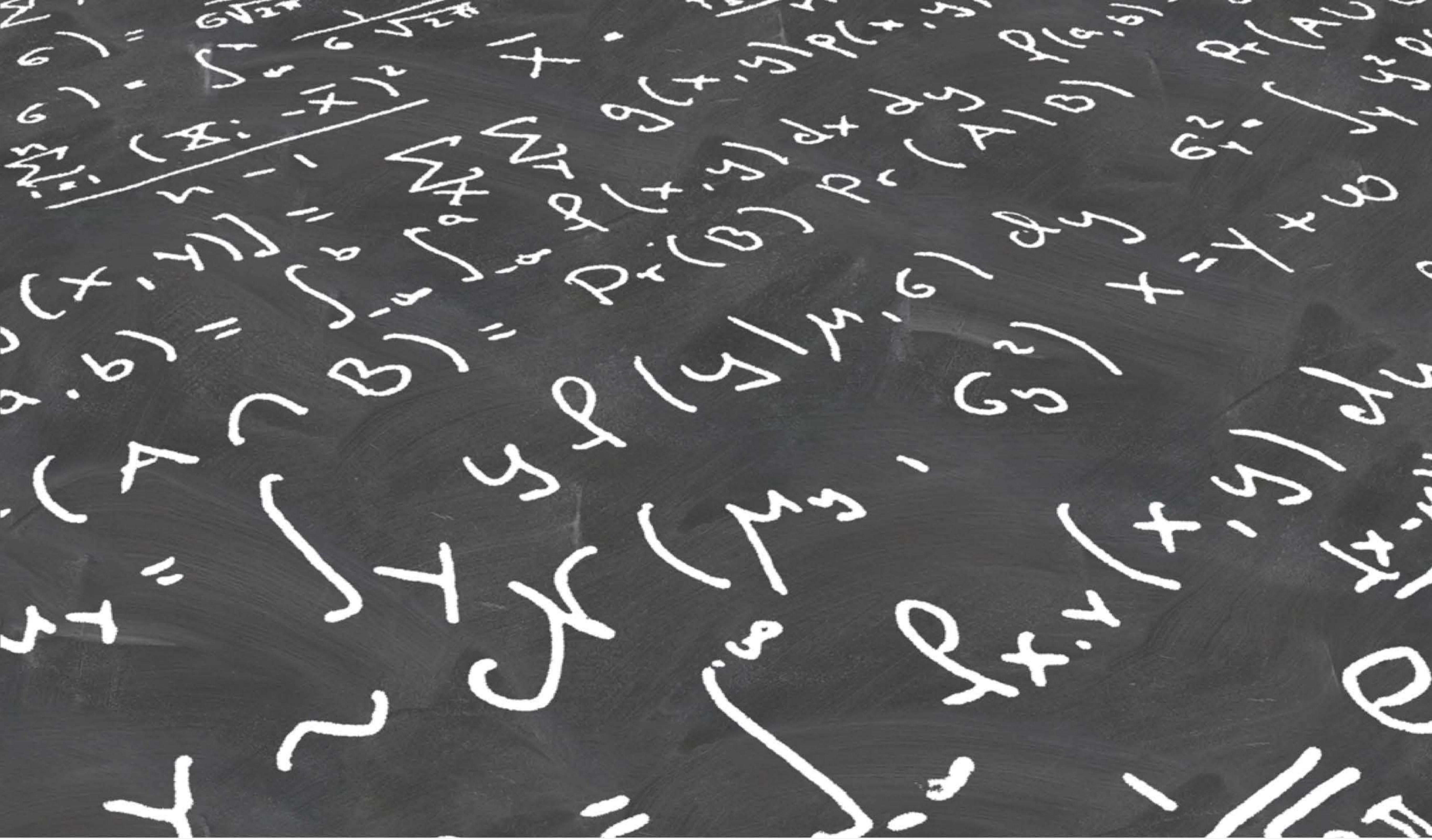


2. Probability Theory and Combinatorics

Gunvor Elisabeth Kirkelund
Lars Mandrup

Agenda for Today

- Repetition from last time
- Bayesian probability calculations and total probability
- Bernoulli trials
- Combinatorics
- An experiment

Basic Probability

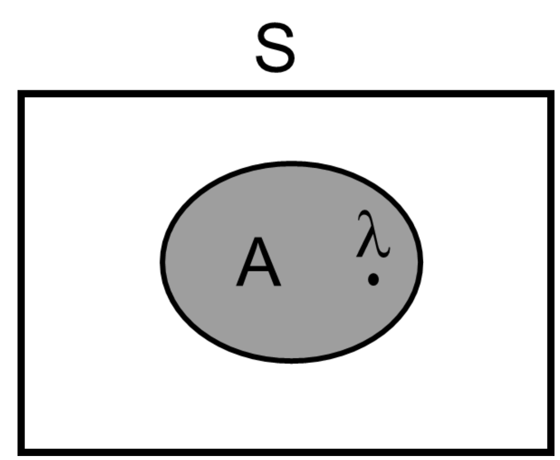
- Probability theory tells us what is in the sample given nature

- Basic Axions:

Axion 1: $\mathbf{0 \le \boldsymbol{Pr}(\boldsymbol{A}) \le 1}$

Axion 2: $\mathbf{\boldsymbol{Pr}(\boldsymbol{S}) = 1}$

S: Sample space
A: Event
λ: Sample point

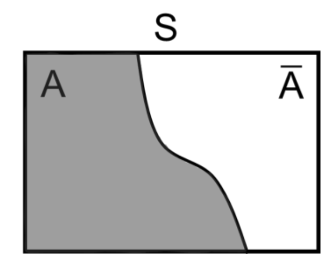


- Often (but not always) we use the relative frequency:

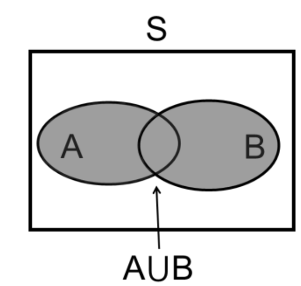
$$\Pr(A) = \frac{N_A}{N}$$

Basic Probability

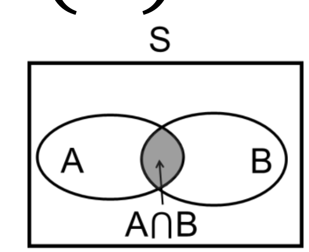
- Complement: $Pr(A) = 1 - Pr(\bar{A})$



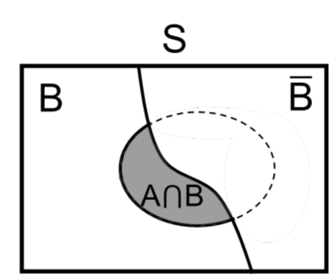
- Union: $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$



- Joint: $Pr(A \cap B) = Pr(A|B) \cdot Pr(B) = Pr(B|A) \cdot Pr(A)$



- Conditional: $Pr(A|B)$



Bayes Rule and Independence

- Bayes Rule:

$$\boxed{Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(B|A) \cdot Pr(A)}{Pr(B)}}$$

- A and B independent:

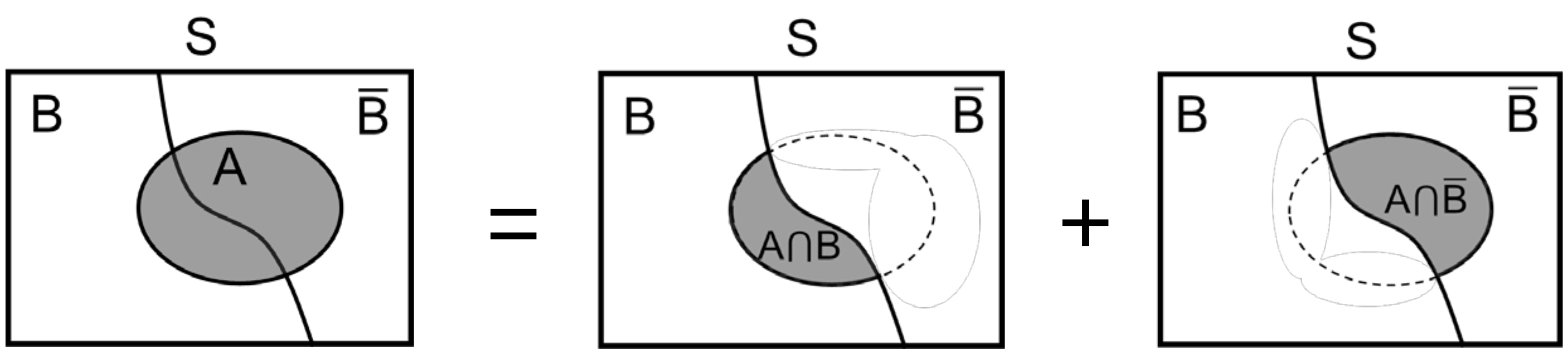
$$\boxed{Pr(A \cap B) = Pr(A) \cdot Pr(B)}$$

$$\boxed{Pr(B|A) = Pr(B) \quad \text{and} \quad Pr(A|B) = Pr(A)}$$

Total Probability

We sometime call it the marginal

- Pr(A) of an event is the total probability of that event.



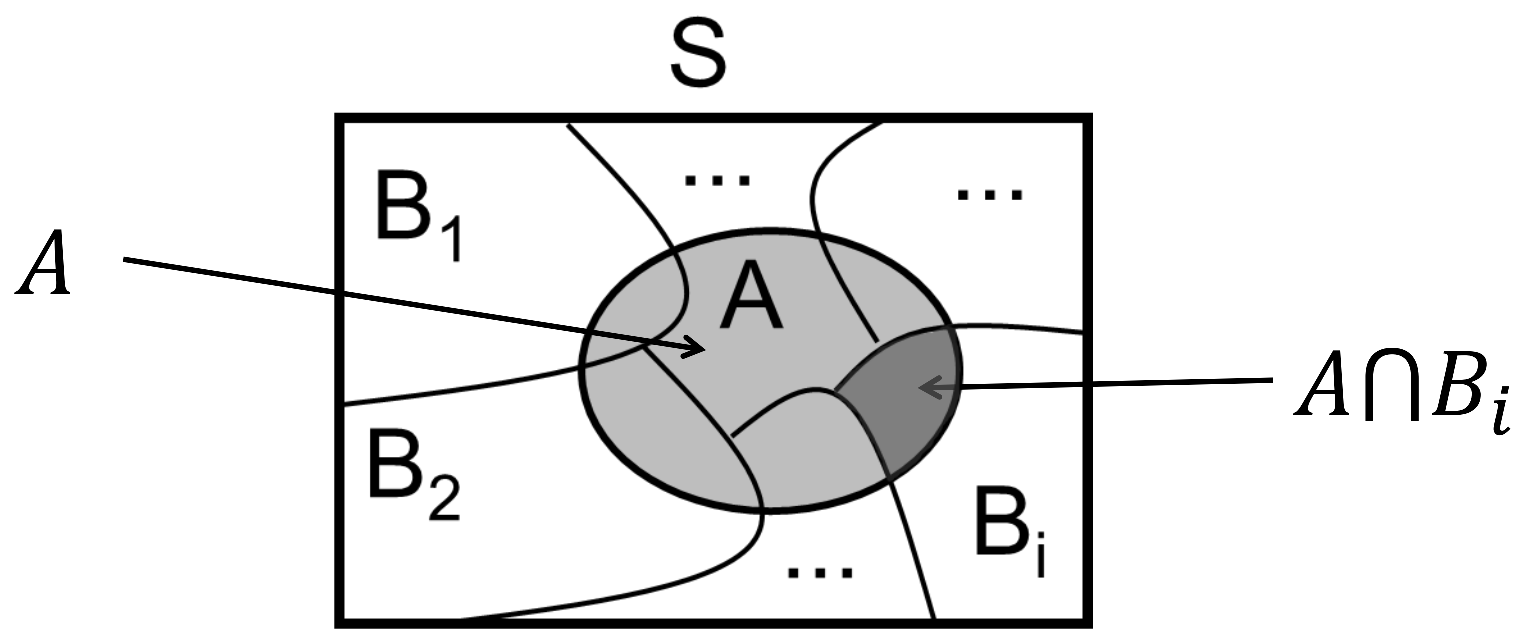
$$\Pr(A) = \Pr(A \cap B) + \Pr\left(A \cap \bar{B}\right)$$

$$= \Pr(A|B) \cdot \Pr(B) + \Pr\left(A|\bar{B}\right) \cdot \Pr(\bar{B})$$

Total Probability

We sometime call it the marginal

- Pr(A) of an event is the total probability of that event.



$$Pr(A) = Pr(A \cap B_1) + Pr(A \cap B_2) + \cdots + Pr(A \cap B_i) + \cdots$$

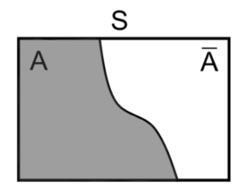
$$= Pr(A|B_1) \cdot Pr(B_1) + Pr(A|B_2) \cdot Pr(B_2) + \cdots$$

where the B_i's are mutually exclusive ($B_i \cap B_j = \varnothing$ for $i \neq j$)
and $S = B_1 \cup B_2 \cup \ldots \cup B_i \cup \ldots$

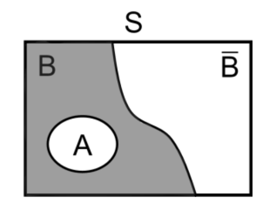
Summary of Probability

Relative frequency: $Pr(A) = \frac{N_A}{N_S}$

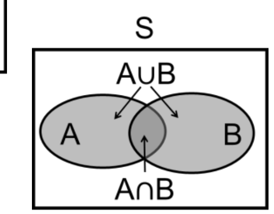
Complement: $Pr(\bar{A}) = 1 - Pr(A)$



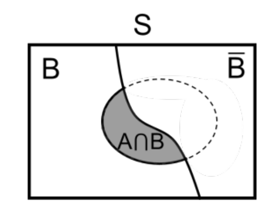
Exclusive: $Pr(\bar{A} \cap B) = Pr(B) - Pr(A)$ if $A \subset B$



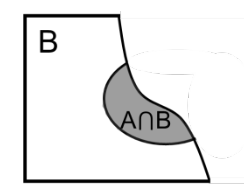
Union: $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$



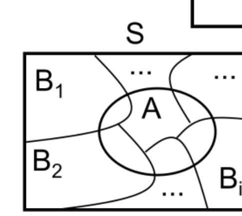
Joint: $Pr(A \cap B) = Pr(A|B) \cdot Pr(B) = Pr(B|A) \cdot Pr(A)$



Conditional: $Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$ if $Pr(B) \neq 0$



Total probability: $Pr(A) = \sum_{i=1}^{n} Pr(A|B_i) \cdot Pr(B_i)$



Bayes rule: $Pr(B|A) = \frac{Pr(A|B) \cdot Pr(B)}{Pr(A)}$

Bayes formula: $Pr(B_i|A) = \frac{Pr(A|B_i) \cdot Pr(B_i)}{\sum_{i=1}^{n} Pr(A|B_i) \cdot Pr(B_i)}$

Independence: $Pr(A \cap B) = Pr(A) \cdot Pr(B)$

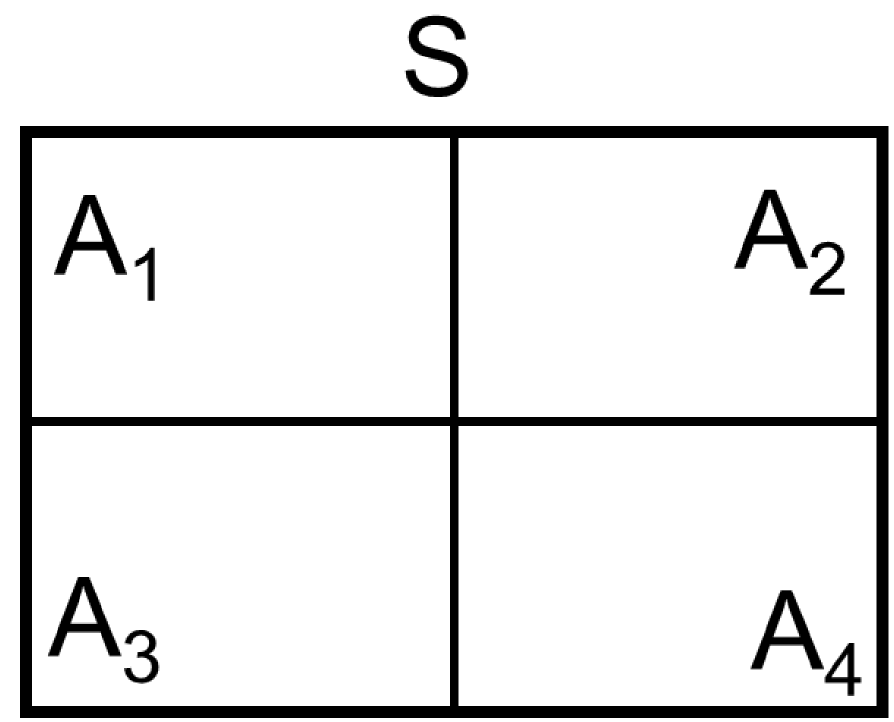
Orca Example

- In a conversation effort, we look for dead orcas when we are visiting an ocean.
- Given (conditioned) that <u>we have selected an ocean to examine</u>, how many males and females orcas will we observe?

Gender\ location	Atlantic (A_1)	Antartica (A_2)	Pacific (A_3)	Seaworld (A_4)
Female ($\bar{\mathbf{B}}$)	2	7	11	9
Male ($\mathbf{B}$)	8	3	1	19
Total	10	10	12	28

Orca Example (Cont'd)

- The probability selecting an ocean is identical.



- Event A_1 : Atlantic
- Event A_2 : Antartica
- Event A_3 : Pacific
- Event A_4 : Seaworld

$$Pr(A_1) = Pr(A_2) = Pr(A_3) = Pr(A_4) = \frac{1}{4}$$

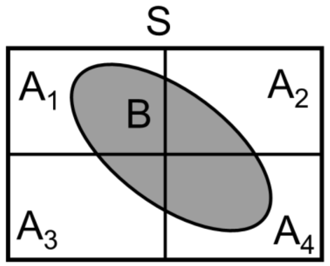
$$Pr(A_1) + Pr(A_2) + Pr(A_3) + Pr(A_4) = 1$$

The events $A_1 - A_4$ are mutually exclusive.

Orca Example Total Probability

- The event B, that the orca is a male, can then be written as:

$$B = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup (B \cap A_4)$$



- The total probability of a found killer whale, being a male, since event $A_1 - A_4$ are mutually exclusive (sum rule):

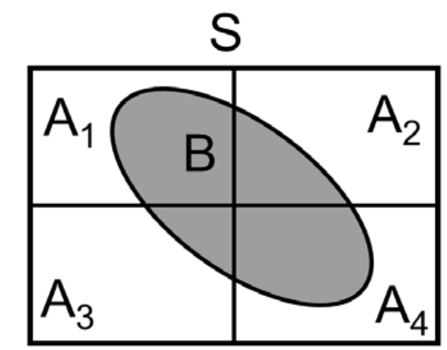
$$\boxed{Pr(B) = Pr(B \cap A_1) + Pr(B \cap A_2) + Pr(B \cap A_3) + Pr(B \cap A_4)}$$

- We rewrite with Bayes rule:

$$\begin{aligned} Pr(B) = {} & Pr(A_1)\, Pr(B|A_1) + Pr(A_2)\, Pr(B|A_2) \\ & + Pr(A_3)\, Pr(B|A_3) + Pr(A_4)\, Pr(B|A_4) \end{aligned}$$

Orca Example Cont'd

- Total Probability:



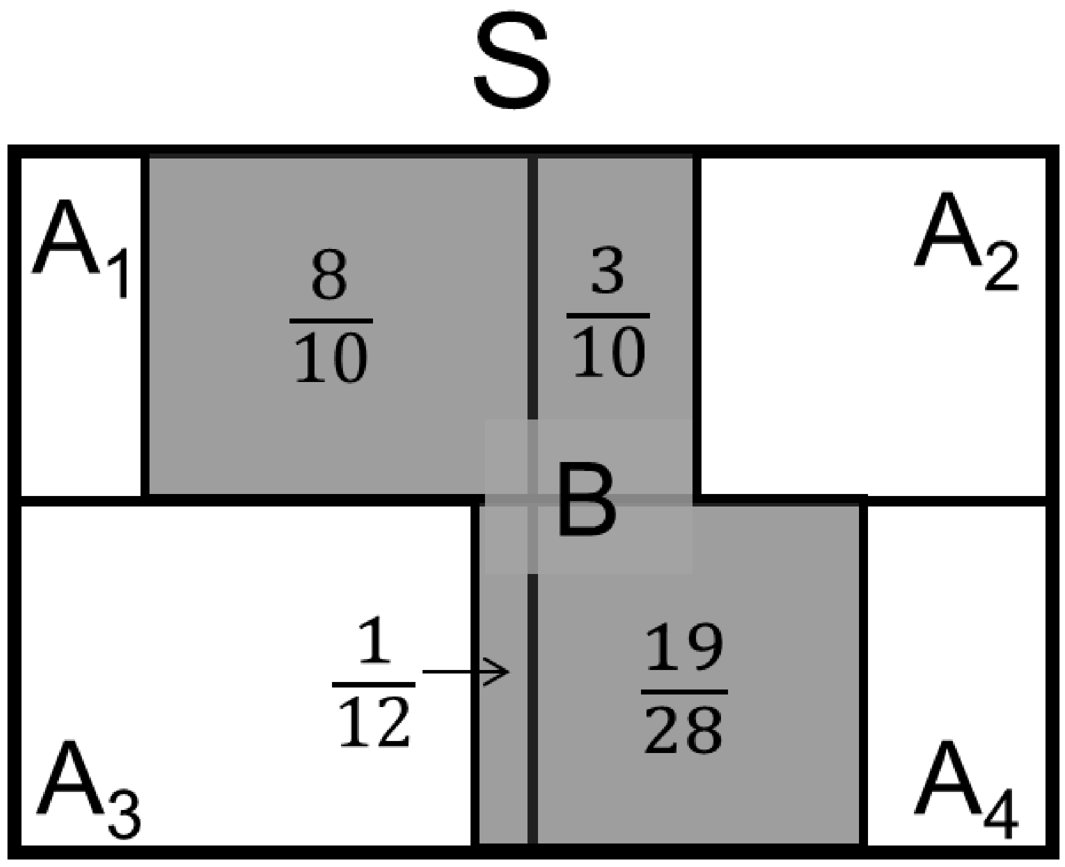
$$Pr(B) = Pr(A_1)\,Pr(B|A_1) + Pr(A_2)\,Pr(B|A_2)$$
$$+ Pr(A_3)\,Pr(B|A_3) + Pr(A_4)\,Pr(B|A_4)$$

Gender\ location	Atlantic (A_1)	Antartica (A_2)	Pacific (A_3)	Seaworld (A_4)
Female ($\bar{B}$)	2	7	11	9
Male (B)	8	3	1	19
Total	10	10	12	28

$$Pr(B) = \frac{8}{10}\cdot\frac{1}{4} + \frac{3}{10}\cdot\frac{1}{4} + \frac{1}{12}\cdot\frac{1}{4} + \frac{19}{28}\cdot\frac{1}{4} = 0,465$$

Orca Example Graphical

- We can also use a Graphical approach with Venn diagrams.



- The total probability of B is given by the marked area divided by the area of S.

Orca Example

- If an orca found is a male, what is the probability of us being in the Antartica?

$$Pr(A_2|B)$$

- We use Bayes rule:

$$Pr(A_2|B) = \frac{Pr(A_2 \cap B)}{Pr(B)} = \frac{Pr(B|A_2)Pr(A_2)}{Pr(B)}$$

- $Pr(B) = 0{,}47;\ \ Pr(A_2) = 0{,}25;\ \ Pr(B|A_2) = 0{,}3$

$$Pr(A_2|B) = \frac{Pr(B|A_2)Pr(A_2)}{Pr(B)} = \frac{0{,}3 \cdot 0{,}25}{0{,}47} = 0{,}16$$

Orca Example

- Is locations of the found orca independent of gender?
- How would you test it?

Gender\ location	Atlantic (A_1)	Antartica (A_2)	Pacific (A_3)	Seaworld (A_4)
Female ($\bar{\mathbf{B}}$)	2	7	11	9
Male ($\mathbf{B}$)	8	3	1	19
Total	10	10	12	28

$$Pr(A_2 \mid \bar{B}) = \frac{Pr(\bar{B}|A_2)Pr(A_2)}{Pr(\bar{B})} = \frac{0{,}7 \cdot 0{,}25}{1 - 0{,}47} = 0{,}33 \neq 0{,}16 = Pr(A_2|B)$$

Orca Example Conclusion

- **Prior**: What is the probability of us being in the Antartica?

$$Pr(A_2) = 0{,}25$$

- **Likelihood**: A tacked orca is found dead in Antartica, what is the probability of it being male?

$$Pr(B|A_2) = 0{,}3$$

- **Posterior**: A tacked orca whale is found dead and is a male, what is the probability of us being in Antartica?

$$Pr(A_2|B) = 0{,}16$$

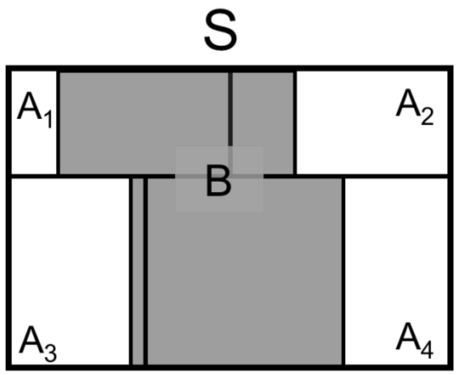
Orca Example – Another test method

- In a conversation effort, we pick up dead orcas from different oceans.
- The dead orcas are marked with the ocean and collected in the same container.
- A dead orca is randomly picked from the container: What is the probability that the orca is a male?

Gender\ location	**Atlantic (A_1)**	**Antartica (A_2)**	**Pacific (A_3)**	**Seaworld (A_4)**
Female ($\bar{B}$)	2	7	11	9
Male (B)	8	3	1	19
Total	10	10	12	28

Orca Example – Another test method

- Total Probability:



$$Pr(B) = Pr(A_1)\,Pr(B|A_1) + Pr(A_2)\,Pr(B|A_2) + Pr(A_3)\,Pr(B|A_3) + Pr(A_4)\,Pr(B|A_4)$$

Gender\ location	Atlantic (A_1)	Antartica (A_2)	Pacific (A_3)	Seaworld (A_4)	Total
Female ($\bar{\mathbf{B}}$)	2	7	11	9	29
Male ($\mathbf{B}$)	8	3	1	19	31
Total	10	10	12	28	60

$$Pr(B) = \frac{10}{60}\cdot\frac{8}{10}+\frac{10}{60}\cdot\frac{3}{10}+\frac{12}{60}\cdot\frac{1}{12}+\frac{28}{60}\cdot\frac{19}{28} = \frac{8+3+1+19}{60} = \frac{31}{60} = 0{,}517$$

Orca Example – Another test method

- If an orca found is a male, what is the probability that it is from the Antartica?

$$Pr(A_2|B)$$

- We use Bayes rule:

$$Pr(A_2|B) = \frac{Pr(A_2 \cap B)}{Pr(B)} = \frac{Pr(B|A_2)Pr(A_2)}{Pr(B)}$$

- $Pr(B) = 0{,}517;\ \ Pr(A_2) = 0{,}167;\ \ Pr(B|A_2) = 0{,}3$

$$Pr(A_2|B) = \frac{Pr(B|A_2)Pr(A_2)}{Pr(B)} = \frac{0{,}3 \cdot 0{,}167}{0{,}517} = \frac{3}{31} = 0{,}097$$

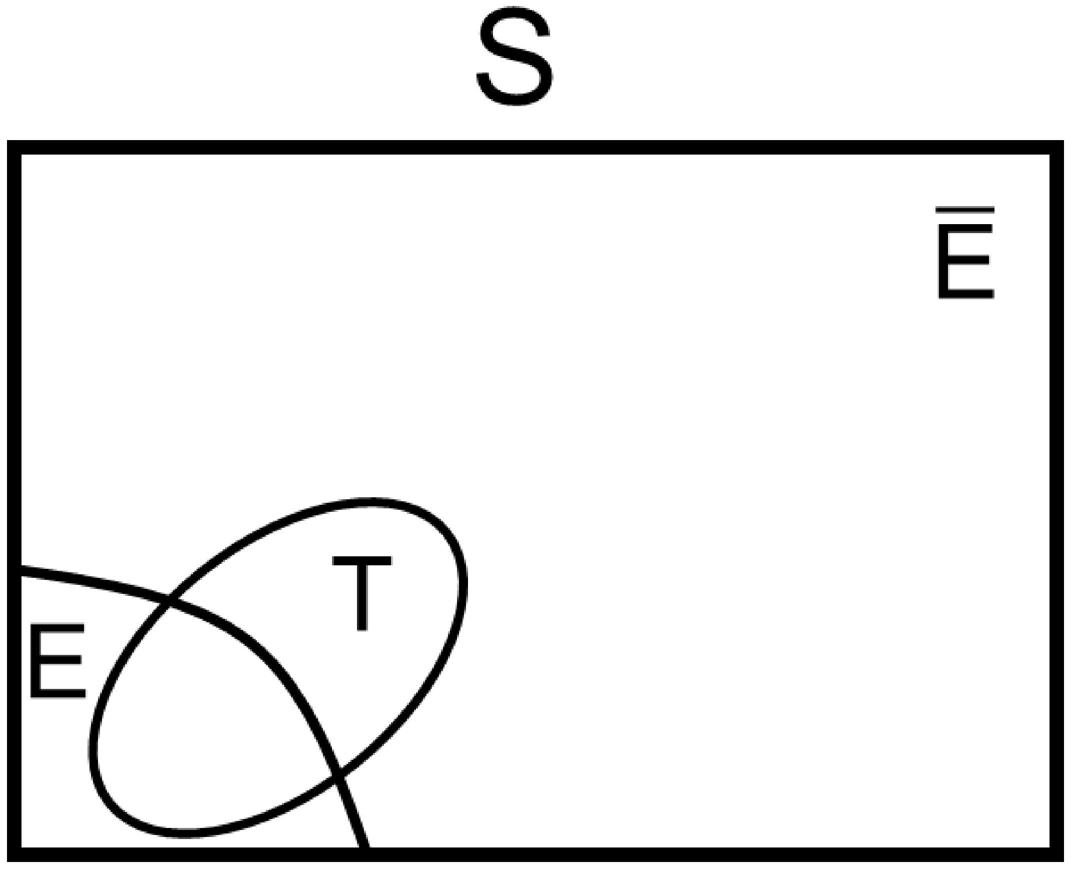
Tests and Types of Errors

- We can classify testing with two outcomes as:

Given / Result	**Disease (True)**	**No disease (False)**
Positive test	Sensitivity	Type I Error
Negative test	Type II Error	Specificity

Example: Ebola Test

- Event E: Patient are infectious with Ebola.
- Event T: The Ebola test is positive.



Example: Ebola Test

- **Prior**: What are the probability of a patient having Ebola?

$$Pr(E)$$

- **Likelihood**: What are the probability of a positive test given infectious with Ebola? Or of a negative test given not infectious with Ebola?

$$Pr(T|E) \quad \text{Sensitivity} \qquad\qquad Pr(\bar{T}|\bar{E}) \quad \text{Specificity}$$

- **Posterior**: What are the probability of being infectious given that a test is positive?

$$Pr(E|T)$$

Example: Ebola Test — Total Probability

- **Prior**: What are the probability of a patient having ebola?

Complement of E

$$Pr(E) = 0,01 \qquad Pr(\bar{E}) = 1 - 0,01 = 0,99$$

- **Likelihood**: What are the probabilities of the tests?

$$Pr(T|E) = 0,9$$ Sensitivity

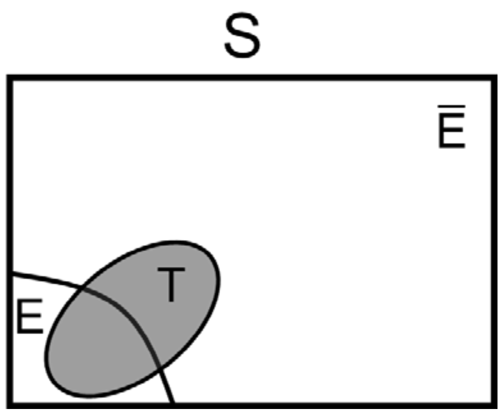
$$Pr(\bar{T}|\bar{E}) = 0,8$$ Specificity

- **Complement**: What are the probability of a patient having a positive test without being infectious?

$$Pr(T|\bar{E}) = 1 - Pr(\bar{T}|\bar{E}) = 0,2$$

Example: Ebola Test — Total Probability

- **Total Probability with the Sum Rule**: What are the probability of a patient having a positive test?



$$Pr(T) = Pr(T \cap E) + Pr(T \cap \bar{E})$$

- **The Product Rule**: We can with Bayes rule find

$$Pr(T) = Pr(T|E)\,Pr(E) + Pr(T|\bar{E})\,Pr\left(\bar{E}\right)$$

$$= 0{,}9 \cdot 0{,}01 + 0{,}2 \cdot 0.99$$

$$= 0{,}207$$

Ebola Example — Posterior

- **We have**: We now know the probabilities:

$$P(E) = 0,01$$ Prior

$$P(T) = 0,207$$ Total probability

$$P(T|E) = 0,9$$ Likelihood

- **Product Rule**: What are the probability of being infectious given that a test is positive?

Bayes rule

$$Pr(E|T) = \frac{Pr(T|E)Pr(E)}{Pr(T)} = \frac{0,9 \cdot 0,01}{0,207} = 0,043$$

Ebola Example — Posterior

- What are the probability of being infectious given that a test is positive?

$$Pr(E|T) = \frac{Pr(T|E)Pr(E)}{Pr(T)} = \frac{0{,}9 \cdot 0{,}01}{0{,}207} = 0{,}043$$

- What are the probability of <u>not</u> being infectious given that a test is positive?

$$Pr(\bar{E}\,\big|\,T) = 1 - Pr(E|T) = 0{,}957$$

- What are the probability of <u>not</u> being infectious given a negative test?

$$Pr\left(\bar{E}\middle|\bar{T}\right) = \frac{Pr(\bar{T}\,\big|\,\bar{E})Pr(\bar{E})}{Pr(\bar{T})} = \frac{0{,}8 \cdot 0{,}99}{0{,}793} = 0{,}999$$

- What are the probability of being infectious given that a test is negative?

$$Pr(\mathrm{E}\,\big|\,\bar{T}) = 1 - Pr\left(\bar{E}\middle|\bar{T}\right) = 0{,}001$$

Ebola Example — Conclusion

- If the test is negative, it is allmost certain (99,9%) that you're not being infectious:

$$Pr\left(\bar{E}\middle|\bar{T}\right) = 0{,}999$$

- If the test is positive, there is still only a small risk (4,3%) that you actually are being infectious:

$$Pr(E|T) = 0{,}043$$

Monty Hall Dilemma

- We have three doors
- Behind two of the doors is a goat
- Behind one door is a million dollars ($)

- What is the chance of guessing behind which door the money is?

$$\Pr(\$|1) = \Pr(\$|2) = \Pr(\$|3) = \frac{1}{3}$$

Monty Hall Dilemma cont’d

- We make a selection of a door, say door 2, without open it.
- The quizmaster eliminates one of the doors ($\overline{\$}$) ,which we did not select, based on his knowledge on the goat situation, say door 1.
- We can now reselect between door 2 and 3.
- What are the probabilities of the money being behind the two doors? Should we switch door?

Monty Hall Dilemma cont’d

- What are the probabilities of the money being behind the two doors? Should we switch door?

The Binomial Distribution

- We have n repeated trials.
- Each trial has two possible outcomes

Bernoulli trial

  - **Success** — probability p
  - **Failure** — probability q=1-p

- What is the probability of having k successes out of n trials?
- We write this question as:

$$\boxed{Pr_n(k) = \frac{n!}{k!\,(n-k)!} p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}}$$

- Faculty: $n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$

  $0! = 1$

Bernoulli Trial

Definition: The binomial coefficient is defined as:

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}$$

Number of ways to select k objects out of a collection of n objects

Example: Out of 10 children, what is the probability that exactly 2 are girls?

$$Pr_n(k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

$$= \frac{10!}{2!(10-2)!} (0,5)^2 (1-0,5)^{10-2} = 0,044$$

Combinatorics

- Take an object from a collection of n objects.
- Repeat the test k times.

Types of Experiments:

- With or without replacement
- Ordered or unordered

Example:

What is the probability that if I have two children that the oldest is a girl and the youngest is a boy?

- Ordered.
- With replacement.

Ordered with Replacement

- Take an object from a collection of n objects.
- **Put it back** each time.
- Repeat the test k times.
- **The sequence** of the objects **matters**.

- The number of combinations is: n^k
  - Each trial has n possible outcomes
  - All the trials are independent

Ordered without Replacement

- Take an object from a collection of n objects.
- **Do not** put it back each time.
- Repeat the test k times.
- **The sequence** of the objects **matters**.

- The number of combinations is:

$$_nP_k = P_k^n = \frac{n!}{(n-k)!} = \mathrm{n} \cdot (n-1) \ldots (n-k+1)$$

- ➢ The 1st trial has n possible outcomes, the 2nd trial has n-1 possible outcomes, … , the k'th trial has n-k+1 possible outcomes

Unordered without Replacement

- Take an object from a collection of n objects.
- **Do not** put it back each time.
- Repeat the test k times.
- **The sequence** of the objects **do not matter**.

- The number of combinations is:

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}$$

  - The k ordered draws can be shuffled in k! different ways (sequences)

Unordered with Replacement

- Take an object from a collection of n objects.
- **Put it back** each time.
- Repeat the test k times.
- **The sequence** of the objects **do not matter**.

- The number of combinations is:

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}$$

> ➢ Each time we draw an object, we should replace an object (except for the last draw). This correspond to we start with n+k-1 object and draw k objects unordered without replacement.

Summary of Combinatorics

- We can summarise the number of possible outcomes of k trials, sampled from a set of n objects.

		Replacement	
		With	Without
Sampling	Ordered	n^k	$P_k^n = \frac{n!}{(n-k)!}$
	Unordered	$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}$	$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$

Experiment: Birthday Example

- k=35 students
- n=365 (number of days in the year)
- What are the probability that at least two have birthday on the same day (E)?

All have different birthdays

Ordered sampling without replacement (k unique birthsdays in n days)

Complement rule

$$\Pr(E) = 1 - \Pr(\bar{E}) = 1 - \frac{\frac{n!}{(n-k)!}}{n^k} = 1 - \frac{\frac{365!}{(365-35)!}}{365^{35}} > 80\%$$

Ordered sampling with replacement
(all possible combinations of k students birthdays in n days)

- k=50 students: Pr(E)>97%
- k=75 students: Pr(E)>99,97%

Words and Concepts to Know

Prior

Type I Error

Binomial coefficient

Sampling

Unordered

Replacement

Specificity

Likelihood

Combinatorics

Bernoulli Trial

Sensitivity

Posterior

Ordered

Type II Error

Binomial distribution