M.Sc. in ‚Transportation Systems'

Applied Statistics in Transport
Theory of Probability

Prof. Regine Gerike
Technische Universität München, mobil.TUM
regine.gerike@tum.de

Munich, 25/10/2011

Plan for Today's Lecture: Theory of Probability

- Questions from last week?
- Basics, Fundamental Principles, Definitions
- Relations Between Random Events
- Relative Frequency and Probability
- Axioms of Probability
- Addition Rules
- Conditional Probability
- Multiplication, Total Probability Rule
- Bayes' Theorem
- Counting Techniques: Permutations, Combinations

- Exercises

Questions for You:

- What scales for variables do you know?
- How many combinations does your bike lock have? Give reasons for the number you indicate.
- Explain the difference between the arithmetic mean and the geometric mean.
- The city of Munich offers welcome information packages with information on public transport services to new citizens. For monitoring purposes two groups of people who recently moved to Munich were surveyed: one group with and the control group without the welcome package. What statistical methods are suitable for analysing whether the mean distance travelled by public transport differs significantly between the two groups?

Definitions I: Probability

General definition:

Likelihood or chance that something is the case or will happen

Common term to phrase suppositions about what will happen in the future

Probability theory: representation of probabilistic concepts in formal terms

Definitions II: Random experiment

- Experiment, trial, or observation that
- can be repeated numerous times
- under the same conditions.
- The outcome of an individual random experiment must be independent and identically distributed (i.i.d.).
- It must in no way be affected by any previous outcome and cannot be predicted with certainty.

To select randomly implies that at each step of the sample, the items that remain in the batch are equally likely to be selected.

Examples of a Random experiment

- The tossing of a coin. The experiment can yield two possible outcomes, heads or tails.
- The roll of a die. The experiment can yield six possible outcomes, this outcome is the number 1 to 6 as the die faces are labelled
- The selection of a numbered ball (1-50) in an urn. The experiment can yield 50 possible outcomes.
- The number of cars arriving at an intersection in a certain period of time.

Definitions III

- **Trial**: Each repetition of a random experiment; we observe an outcome for each trial.
- **Outcome**: Result of a trial.
- **Sample Space** S: Set of all possible outcomes of a random experiment.
- **Frequency** F: Number of times an event occurs (absolute, relative)
- **Event** E: Outcome or a combination of outcomes of an experiment; subset of the sample space.
- Relation **Outcome – Event**: Whenever an outcome satisfies the conditions, given in the event, we say that the event has occurred.
- **Discrete** sample space: Finite or countable infinite set of outcomes.
- **Continuous** sample space: Interval (either finite or infinite) of real numbers.

Example Sample Space:

- The person who was responsible for an road accident failed to stop after the accident.
- The witness to the accident saw on the vehicle number plate: the city registration M for Munich; EU, EV or EY; three numbers starting with 3, at least one of them being a 4.
- He regards all those events as equally likely.
- Seeing the situation and having the statement of the witness, the situation can be interpreted as random, with the whole number of possible events being the sample of random outcomes / events:

- Which and which number of number plates are possible if the witness is right?

Definitions IV:

- **With replacement**: A method to select samples in which items are replaced between successive selections.
- **Without replacement**: A method to select samples in which items are not replaced between successive selections.
- **Ordered/unordered** sample space

- Ex.: Select two out of three items {a,b,c} in the batch:

	With Replacement	Without Replacement
Ordered		
Unordered		

Definitions IV:

- **With replacement**: A method to select samples in which items are replaced between successive selections.
- **Without replacement**: A method to select samples in which items are not replaced between successive selections.
- **Ordered/unordered** sample space

- Ex.: Select two out of three items {a,b,c} in the batch:

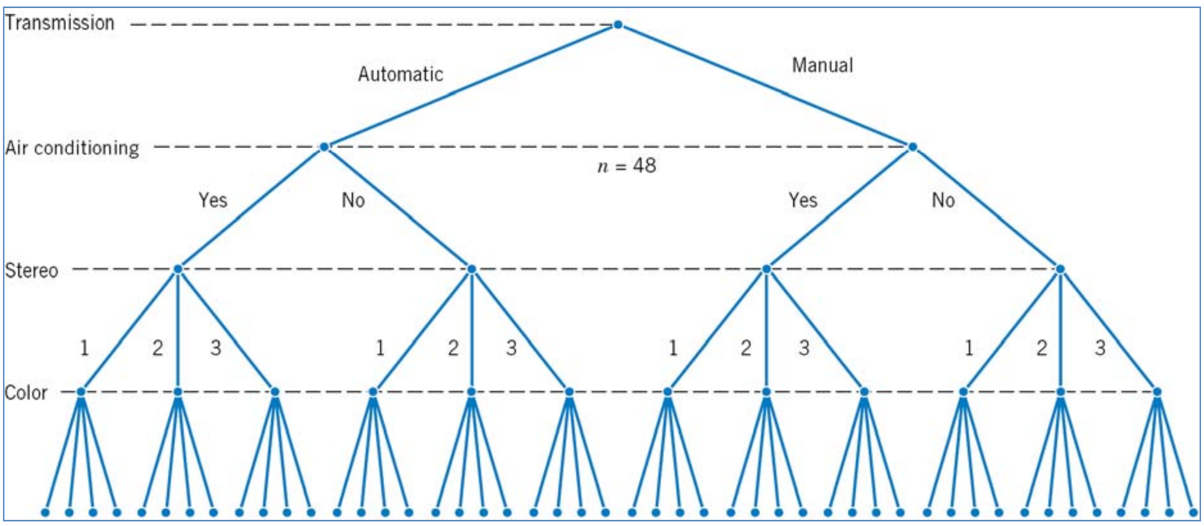
	With Replacement	Without Replacement
Ordered	{{a,a},{a,b},{a,c},{b,b},{b,a}.{b,c},{c,c},{c,a},{c,b}}	{{a,b},{a,c},{b,a}.{b,c},{c,a},{c,b}}
Unordered	{{a,a},{a,b},{a,c},{b,b},{b,c},{c,c}}	{{a,b},{a,c},{b,c}}

Counting Techniques

	Ordered {a,b}≠{b,a}	Unordered {a,b}={b,a}
With Replacement	n^k	$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k! * (n-1)!}$
Without Replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{k! * (n-k)!}$

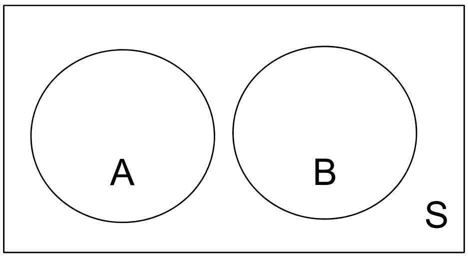
Tree Diagrams: Means for graphical description of sample spaces

Example: Vehicle Equipment:



Relations Between Random Events

- Venn diagrams (set diagrams): Diagrams that show all hypothetically possible logical relations between a finite collection of sets (groups of things, events, observations, etc.). Venn diagrams were invented around 1880 by John Venn. They are used in many fields, including set theory, probability, logic, statistics, and computer science.



Relative Frequency and Probability

Absolute frequency: Number of times the event occurred in a series of trial of a random experiment.

Relative frequency: Proportion of times the event occurred in a series of trial of a random experiment.

Classical definition of probability:

- If a trial results in **n**-exhaustive, mutually exclusive and equally likely cases and **m** of them are favourable to the occurrence of an event A, then the probability of the happening of A, denoted by P(A), is given by:
- $P(A) = m/n$.

Note: Probability is always between 0 and 1:

0 indicates that an event never occurs,
1 indicates that an event will occur certainly.

Example Sample Space, Classical Definition of Probability :

- The witness to the accident saw on the vehicle number plate: the city registration M for Munich; EU, EV or EY; three numbers starting with 3, at least one of them being a 4.

- What is the probability of:

a) EY
b) 34 being the first two numbers
c) 47 being the last two
d) last number 4
e) that there is one zero among the three numbers
f) that the last number is higher than the other two

Axioms of Probability

- **Axioms of probability**: A set of rules that probabilities defined on a sample space must follow.
- **Probability**: A numerical measure between 0 and 1 assigned to events in a sample space. Higher numbers indicate the event is more likely to occur.

$$P(A) \geq 0$$

$$P(\Omega) = 1$$

$$P(A \cup B) = P(A) + P(B) \text{ with } A \cap B = \emptyset$$

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

Addition Rules

$0 \leq P(A) \leq 1 with\ A \subset \Omega$

$P(\emptyset) = 0$

$P(A) \leq P(B)\ with\ A \subset B\ and\ A, B \subset \Omega$

$P(A) = 1 - P(\overline{A})\ 1 = P(A) + P(\overline{A})\ with\ \overline{A} = \Omega \backslash A$

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

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

Conditional Probability, Multiplication Rule

- **Conditional Probability**: Probability of an event given that the random experiment produces an outcome in another event.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}\ (with\ P(B) > 0\ and\ A, B \subset \Omega)$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}\ (with\ P(A) > 0\ and\ A, B \subset \Omega)$$

- **Multiplication Rule**:

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

Independent Events

- Independence:
- A property of a probability model and two (or more) events
- that allows the probability of the intersection to be calculated as the product of the probabilities.

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

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

Example: Independent Events

Samples of emissions from three suppliers are classified for conformance to air quality specifications. The results from 100 samples are summarised as follows:

		Conforms		
		Yes	No	
	1	22	8	30
Supplier	2	25	5	30
	3	30	10	40
Total		77	23	100

Let S1 denote the event that a sample is from supplier 1, and let CY denote the event that a sample conforms to specifications.

1. Are the events S1 and CY independent?
2. Determine P(CY I S1).

Total Probability Rule - Example

Suppose that in semiconductor manufacturing the probability is 0.10 that a chip that is subjected to high levels of contamination during manufacturing causes a product failure.

The probability is 0.005 that a chip that is not subjected to high contamination levels during manufacturing causes a product failure.

In a particular production run, 20% of the chips are subject to high levels of contamination.

What is the probability that a product using one of these chips fails?

Total Probability Rule

For two events:

$$P(B) = P(B \cap A) + P\left(B \cap \overline{A}\right) = P(B|A) * P(A) + P\left(B|\overline{A}\right) * P\left(\overline{A}\right)$$

For k events:

$$P(B) = \sum_{i=1}^{k} P(B|A_i) * P(A_i)$$

Bayes' Theorem

$$P(A_i|B) = \frac{P(A_i) * P(B|A_i)}{\sum_{i=1}^{n} P(A_i) * P(B|A_i)}$$

Thank you for your attention.

Regine Gerike
Technische Universität München
mobil.TUM
Office: 1753
Tel +49.89.289.28575
regine.gerike@tum.de