REGIONAL TRAINING COURSE ON APPLIED ECONOMETRIC ANALYSIS (SUMMER SCHOOL) FOR YOUNG ECONOMISTS / RESEARCHERS ORGANIZED BY WIUT AND IFPRI JUNE 4-15, 2018

Review of Probability

Farrukh Ataev
Lecturer, WIUT
5 June 2018

Outline

- **Session 1 Review of probability**
  - (Tue 5 June at 13:30-15:00)
- **Session 2 Review of probability (continued)**
  - (Tue 5 June at 15:30-17:00)
- **Session 3 Probability distributions**
  - (Wed 6 June at 9:00-10:30)
- **Session 4 Probability distributions (continued)**
  - (Wed 6 June at 11:00-12:30)

Sessions 1 Outline

- Concepts of experiment, outcome, event
- Approaches to probability (Classical and Empirical)
- Relationships of events (Mutually Exclusive, Dependent and Independent, Joint probability, Bayes' Formula)

Probability

Any activity or process involves a chance (or probability, possibility, likelihood)

Examples:

- Forecasting (chance of raining or not)
- Lottery or any game (chance of winning, losing or making a draw)
- Running a business (chance of succeeding or failing)
- Investing (chance of making extra money or not)
- Buying a product (chance of buying normal or defective product)
- Taking an exam (chance of passing or failing)

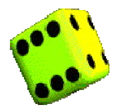
Exercise: You give an example

Experiment and Sample space

Experiment is any activity or process that generates well-defined outcomes (or results).

Sample space for an experiment is the set of all experimental outcomes.

Experiment	Experimental outcomes	Sample space
Toss a coin	Head, Tail	S = {Head, Tail}
Apply for a job	Hired, Not hired	S = {Hired, Not Hired}
Roll a die	1, 2, 3, 4, 5, 6	S = {1, 2, 3, 4, 5, 6}
Run a business	Profit, Loss, Even	S = {Profit, Loss, Even}
Sales (units)	0, 1, 2, 3, 4, …	S = {0,1,2,3,4,…}



Event

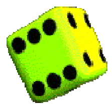
Event is a subset of the sample space. Alternatively, it is a collection (group) of outcomes.

Simple event contains a single outcome.

Example:

The experiment of rolling a die has a sample space of six outcomes:

$$S = \{1, 2, 3, 4, 5, 6\}.$$



1) Simple events: $E_1 = \{1\}$, $E_2 = \{2\}$, $E_3 = \{3\}$, $E_4 = \{4\}$, $E_5 = \{5\}$, $E_6 = \{6\}$.
2) Event of an even number: $A = \{2, 4, 6\}$.
3) Event of an odd number: $B = \{1, 3, 5\}$.

Probability

Probability is a numerical measure of the chance (or likelihood) that a particular event will occur.

Probability values are always assigned on a scale of 0 to 1:

$$0 \le P(E) \le 1$$

A probability near 0 indicates that an event is unlikely to occur, whereas near 1 indicates that an event is likely to occur.

Total sum of probabilities of n simple events is equal to 1:

$$\sum_{k=1}^{n} P(E_k) = P(E_1) + P(E_2) + \cdots + P(E_n) = 1$$

Two approaches to probability

- **Classical**
- **Empirical (Frequency)**

Classical probability

$$P(E) = \frac{n(E)}{n(S)} = \frac{n(favorable\ outcomes)}{n(possible\ outcomes)}$$

Note: It is the same as the relative frequency of the event of equally likely events.

Example:

2 out of 20 products are defective at a store. If a customer randomly buys 1 product, what is $P(D)$?

$$P(D) = \frac{n(D)}{n} = \frac{2}{20} = 0.1$$

Exercise: Find:

a) $P(N)$

b) $P(D\ and\ N) = P(D \cap N)$

c) $P(D\ or\ N) = P(D \cup N)$.

Mutually exclusive (disjoint) events:

$$A \cap B = \emptyset \quad (P(A \cap B) = 0)$$

Empirical probability

Empirical probability of an event refers to the frequency of similar events happened in the past.

Example: A radio station asked a random sample of 1,000 out of the 1,000,000 listeners to find out how many preferred one of the four types of music.

Type of music	Number of listeners
Popular	521
Big band	188
Classical	207
Other	84
Total	1,000

What is the probability that a randomly chosen listener from the sample likes pop music? (Descriptive statistics)

$$P(pop) = \frac{n(pop)}{n} = \frac{521}{1{,}000} = 0.521.$$

Question: What is the probability that a randomly chosen listener from the population likes pop music? (Inferential statistics)

Probability formulas

1) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. (for disjoint events: $P(A \cup B) = P(A) + P(B)$)
2) $P(A \cap B) = P(A) \cdot P(B|A)$. (for independent events: $P(A \cap B) = P(A) \cdot P(B)$)
3) $P(B|A) = \frac{P(A \cap B)}{P(A)}$.
4) $P(A) = 1 - P(A')$.

Example: 2 out of 20 products are defective at store A and 4 out of 50 products are defective at store B. If a customer randomly buys 1 product from each store, find the probability of:

a) both defective: $P(D_A \cap D_B)$. b) at least 1 defective: $P(D_A \cup D_B)$.

a) $P(D_A \cap D_B) = P(D_A) \cdot P(D_B|D_A) = P(D_A) \cdot P(D_B) = \frac{2}{20} \cdot \frac{4}{50} = 0.008.$

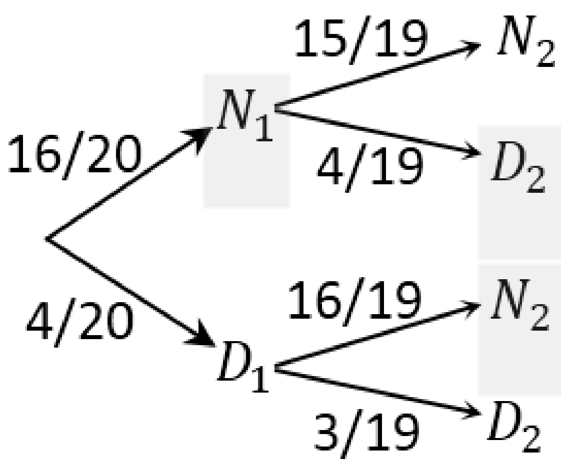
b) $P(D_A \cup D_B) = P(D_A) + P(D_B) - P(D_A \cap D_B) = \frac{2}{20} + \frac{4}{50} - \frac{2}{20} \cdot \frac{4}{50} = 0.172$

$P(D_A \cup D_B) = 1 - P(N_A \cap N_B) = 1 - \frac{18}{20} \cdot \frac{46}{50} = 0.172.$

Probability tree diagram

Probability tree diagram is used to show the outcomes of an experiment and their relative probabilities.

Example: 4 out of 20 products are defective at a store. If a customer randomly buys 2 products, find: a) $P(D_1 \cap D_2)$. b) $P(N_1 \cup N_2)$.



$$P(N_1 \cap N_2) = P(N_1) \cdot P(N_2|N_1) = \frac{16}{20} \cdot \frac{15}{19} = 0.63$$

$$P(N_1 \cap D_2) = P(N_1) \cdot P(D_2|N_1) = \frac{16}{20} \cdot \frac{4}{19} = 0.17$$

$$P(D_1 \cap N_2) = P(D_1) \cdot P(N_2|D_1) = \frac{4}{20} \cdot \frac{16}{19} = 0.17$$

$$P(D_1 \cap D_2) = P(D_1) \cdot P(D_2|D_1) = \frac{4}{20} \cdot \frac{3}{19} = 0.03$$

Joint probability table

Contingency table is a table of joint-occurrence frequencies of two variables.

Joint probability table is a table with probabilities calculated based on the contingency table.

Example: The table shows the numbers of full-time (F) and part-time (P) workers among 100 employees.

	F	P	Totals
Men	50	20	70
Women	25	5	30
Totals	75	25	100

$\Rightarrow$

	F	P	Totals
Men	0.5	0.2	0.7
Women	0.25	0.05	0.3
Totals	0.75	0.25	1

Find the probabilities that: **a)** an employee is a man and part-timer;

b) a man is part-timer; **c)** a woman is full-timer.

Bayes' Formula

Bayes' formula is used to calculate the posterior (updated, revised with an additional) probabilities:

Suppose $A_1, A_2, \dots, A_n$ are pairwise disjoint events, whose union is the sample space. Then for an event E and for each i with $1 \le i \le n$:

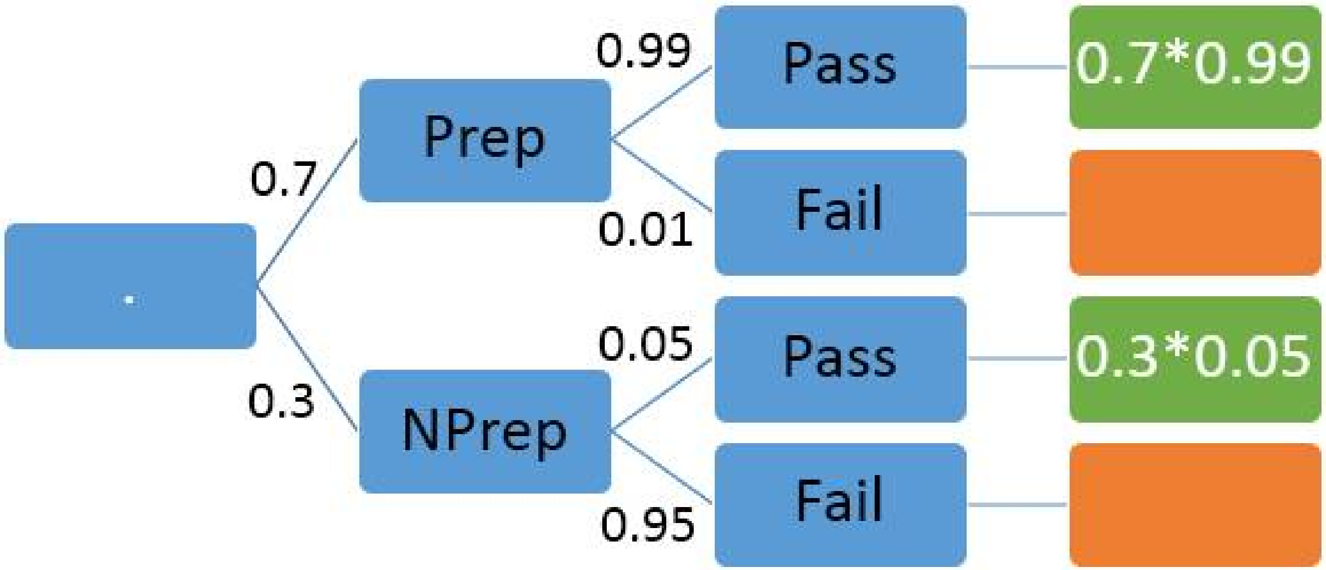
$$P(A_i|E) = \frac{P(A_i) \cdot P(E|A_i)}{P(A_1) \cdot P(E|A_1) + \cdots + P(A_n) \cdot P(E|A_n)}$$

Example: Given that a student had prepared, the probability of passing a Math exam is 0.99. Given that a student did not prepare, the probability of passing the Math exam is 0.05. Assume that the probability of preparing is 0.7. The student passes in the Math exam. What is the probability that the student did not prepare?

$P(Pass|Prep) = 0.99$, $P(Pass|NPrep) = 0.05$, $P(Prep) = 0.7$.

$P(NPrep|Pass) = ?$

Solution



$$P(NPrep|Pass) = \frac{P(NPrep \cap Pass)}{P(Pass)} = \frac{P(NPrep) \cdot P(Pass|NPrep)}{P(Prep \cap Pass) + P(NPrep \cap Pass)} =$$

$$\frac{P(NPrep) \cdot P(Pass|NPrep)}{P(Prep) \cdot P(Pass|Prep) + P(NPrep) \cdot P(Pass|NPrep)} = \frac{0.3 \cdot 0.05}{0.7 \cdot 0.99 + 0.3 \cdot 0.05} = 0.02.$$

Question: $P(Prep|Pass) = ?$

Bayes' Formula

Example: A firm receives 65% of its parts from one supplier A_1 and 35% from a second supplier A_2. The table shows the percentages of good G and defective D parts received from the two suppliers.

	G	D
A_1	98	2
A_2	95	5

Given:

$P(A_1) = 0.65$, $P(A_2) = 0.35$,

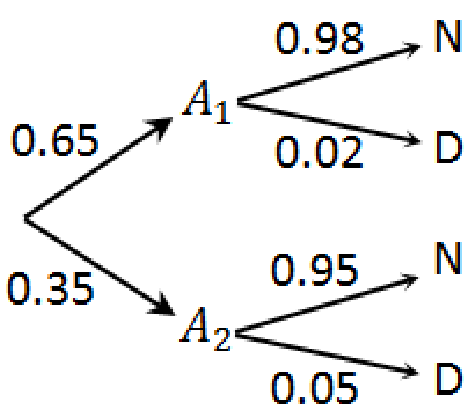
$P(G|A_1) = 0.98$, $P(D|A_1) = 0.02$,

$P(G|A_2) = 0.95$, $P(D|A_2) = 0.05$.

Suppose that a machine broke down due to a defective part and the firm wants to know the probability that the part came from A_1 and the probability that it came from A_2. That is:

$$P(A_1|D) = ? \qquad P(A_2|D) = ?$$

Solution



$$P(A_1 \cap N) = P(A_1) \cdot P(N|A_1) = 0.65 \cdot 0.98 = 0.637$$

$$P(A_1 \cap D) = P(A_1) \cdot P(D|A_1) = 0.65 \cdot 0.02 = 0.013$$

$$P(A_2 \cap N) = P(A_2) \cdot P(N|A_2) = 0.35 \cdot 0.95 = 0.3325$$

$$P(A_2 \cap D) = P(A_2) \cdot P(D|A_2) = 0.35 \cdot 0.05 = 0.0175$$

$$P(A_1|D) = \frac{P(A_1) \cdot P(D|A_1)}{P(A_1) \cdot P(D|A_1) + P(A_2) \cdot P(D|A_2)} = \frac{0.65 \cdot 0.02}{0.65 \cdot 0.02 + 0.35 \cdot 0.05} = 0.426$$

Exercise: Find $P(A_2|D)$.

Reading

1) Murray R. Spiegel, *Schaum's outline of Theory and Problems of Probability and Statistics*, McGraw-Hill, 23 edition, 1998.

2) Nitis Mukhopadhyay, Probability and Statistical Inference, Marcel Dekker, Inc. 2000.