Theoretical Foundations
Probabilities

Monia Ranalli
monia.ranalli@uniroma2.it

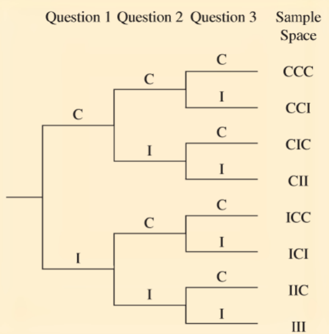
Objectives

- understand the **probability basics**
- quantify **random phenomena** through probability
- classify different **types of variables**
- understand what phenomena can be modeled by **binomial distribution** and how to calculate **binomial probabilities**
- understand what phenomena can be modeled by **normal distribution** and how to calculate **normal probabilities**

Finding Probabilities

- **Sample Space**: the set of all possible outcomes
- **Event**: a subset of the sample space; it corresponds to a particular outcome or a group of possible outcomes
- The **probability** that event E occurs is denoted by $P(E)$

Experiment: Three questions. Students can answer Correctly or Incorrectly



- **Event A**: student answers all 3 questions correctly = (CCC)
- **Event B**: student passes (at least 2 correct) = (CCI, CIC, ICC, CCC)

What is probability?

- **Classical Rule.** When all outcomes are equally likely, then

$$P(E) = \frac{\text{number of outcomes in E}}{\text{number of possible outcomes}}$$

  Example: What is the chance of getting a head (H)? $P(H) = 1/2$
- **Relative Frequency (Empirical Approach).** Flip the coin a very large number of times and count the number of H out of the total number of flips,

$$P(E) \approx \frac{\text{number of outcomes in E}}{\text{number of possible outcomes}}$$

  Example: if we flip the given coin 10,000 times and observe 4555 heads and 5445 tails, then for that coin, $P(H)$ 0.4555.
- **Subjective Probability.** It reflects personal belief which involves personal judgment, information, intuition, etc.

  Example: what is P(you will get an A in the Statistics course)?

  Each student may have a different answer to the question.

  Side note: Bayesian statistics is a branch of statistics that uses subjective probability as its foundation

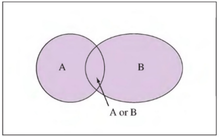
Examples for Using the Classical Rule to Find the Probability

- Find the probability that exactly one head appears in two flips of a fair coin.
  Sample Space: $\{(H,H),(H,T),(T,H),(T,T)\}$
  $P(\text{getting exactly one } H \text{ in two flips of a fair coin}) =$
  $P(\{(H,T),(T,H)\}) = 2/4 = 1/2$
- Find the probability that the sum of two faces is greater than or equal to 10 when one rolls a pair of fair dice.
  Sample Space:
  $\{(1,1),(2,1),(3,1),(4,1),(5,1),(6,1),(1,2),(2,2),(3,2),(4,2),(5,2),(6,2),$
  $(1,3),(2,3),(3,3),(4,3),(5,3),(6,3),(1,4),(2,4),(3,4),(4,4),(5,4),(6,4),$
  $(1,5),(2,5),(3,5),(4,5),(5,5),(6,5),(1,6),(2,6),(3,6),(4,6),(5,6),(6,6)\}$

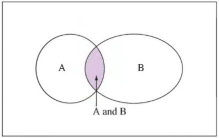
  Let S be the sum of the points in the two faces:
  $P(S \text{ greater than or equal to } 10) =$
  $P(S=10)+P(S=11)+P(S=12) = P(\{(4,6),(5,5),(6,4)\}) +$
  $P(\{(5,6),(6,5)\}) + P(\{(6,6)\}) = 3/36 + 2/36 + 1/36 = 1/6$

Set Operations

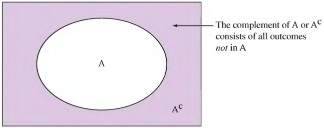
- **Union.** A or B also written as $A \cup B$ = outcomes in A or B
  $P(A \cup B) = \frac{\text{number of outcomes in A or B}}{\text{total number of individual outcomes}}$



- **Intersection.** A and B also written as $A \cap B$ = outcomes in A and B
  $P(A \cap B) = \frac{\text{number of outcomes in A and B}}{\text{total number of individual outcomes}}$



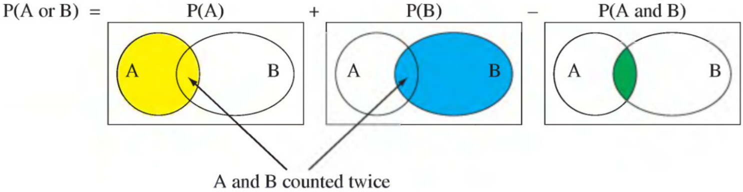
- **Complement.** A^c also written as $\bar{A}$ = outcomes not in A



A and B are called **mutually exclusive** (disjoint) if the occurrence of outcomes in A excludes the occurrence of outcomes in $B \Rightarrow$ there are no elements in $A \cap B$ and thus $P(A \cap B) = 0$: A and $\bar{A}$ are mutually exclusive.

Probability Properties

1. $0 \leq P(A) \leq 1$
2. $P(\bar{A}) = 1 - P(A)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



Example

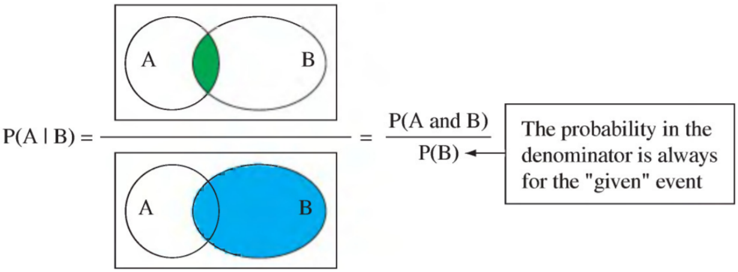
Experiment: draw at random a person from the following population

		Smoker Yes	Smoker No	
Gender	**M**	100	60	160
	F	10	30	40
		110	90	200

- $P(M) = 160/200 = 0.80 = P(F^c) = 1 - 0.20 = 0.80$
- $$\begin{aligned} P(M \cup Yes) &= 170/200 = 0.85 \\ &= P((F \cap NO)^c) = 1 - 0.15 = 0.85 \\ &= P(M) + P(YES) - P(M \cap YES) = 0.80 + 0.55 - 0.50 \\ &= 0.85 \end{aligned}$$
- $P(M \cap YES) = 100/200 = 0.50$

Conditional Probability

- $P(A \mid B)$ is interpreted as the "Probability event A happens given that event B has happened"



- **Remark**: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$ and $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$
  $\Rightarrow P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A)$.
  Usually $P(A \mid B) \neq P(B \mid A)$

Example

Experiment: draw at random a person from the following population

		Smoker		
		Yes	**No**	
Gender	**M**	100	60	160
	F	10	30	40
		110	90	200

- $P(YES \mid M) = P(YES \cap M)/P(M) = (100/200) \times (200/160) = 100/160$
- $P(NO \mid M) = P(NO \cap M)/P(M) = (60/200) \times (200/160) = 60/160$

Example - Multiplication rule

Two cards are drawn at random from a deck (without replacement). Compute the probabilities:
1) (1♡, 3♢); 2) I card 4♠; 3) II card 4♠; 4) P(II 4♠ | I 3♢).

Solutions

1. $P(1\heartsuit, 3\diamondsuit) = P(\text{II}3\diamondsuit \mid \text{I}1\heartsuit)P(\text{I}1\heartsuit) = \frac{1}{51}\frac{1}{52}$
2. $P(\text{I}4\spadesuit) = \frac{1}{52}$
3. $P(\text{II}4\spadesuit) = \frac{51}{51 \times 52} = \frac{1}{52}$
4. $P(\text{II}4\spadesuit \mid \text{I}3\diamondsuit) = \frac{1}{51}$

Independent Events

- Two events A and B are **independent** if the probability that one occurs is not affected by whether or not the other event occurs

$$\Downarrow$$

For any given probabilities for events A and B, the events are independent if any **ONE** of the following are true

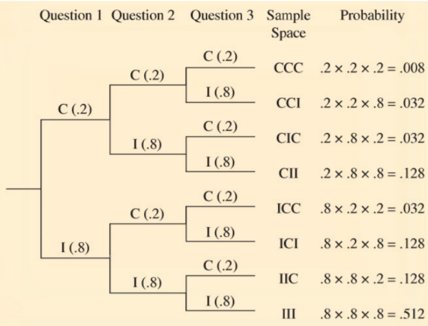
1. $P(A \cap B) = P(A) \times P(B)$
2. $P(A \mid B) = P(A)$
3. $P(B \mid A) = P(B)$

- **Remark**: Independent is very different from mutually exclusive. In fact, **mutually exclusive events are dependent**. If A and B are mutually exclusive events, there is nothing in $A \cap B$, and thus:$P(A \cap B) = 0 \neq P(A) \times P(B)$

Example

Knowing that probability of guessing correctly is 0.2 and assuming that each answer is independent of the other...

1. What is the probability of getting 3 questions correct by guessing? **0.008**
2. What is the probability of getting 2 questions correct by guessing? $\mathbf{0.032 + 0.032 + 0.032 = 0.096}$
3. What is the probability of getting at least 2 questions correct by guessing? $\mathbf{0.032 + 0.032 + 0.032 + 0.008 = 0.104}$



Example

Experiment: draw at random a person from the following population

		Smoker		
		Yes	**No**	
Gender	**M**	100	60	160
	F	10	30	40
		110	90	200

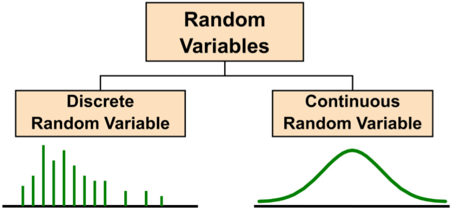
Are the events M and YES independent?

- $P(YES \cap M) = 100/200 = 0.5 \neq 0.44 = 110/200 \times 160/200 = P(YES) \times P(M)$
- $P(YES \mid M) = 100/160 \neq 110/200 = P(YES)$

They are not independent

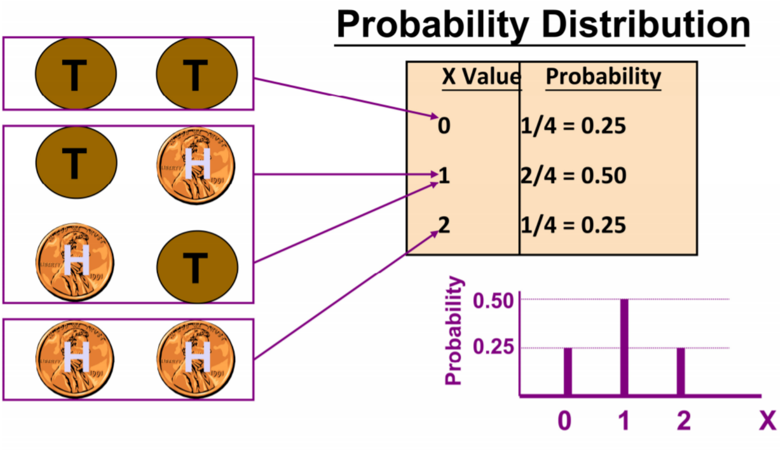
Types of Random Variables

- **Random Variable**: a numerical measurement of the outcome of a random experiment (phenomenon).
  - *Discrete Random Variable*: When the random variable can assume only a countable (such as $0, 1, 2, \ldots$), sometimes infinite, number of values (such as the number of tosses to get the first Head when flipping a fair coin).
  - *Continuous Random Variable*: When the random variable can take any value in a real interval (such as height or weight of a newborn baby).
- **The probability distribution** of a random variable specifies its possible values and their probabilities



Example

Experiment: Toss 2 coins. $X = \#$ heads.



Discrete Random Variables

- A discrete random variable X assigns a **probability** $P(x)$ to each possible value x:
  - For each x, the probability $P(x)$ falls between 0 and 1
  - The sum of the probabilities for all the possible x values equals 1
- The **mean** of a probability distribution, also called **expected value**, for a discrete random variable is

$$\mu = \sum xp(x).$$

The expected value reflects not what we will observe in a single observation, but rather what we expect for the average in a long run of observations. It is not unusual for the expected value of a random variable to equal a number that is NOT necessarily a possible outcome.

- The **variance** and **standard deviation** of a probability distribution, denoted by the parameter σ^2 and σ, respectively, measures its variability

$$\sigma^2 = \sum_x (x - E(x))^2 p(x), \quad \sigma = \sqrt{\sum_x (x - E(x))^2 p(x)}.$$

Larger values of σ correspond to greater spread. Roughly, σ describes how far the random variable falls, on the average, from the mean of its distribution

Example

Experiment: Toss 2 coins. $X = \#$ heads.

$X = x$	$P(x)$
0	0.25
1	0.50
2	0.25

- $E(X) = (0 \times 0.25) + (1 \times 0.50) + (2 \times 0.25) = 1$
- $\sigma = \sqrt{(0-1)^2 \times 0.25 + (1-1)^2 \times 0.50 + (2-1)^2 \times 0.25} = \sqrt{0.50} = 0.707$

Binomial Distribution

Conditions

- The experiment consists of n **identical trials**
- Each trial results to have two distinct complimentary outcomes, a **success** (π) and a **failure** $(1-\pi)$
- **The probability of success**, denoted π, **remains the same** from trial to trial
- The n trials are **independent** $\Rightarrow$ the outcome of any trial does not affect the outcome of the others

$$\Downarrow$$

- The **probability** of x successes equals:
  $P(x) = \frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{n-x}$
- The **mean** is $\mu = E(X) = n\pi$
- The **variance** is $\sigma^2 = n\pi(1-\pi)$

Example

Knowing that 80% of voters said they voted for the White party, what is the probability that randomly drawing (with replacement) 6 voters

1. all claim to vote for whites
2. five claim to vote for whites
3. one claim to vote for whites

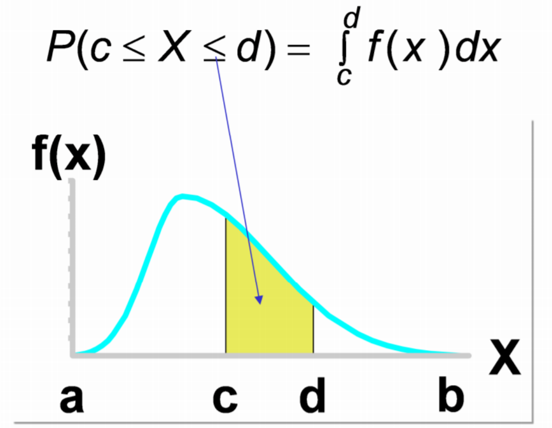
Solution

1. $P(6) = \frac{6!}{6!(6-6)!}(0.80)^6(0.2)^{6-6} = 1 \times 0.262144 \times 1 = 0.262144$
2. $P(5) = \frac{6!}{5!(6-5)!}(0.80)^5(0.2)^{6-5} = 6 \times 0.32768 \times 0.2 = 0.393216$
3. $P(1) = \frac{6!}{1!(6-1)!}(0.80)^1(0.2)^{6-1} = 6 \times 0.8 \times 0.00032 = 0.001536$

Continuous Random Variables

- It assumes values in an interval $\Rightarrow$ If X is continuous $P(X = x) = 0$ for any given value x
- Its **probability distribution** is specified by a **density curve**. The probability of an interval is given by the area under the curve over the interval
- Each interval has probability between 0 and 1 but the density can be greater than 1. The interval containing all possible values has probability equal to 1.
- **Normal distribution** is a family of continuous distributions commonly used to model many histograms of real-life data which are mound-shape and symmetric (for example, height, weight, etc.).
  - A normal curve has two parameters: mean μ (center of the curve), that is also the mode and the median, and standard deviation σ(spread about the center)
  - A normal distribution with $\mu = 0$ and $\sigma = 1$ is called a standard normal curve, usually denoted as Z

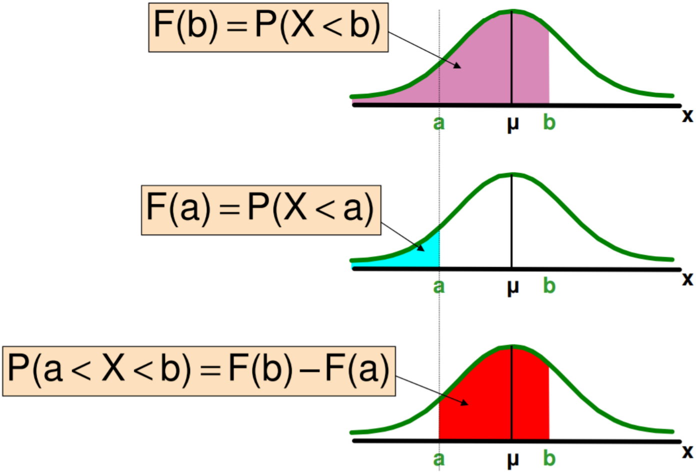
How to compute the probability



- The random variable has an infinite theoretical range: $-\infty$, $+\infty$
- $f(x)$ cannot be negative and the total area under the curve must be 1
- $f(x)$ it is not a probability and it can be greater than 1

How can we compute probabilities of intervals?

Use of the cumulative probabilities of the standard normal distribution



How to use the Standard Normal Cumulative Table

Second Decimal Place of z										
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
...										
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9139	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441

Examples

- $P(Z < 1.43) = 0.9236$
- $P(0 < Z < 1.43) = P(Z < 1.43) - P(Z < 0) = 0.9236 - 0.5000 = 0.4236$
- $P(1.30 < Z < 1.54) = P(Z < 1.54) - P(Z < 1.30) = 0.9382 - 0.9032 = 0.0350$

Z-Scores and the Standard Normal Distribution

- The z-score for a value x of a random variable is the number of standard deviations that x falls from the mean: $(x - \mu)/\sigma$
- For NON standard normal distributions $\rightarrow$ transform the normal distribution into a standard normal distribution by applying the z-score transformation
- The z-scores have the standard normal distribution, i.e. a normal distribution with $\mu = 0$ and $\sigma = 1$

Example

In a population the height distribution is well approximated by a normal with $\mu = 170$ cm and $\sigma = 3$. Compute the probability that the height (X) of a person drawn at random is within:

1.

$$P(167 < X < 173) = P\left(\frac{167-170}{3} < \frac{X-170}{3} < \frac{173-170}{3}\right)$$
$$= \quad P(-1 < Z < 1) = P(Z < 1) - P(Z < -1) = 0.8413 - 0.1587 = 0.6826$$

2.

$$P(167 < X < 170) = P\left(\frac{167-170}{3} < \frac{X-170}{3} < \frac{170-170}{3}\right)$$
$$= \quad P(-1 < Z < 0) = P(Z < 0) - P(Z < -1) = 0.5 - 0.1587 = 0.3413$$

3.

$$P(170 < X < 173) = P\left(\frac{170-170}{3} < \frac{X-170}{3} < \frac{173-170}{3}\right)$$
$$= \quad P(0 < Z < 1) = P(Z < 1) - P(Z < 0) = 0.8413 - 0.5 = 0.3413$$

4.

$$P(166 < X < 174) = P(-1.33 < Z < 1.33) = 0.9082 - 0.0918 = 0.8164$$

Empirical Rule

For any data set having **approximately a bell-shaped distribution**: roughly **68%** of the observations lie within **one standard deviation** to either side of the mean; roughly **95%** of the observations lie within **two standard deviations** to either side of the mean; roughly **99.7%** of the observations lie within **three standard deviations** to either side of the mean.