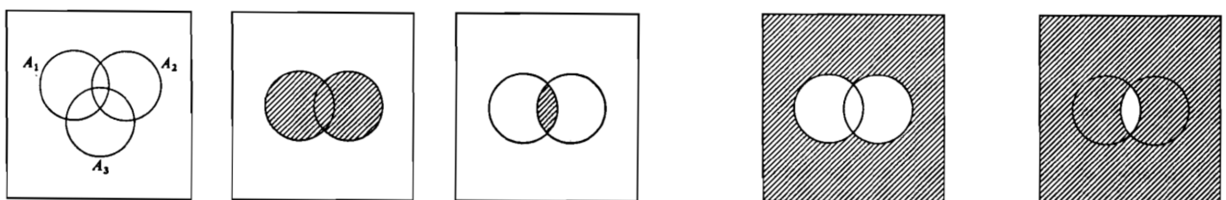
- The intersection symbol is "$\cap$", which is also the logical **AND**. It means that both intersecting events must be ***satisfied simultaneously***.
  - Let $A = \{\omega_1, \omega_2, \omega_4\}$ and $B = \{\omega_1, \omega_2, \omega_3\}$ then $A \cap B = \{\omega_1, \omega_2\}$
  - If $A \cap B \neq \varnothing$ then $A \cap B \subset A$ and $A \cap B \subset B$

Definition: Mutually Exclusive Events

- The intersection of ***mutually exclusive events*** is the empty set: $A \cap B = \varnothing$

Definition: Union

- The Union symbol is "$\cup$", which is also the logical **Either-OR**. Thus either event A or event B or both events happen together.
- Let $A = \{\omega_1, \omega_2, \omega_4\}$ and $B = \{\omega_1, \omega_2, \omega_3\}$ then $A \cup B = \{\omega_1, \omega_2, \omega_3, \omega_4\}$
- Some ***Venn-Diagrams*** highlighting the intersection, union and complement:



Ω is the entire square; $A_1 \cup A_2$; $A_1 \cap A_2$; $\overline{A_1 \cup A_2} = \overline{A_1} \cap \overline{A_2}$; $\overline{A_1 \cap A_2} = \overline{A_1} \cup \overline{A_2}$

Probabilities

Postulated Properties of Probabilities (pp. 203-204) (Kolmogorov's axioms[i]):

1. $0 \le \Pr(\omega_i) \le 1$ for all $\omega_i \in \Omega$. In a deterministic world (no uncertainty) either $\Pr(\omega_i) = 0$ or $\Pr(\omega_i) = 1$
2. $\Pr(A) = \sum_{\omega_i \in A} \Pr(\omega_i)$. This property requires that the elementary outcomes, which constitute the event A, are mutually exclusive.
3. $\Pr(\Omega) = 1$ and $\Pr(\varnothing) = 0$. This can be derived by deductive logic from the previous properties.

Definition of Probabilities:

- Probabilities, as measure of ***uncertainties*** or ***likelihood*** of an experiment with ***many possible random outcomes***, can be obtained from several perspectives:

  [a] Analytical: a probability model based on counting rules

  This perspective is usually based on the appealing assumption that ***all elementary have equal probability*** $\Pr(\omega_i) = \frac{1}{n}$.

  The equal probability assumption leads to the ***classical*** definition $\Pr(A) = \dfrac{|A|}{|\Omega|}$

  with $|\cdot|$ ***denoting the number of elements in a set***

  Criticism: [a] Circular definition because of the use of $\Pr(\omega_i) = \frac{1}{n}$; [b] the event and sample spaces need to be countable.

  [b] Relative Frequency:

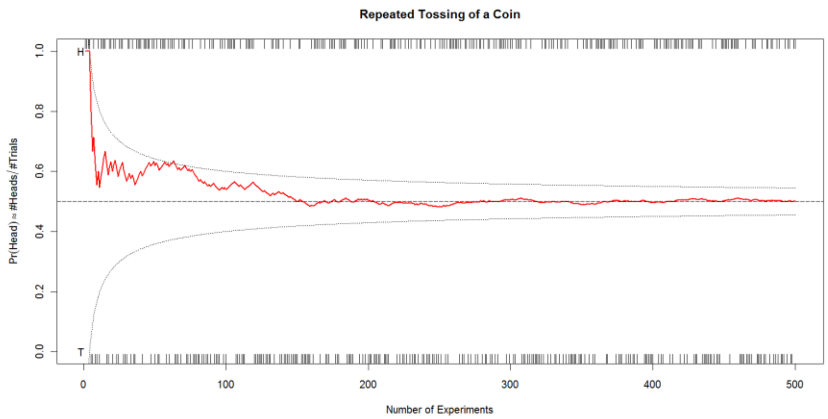
  Probabilities are obtained by repeating a random experiment under fixed conditions over a very large number of trials; e.g., $\Pr(success) = \lim_{n \to \infty} \left(\frac{number\ of\ successes}{total\ number\ of\ experiments} \right)$.

  Larger numbers of repetitions lead to more accurate estimates.

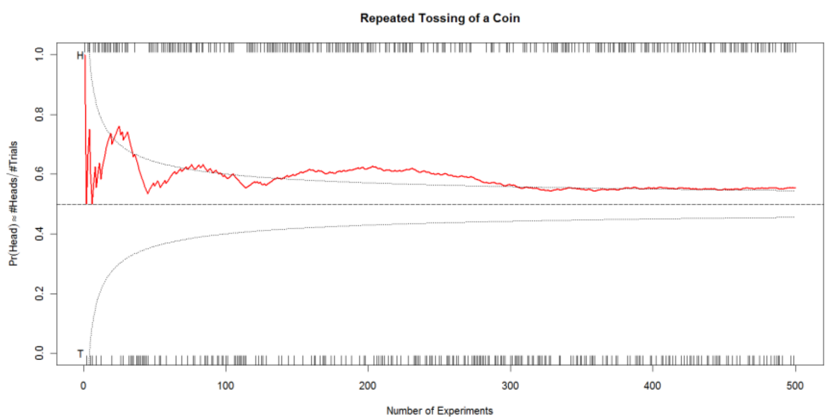
  Criticism: [a] Theoretically requires infinite number of trials. [b] Conditions underlying each trial cannot be held indefinitely constant. [c] Some experiments cannot be repeated indefinitely (e.g., lifespan of light bulbs in a production process)

Example (see script **PROBFREQUENT.R**):

- TA's experiment: Outcome of an process with 500 random tosses of a fair coin:



- Instructor's Experiment: Outcome of 500 random tosses of a fair coin:



[c] <u>Subjective Probabilities:</u> A person assigns subjectively a probability to a random event. These subjective assessments may originate from ***personal experiences*** or "divine

intuition".
For experience based probabilities one assumes, that under similar circumstances one has observed an event occurring with a particular frequency.
Criticism: [a] The problem with subjective probabilities is that an external observer cannot ***reproduce*** the subjective probability (i.e., we cannot look into a person's head); [b] Subjective probabilities may not necessarily satisfy the Kolmogorov's axioms.

Counting Rules for Computing Probabilities:

Counting rules can be used to evaluate theoretically the number of elementary outcomes $|A|$ that satisfy the criteria of an event A (i.e., the ***event space***) and the number of all possible elementary outcomes $|\Omega|$ in the ***sample space***. These allow developing, depending on the underlying assumptions, models for specific probability distributions.

The analytic/geometric probability $\Pr(A) = \frac{|A|}{|\Omega|}$ of event A provides the ***foundation*** of a counting rule based probabilities.

One just needs to ***enumerate*** all possible elementary events in numerator $|A|$ and all possible events in the denominator $|\Omega|$.

Combinatorics

The rules of combinatorics can be used to evaluate the size of the numerator and denominator.

<u>Definition:</u> <u>Combination.</u> A set C of distinguishable objects ***regardless*** of their order.

Example: $\{a,b\}=\{b,a\}$. Both count only as one event.

<u>Definition:</u> <u>Permutations.</u> A set P of distinguishable objects with ***distinct ordering*** (order is relevant here).

Example: $\{a,b\}\neq\{b,a\}$. Both count as separate events.

<u>Definition</u><u>: Product Rule (p 247):</u> Suppose there are r sets of objects. Each set has n_i objects. If we select one object from each set, then there are in total $\prod_{i=1}^{r} n_i = n_1 \cdot n_2 \cdots n_r$ distinct combinations.

Example: we have n objects (set 1) and we sample one object. From the remaining $n-1$ objects (now set 2) we sample again one object. Then there are $n\cdot(n-1)$ distinct combinations.

Classification scheme:

Sampling <u>*with or without*</u> *replacement* and <u>*with*</u> *(permutation)* <u>*or without*</u> *(combination) considering the order* of events.

- Potential permutations by sampling twice from the set $S=\{A,B,C,D,E\}$:

$$S \times S = \begin{Bmatrix} \{AA\} & \{AB\} & \{AC\} & \{AD\} & \{AE\} \\ \{BA\} & \{BB\} & \{BC\} & \{BD\} & \{BE\} \\ \{CA\} & \{CB\} & \{CC\} & \{CD\} & \{CE\} \\ \{DA\} & \{DB\} & \{DC\} & \{DD\} & \{DE\} \\ \{EA\} & \{EB\} & \{EC\} & \{ED\} & \{EE\} \end{Bmatrix}$$

  The symbol $\times$ denote the *Cartesian Product* of two sets

- Recall the definition of the factorial: $n!=1\cdot 2\cdots n$ with $0!=1$ by convention.

- Classification scheme for ***sampling twice*** from $S=\{A,B,C,D,E\}$:

  n expresses the number of different elements in the sample space and p is the size of sample (repeated samples from the set S).

  - With replacement and with considering order: $5\cdot 5=25$

    General: n^p

  - Without replacement and with considering the order: $5\cdot 4=20$

    General *Permutations*: $P_r^n = n\cdot(n-1)\cdots(n-r+1)=\dfrac{n!}{(n-r)!}$

  - Without replacement and without considering order: $5\cdot 4/2=10$

    General *Combinations*: $C_r^n = P_r^n / r! = [n\cdot(n-1)\cdots(n-r+1)]/[1\cdot 2\cdots r] = \dfrac{n!}{(n-r)!\cdot r!} = \dbinom{n}{r}$

- With replacement and without considering order: $6 \cdot 5/2 = 15$
  General: $\binom{n+r-1}{r} = \frac{(n+r-1)!}{(n-1)! \cdot r!}$

- Hypergeometric Rule (p. 249) is a combination of the product rule and the combination rule. Suppose there are p sets of objects. Each set has n_i objects with the total number of objects being $\sum_{i=1}^{p} n_i = n$.
  From each set we select r_i with $r_i \leq n_i$ objects. The different number of possible combinations is $\binom{n_1}{r_1} \cdot \binom{n_2}{r_2} \cdot \ldots \cdot \binom{n_p}{r_p}$

Basic Probability Theorems

The sample space $\Omega = \{A_1 \cap B_1, A_1 \cap B_2, A_1 \cap B_3, A_2 \cap B_1, A_2 \cap B_2, A_2 \cap B_3\}$ is derived by evaluating all pairwise combinations events $\{A_1, A_2\}$ with events $\{B_1, B_2, B_3\}$.

Let us arrange the probabilities of these events and their intersection in a cross-tabulation:

	B_1	B_2	B_3	$\sum$
A_1	$\Pr(A_1 \cap B_1)$	$\Pr(A_1 \cap B_2)$	$\Pr(A_1 \cap B_3)$	$\Pr(A_1)$
A_2	$\Pr(A_2 \cap B_1)$	$\Pr(A_2 \cap B_2)$	$\Pr(A_2 \cap B_3)$	$\Pr(A_2)$
$\sum$	$\Pr(B_1)$	$\Pr(B_2)$	$\Pr(B_3)$	$\Pr(\Omega) = 1.0$

- How are the marginal probabilities of individual events A_1 or B_2 calculated from probabilities of the intersections?
  $\Pr(A_i) = \Pr(A_i \cap B_1) + \Pr(A_i \cap B_2) + \Pr(A_i \cap B_3)$ and $\Pr(B_j) = \Pr(A_1 \cap B_j) + \Pr(A_2 \cap B_j)$
  Why are we allowed to do this? $\Rightarrow$ The intersections $\Pr\left(A_i \cap B_j\right)$ are based mutually exclusive pairs of events.

Definition: Addition Theorem (p 207)

- Warning: Look out for the intersection of events: $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
  Example: $\Pr(A_1 \cup B_2) = \Pr(A_1) + \Pr(B_2) - \Pr(A_1 \cap B_2)$
- Special rule for mutually exclusive events: $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

Definition: Complementation Theorem (p. 208): $\Pr(\bar{A}) = \underbrace{1}_{=\Pr(\Omega)} - \Pr(A)$

Definition: Conditional Probability (p 208)

- The probability of an event may ***change*** once another event has ***taken place***.
  This allows predicting the probabilities of events with the knowledge of an earlier event:
  General rule: $\Pr(A \mid B) = \dfrac{\Pr(A \cap B)}{\Pr(B)}$.
  How to remember the rule: The event that conditions the other event (written after "|") is in the denominator.
- **Note:** Some books use the notation $\Pr(A,B) \Leftrightarrow \Pr(A \cap B)$ for the probability of the intersection of events

Definition: Statistical Independent Events (p 209)

- Under statistical independence, the occurrence of an conditioning event ***does not change*** the probability for another event: $\Pr(A) = \Pr(A \mid B)$.
- Equivalently, we can say that $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$

Definition: Multiplication Theorem (p 209)

- Simple algebraic transformation of the definition of the conditional probabilities gives:
  $$\begin{aligned}\Pr(A \cap B) &= \Pr(A \mid B) \cdot \Pr(B) \\ &= \Pr(B \mid A) \cdot \Pr(A)\end{aligned}$$

- For independent probabilities the multiplication theorem simplifies to
  $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$ because $\Pr(A \mid B) = \Pr(A)$

The Bayes' Theorem (left for the home work)

- From the multiplication theorem we have:
  $\Pr(A \cap B) = \Pr(A \mid B) \cdot \Pr(B) \quad \text{or} \quad \Pr(A \cap B) = \Pr(B \mid A) \cdot \Pr(A)$

  $\Leftrightarrow \Pr(A \mid B) \cdot \Pr(B) = \Pr(B \mid A) \cdot \Pr(A)$

  $\Rightarrow \Pr(A \mid B) = \dfrac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B)} \quad \text{or alternatively,} \quad \Pr(B \mid A) = \dfrac{\Pr(A \mid B) \cdot \Pr(B)}{\Pr(A)}$

- For the Bayesian equation $\Pr(B \mid A) = \dfrac{\Pr(A \mid B) \cdot \Pr(B)}{\Pr(A)}$ the probability $\Pr(B)$ is called the

  a priori probability (first believe in the probability) of event *B*

  $\Pr(B \mid A)$ is called the ***posteriori*** probability (revised probability) of event *B* after we have observed event *A*.

  The probability $\Pr(A \mid B)$ is called the ***likelihood*** of event *A* assuming event *B* has taken place. The likelihood is assumed to be externally known.

- The ***total probability*** $\Pr(A)$ for event *A* can be calculated using marginal probability equation

$$\Pr(A) = \sum_{j=1}^{r} \Pr(A \cap B_j) = \sum_{j=1}^{r} \Pr(A \mid B_j) \cdot \Pr(B_j)$$

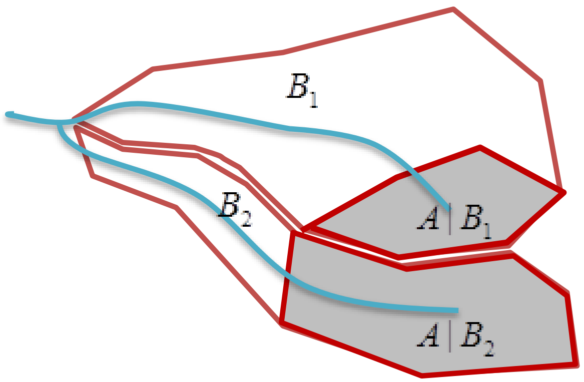
assuming $\Pr(A \mid B_j)$ and $\Pr(B_j)\ \forall j$ are externally given.

- This gives the general form of the Bayes' theorem:

$$\Pr(B_j \mid A) = \frac{\Pr(A \mid B_j) \cdot \Pr(B_j)}{\sum_{j=1}^{r} \Pr(A \mid B_j) \cdot \Pr(B_j)} \text{ for all events } B_j$$

- Example: A paleontologist found downstream of a river some Mosasaur fragment.
  He/she wishes to conduct an expedition into either basin B_1 (with the area of 18 acres) or basin B_2 (with the area of 10 acres) from where these fragments may have originated.
  Research question: Which basin maximizes the likelihood of finding the skeleton?

- The *prior probabilities* of the basins are proportional to their size:
  $\Pr(B_1) = 18/(18+10) = 0.64$ and $\Pr(B_2) = 10/(18+10) = 0.36$
  The higher prior probabilities direct the paleontologist to basin B_1.
- Skeletons are only found in Cretaceous rock (event A). The proportions of Cretaceous rock in either basin are:
  $\Pr(A \mid B_1) = 0.35$ and $\Pr(A \mid B_2) = 0.80$
- Schematic Diagram:



- The total probability of Cretaceous rock in both basins is
  $\Pr(A) = \Pr(A \mid B_1) \cdot \Pr(B_1) + \Pr(A \mid B_2) \cdot \Pr(B_2) = 0.35 \cdot 0.64 + 0.80 \cdot 0.36 = 0.512$

- Posteriori Probabilities:

$$\Pr(B_1 \mid A) = \frac{\Pr(A \mid B_1)\cdot\Pr(B_1)}{\Pr(A)}$$
$$= \frac{0.35\cdot 0.64}{0.512} = 0.4375$$
$$\Pr(B_2 \mid A) = \frac{0.80\cdot 0.36}{0.512} = 0.5625$$

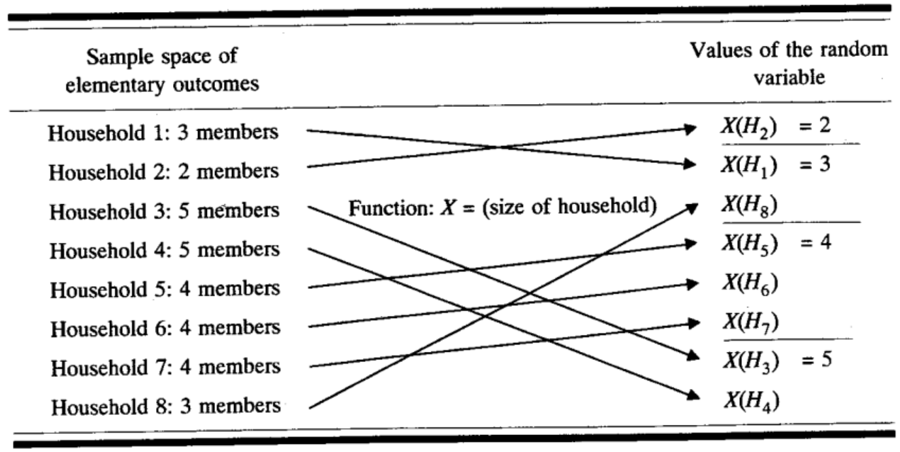
- The posteriori probabilities direct the paleontologist towards searching in basin B_2.

Concept of a Random Variable

- Assumption: A numerical characteristic is measurable for each object (i.e., elementary outcome) in a population (i.e., the sample space)

- We can enumerate (or count) the different occurrence of this characteristic over all objects in the population.

TABLE 6-1
The Concept of a Random Variable



- <u>**Definition:** Random Variable (p 210).</u> The ***function*** $X(\,)$ ***measures the numerical population characteristic***. It is usually denoted in the statistical literature by a capital letter, e.g., X.
- Why are events translated into numerical values? Answer: In order to perform calculations, it is easier to work with numbers rather than with elements in sets.

- The probably of an event A and its associated random variable $X(A)$ must be identical: $\Pr(A) = \Pr\left(X(A)\right)$

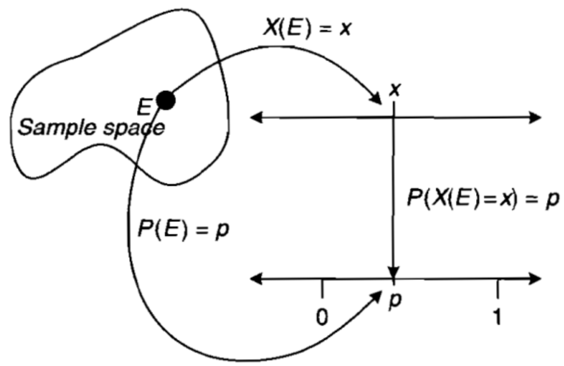


Figure 4.11 A random variable is a mapping of events to the real line.

---

i Def. Axiom: A self-evident principle or one that is intuitively accepted as being true without proof. In mathematics the axioms establish the basis for an argument.
Def. Theorem: A proposition that can be proven to the true by using explicit statements obtained from accepted axioms as argument.