5LN445 Mathematics for Language Technologists

Q & A, Exam prep

Today

Learning outcomes

Brief review of the content

- Set theory
- Functions
- Foundations of probability
- Conditional probability
- Independence
- Total probability theorem
- Bayes' theorem
- Combinatorics
- Distributions

Solve old exams

Exam on Wednesday

This level of fancy is enough.

Room 16-0043 (Still, check your schedules!)

You bring:
Calculator
Pen
Notes written by hand

You will be given:
Probit table

Remote participants must have their cameras turned on and stay connected during the exam.

The exam will cover sets, functions and probability theory. Max 40 points.

Learning outcomes

Sets: discuss and apply elementary concepts in set theory such as subset, intersection, union, relation, and function. Readings: Slides.

- Sets
  - Important examples: the empty set, finite sets, infinite sets of numbers.
  - Set comparison: equality, subset, proper subset.
  - Set operations: union, intersection, set difference, complement, cardinality.
  - Set builder notation.
- Functions
  - Domain and codomain.
  - Function composition
  - Recursive functions on natural numbers

Probability: discuss and apply elementary concepts in probability theory such as unconditional and conditional probability, Bayes' theorem, and the law of total probability. Readings: OpenIntro Statistics chapter 3.

- Probability Space
  - Sample Space, Doman and Events
  - Law of Large Numbers
  - General Addition Rule
  - Tree Diagrams
  - Basic combinatorics
  - Independence
  - Conditional Probability
  - Bayes' theorem
  - Theorem of Total Probability
- Random variables
  - Discrete/Continuous Random Variables
  - Probability/Distribution Functions
  - ~~Mean, Median, Expectation and Variance~~
- Random process

Set operations

$A = \{a, b, c, d\}$, $B = \{c, d, e, f, g\}$, $C = \{a, b\}$

Subset: $A \subseteq B = \{x \in A \mid x \in B$ **and** $|A| \leq |B|\}$
Ex: $A \subseteq A$, $C \subset A$

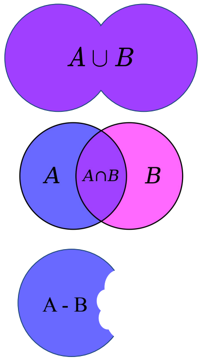
Union: $A \cup B = \{x \mid x \in A$ **or** $x \in B\}$
Ex: $A \cup B = \{a, b, c, d, e, f, g\}$

Intersection: $A \cap B = \{x \mid x \in A$ **and** $x \in B\}$
Ex: $A \cap B = \{c, d\}$

Difference: $A - B = \{x \mid x \in A$ **and not** $x \in B\}$
Ex: $A - B = \{a, b\}$

Cardinality: $|A|$ is the number of elements in set A
Ex: $|A| = 4$, $|B| = 5$

Set builder: $\{x \mid$ condition for $x\}$



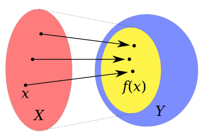
Elements/Members in a set are **always** unordered and unique.

Functions

Notation

A function called f takes elements from its domain A and maps them to elements in its co-domain B: "f: A → B"

Ex: $g: \mathbb{N} \rightarrow \mathbb{R}$, $g(x) = 1/x$



$f: X \rightarrow Y$

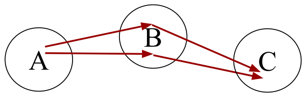
Composition of functions

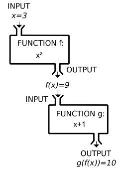
$f: A \rightarrow B$

$g: B \rightarrow C$

$g \circ f: A \rightarrow C$

$(g \circ f)(x) = g(f(x))$





If $f(x) = x^2$ and $g(x) = x + 1$ then $(g \circ f)(x) = x^2 + 1$

Functions

Sigma notation for summation

$\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4$

$\sum_{k=5}^{7} f(k) = f(5) + f(6) + f(7)$

$\sum_{x=1}^{3} x^2 = 1^2 + 2^2 + 3^2 = 14$

Absolute value

$$|x| = \begin{cases} x & ,x \geq 0 \\ -x & ,x < 0 \end{cases}$$

$|-1|=1$, $|1|=1$

Pi notation for multiplications

$\prod_{i=1}^{3} x_i = x_1 \cdot x_2 \cdot x_3$

$\prod_{x=1}^{3} x^2 = 1^2 \cdot 2^2 \cdot 3^2 = 1 \cdot 4 \cdot 9 = 36$

$\prod_{n=1}^{3} x^n = x^1 \cdot x^2 \cdot x^3 = x^6$

Recursive function (only over $\mathbb{N}$ in the course)

$$f(n) = \begin{cases} 1 & ,n = 1 \\ 1 & ,n = 2 \\ f(n-1) + f(n-2) & ,n \geq 2 \end{cases}$$

$$f(n) = \begin{cases} 1 & ,n = 1 \\ f(n-1) \cdot n & ,n > 1 \end{cases}$$

$$f(n) = \begin{cases} 1 & ,n = 1 \\ f(n-1) + n^2 & ,n \geq 1 \end{cases}$$

https://en.wikipedia.org/wiki/Multiplication#Capital_pi_notation
https://en.wikipedia.org/wiki/Function_composition

Foundations

The *sample space* Ω is the set of all possible *outcomes*. Their total probability add up to 1 (i.e. P(Ω)=1).

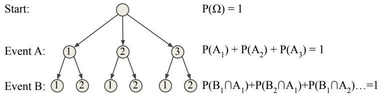
An *event* is a non-empty subset of outcomes in Ω. The of all events is called F.

The *probability measure* P is a function that assigns a probability value (between 0 and 1) to and event as:

$P: F \rightarrow \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$

Special case: If all elements of Ω are equally likely to occur, then the probability of some event A is $P(A) = |A|/|\Omega|$

A tree diagram of possible outcomes is called an *event tree*. The probability start at 1 and "flows" down the tree.



Two events A and B are independent when:

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

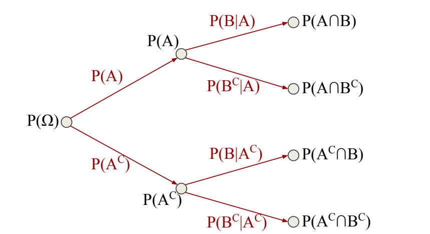
(Knowing that A happened says nothing about B)

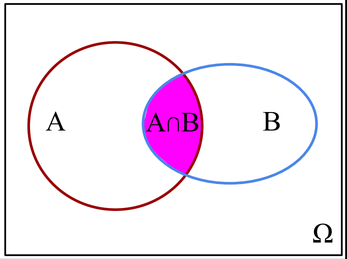
General addition rule:

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

Conditional probability

Let A and B be arbitrary events in a given probability space, with $P(B) > 0$. Then we define the conditional probability of A, given B, as: $P(A|B) = P(A \cap B)/P(B)$





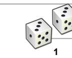
$P(\Omega)P(A)P(B|A) = P(A \cap B)$ and $P(\Omega)P(A^C)P(B|A^C) = P(A^C \cap B)$

Solution algorithm

1. Find the minimum number of variables for describing the outcomes of the problem
2. Define a sample space
3. Define a set of events
4. What do we know?
5. What do we want to know?
6. Link points 4 and 5 using theorems and assumptions.

From the lab:

1. Die 1 and die 2, $T = \{x \in \mathbb{N} | x \leq 6\} = \{1, 2, 3, 4, 5, 6\}$
2. $\Omega = T^2 = \{(1, 1), (1, 2), (1, 3), \dots\}$
3. $Y \in F$, $Y = \{(a, b) \in \Omega \mid 3 \leq |a - b|\}$ *(Yoshi wins)*
   $Y^C \in F$ *(Peach wins)*
4. $P(Y^C) = 1 - P(Y)$
5. P(Yoshi wins)
6. $|\Omega|=36$, $|Y|=12$, assume fair dice
   $P(Y^C) = 1 - P(Y) = 1-|Y|/|\Omega|=24/36=2/3$



	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

Solution algorithm

1. Find the minimum number of variables for describing the outcomes of the problem
2. Define a sample space
3. Define a set of events
4. What do we know?
5. What do we want to know?
6. Link points 4 and 5 using theorems and assumptions.

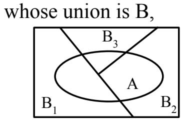
From every statistics book:

2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣	A♣
2♢	3♢	4♢	5♢	6♢	7♢	8♢	9♢	10♢	J♢	Q♢	K♢	A♢
2♡	3♡	4♡	5♡	6♡	7♡	8♡	9♡	10♡	J♡	Q♡	K♡	A♡
2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠	A♠

1. Number, Colour
2. $\Omega = \{x \in \mathbb{N} \mid 1 \leq x \text{ and } x \leq 13\} \times \{♣, ♠, ♦, ♥\} = \{(1, ♣), (2, ♣), \dots\}$
3. $A = \{x \in \Omega \mid x_2 = ♥\} = \{(1, ♥), (2, ♥), (3, ♥), (4, ♥), \dots\}$ *(hearts)*
   $B = \{x \in \Omega \mid x_1 \geq 11\} = \{(11, ♣), (12, ♣), (13, ♣), (11, ♠), \dots\}$ *(face cards)*
4. $|\Omega| = 52$, $|A| = 13$, $|B| = 12$
5. $P(A|B)$
6. $P(A|B) = P(A \cap B)/P(B) = (|A \cap B|/|\Omega|)/(|B|/|\Omega|) = 3/12 = 1/4$, assumption: each card is equally likely

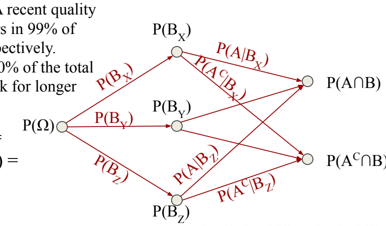
Theorem of Total Probability

If $B_1, B_2, \dots, B_n$ are mutually exclusive events with nonzero probabilities, whose union is B, and A is any event, then:



$$P(A \cap B) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n)$$

A company is making light bulbs in three factories (X, Y and Z). A recent quality inspection showed that Factory X's bulbs work for over 5000 hours in 99% of cases. The same figure is 95% and 97% for factories Y and Z, respectively. Factory X supplies 50%, Y supplies 30% and factory Z supplies 20% of the total bulbs available. What is the chance that a purchased bulb will work for longer than 5000 hours?



$P(A \cap B) = P(A|B_X)P(B_X)+P(A|B_Y)P(B_Y)+P(A|B_Z)P(B_Z) =$
$= (99/100) \cdot (5/10)+(95/100) \cdot (3/10)+(97/100) \cdot (2/10) =$
$= 974/1000 = 97.4\%$

Note: $P(B_X) + P(B_Y) + P(B_Z) = 1$, $P(A) + P(A^C) = 1$

https://en.wikipedia.org/wiki/Law_of_total_probability

Bayes' Theorem

If A is any event with P(A) > 0 and B_1, B_2 , … , B_n are mutually exclusive events with nonzero probabilities, whose union is Ω or contains A, then

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+...+P(A|B_n)P(B_n)}$$

A blood test, when given to a person who might have a certain disease, is right in 99% of cases. What is the probability that a person really has the disease if the test says so given that its prevalence is 0.1% in the population?

B = "the person has the disease" P(A|B) = .99 $P(A|B^C)$ = .01 *(performance symmetry)*

B^C = "the person *does not* have the disease" P(B) = .001 $P(B^C)$ = .999

A = "the test gives a positive result."

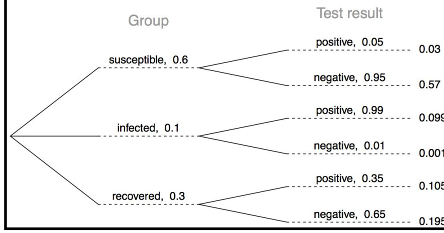
$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^C)P(B^C)}$$

$$= \frac{.99 \cdot 0.001}{(.99 \cdot 0.001)+(0.01 \cdot 0.999)} = \frac{0.00099}{0.00099+0.00999} \approx 0.09$$

P(B|A) is unknown

Inverting probabilities

Imagine a population in the midst of an epidemic where 60% of the population is considered susceptible, 10% is infected, and 30% is recovered. The only test for the disease is accurate 95% of the time for susceptible individuals, 99% for infected individuals, but 65% for recovered individuals. If the individual has tested positive, what is the probability that they are actually infected?



Inverting probabilities

Imagine a population in the midst of an epidemic where 60% of the population is considered susceptible, 10% is infected, and 30% is recovered. The only test for the disease is accurate 95% of the time for susceptible individuals, 99% for infected individuals, but 65% for recovered individuals. If the individual has tested positive, what is the probability that they are actually infected?

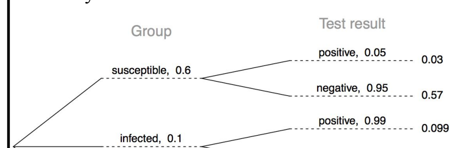
$$P(inf|+) = \frac{P(inf \cap +)}{P(+)}$$

$$= \frac{P(inf \cap +)}{P(sus \cap +)P(inf \cap +)P(rec \cap +)}$$

$$= \frac{0.099}{0.03+0.099+0.105} \approx 0.423$$

Inverting probabilities

Imagine a population in the midst of an epidemic where 60% of the population is considered susceptible, 10% is infected, and 30% is recovered. The only test for the disease is accurate 95% of the time for susceptible individuals, 99% for infected individuals, but 65% for recovered individuals. If the individual has tested positive, what is the probability that they are actually infected?



$$P(inf|+) = \frac{P(inf \cap +)}{P(+)}$$

$$= \frac{P(inf \cap +)}{P(sus \cap +)P(inf \cap +)P(rec \cap +)}$$

$$= \frac{0.099}{0.03+0.099+0.105} \approx 0.423$$

$$P(inf|+) = \frac{P(+|inf)P(inf)}{P(+)} = \frac{P(+|inf)P(inf)}{P(+|sus)P(sus)+P(+|inf)P(inf)+P(+|rec)P(rec)}$$

Combinatorics

Example given that Ω = {A, B, C, D, E}, n=5 k=2

Permutations from a set

Permutations *(ordered)* of all elements from a set of n

$n!$

$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$

Permutations from a subset

Permutations *(ordered)* of k elements from a set of n

$\frac{n!}{(n-k)!}$

$\frac{5!}{(5-2)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1} = 5 \cdot 4 = 20$

Combinations from a subset

Combinations *(unordered)* of k elements from a set of n, "n choose k"

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

$\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1)(2 \cdot 1)} = \frac{5 \cdot 4}{2 \cdot 1} = \frac{20}{2} = 10$

n · (n-1) · (n-2) … 4 · 3 · 2 · 1 = n! *(n factorial)*

Ex: 3!=6, 2!=2, 1!=1, 0!=1 *(note: 0!=1)*

Examples

Given the standard english alphabet, how many 3 letter words can be formed? (Ignoring limitations due to possible pronunciation)

C = {a, b, c, d, …, z}

$|C^3| = |C|^3$

$|C| = 26 \rightarrow |C^3| = 17576$

How many different three-card hands can be drawn from a standard deck of cards?

$\binom{n}{k} = \frac{n!}{k!(n-k)!}, n = 52, k = 3$

$\frac{52!}{49!3!} = \frac{52 \cdot 51 \cdot 50}{6} = 22100$

In some group there are 30 men and 20 women. In how many ways can a committee of two men and two women be chosen?

$\binom{n_1}{k_1}\binom{n_2}{k_2}, n_1 = 30, k_1 = 2, n_2 = 20, k_2 = 2$

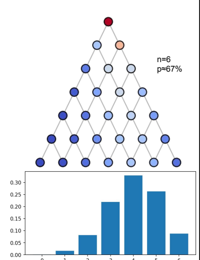
$\frac{30!}{28!2!}\frac{20!}{18!2!} = \frac{30 \cdot 29}{2}\frac{20 \cdot 19}{2} = 82650$

Discrete distributions

Discrete random variable: Outputs are separate and distinct.
Continuous random variable: Outputs are *not* separate and distinct.

Probability mass function (PMF)
"What is the probability of getting this realisation k?"
PMF ⇔ Probability measure P from earlier, P(X=k)
(Note: The probabilities of all outcomes must still sum to 1, i.e. $\Sigma_K P(X=k) = 1$)

Cumulative distribution function (CDF)
"What is the probability of getting k or something below k?"
Discrete CDF ⇔ $P(X \leq k)$, i.e. $F(k) = \Sigma f(k)$ *(sum up to, and including, k)*



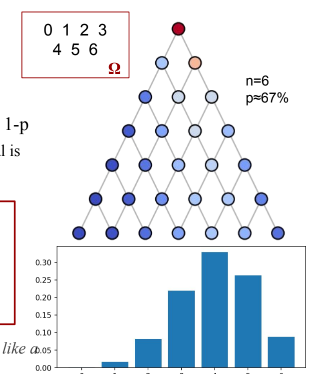
Note that the CDF (but not the PMF) requires an ordering of the outcomes. Why?

Binomial distribution

$\binom{n}{k} p^k (1-p)^{n-k}$

n choose k	*The number of "chains" ending here*
p^k	k number of "positives" with probability p
$(1-p)^{n-k}$	n-k number of "negatives" with probability 1-p

We *don't care* about the underlying process anymore. Each game/trial is modelled as a Bernoulli trial.



For a Bernoulli trial, the following must hold:

- The trials are independent.
- The number of trials, n, is fixed.
- Each trial outcome is binary (e.g. "success" or "failure").
- The probability of a success, p, is the same for each trial.

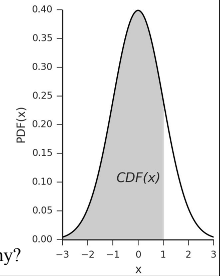
Note that at higher n, the binomial distribution looks more and more like a bell curve.

Continuous distributions

Discrete random variable: Outputs are separate and distinct.
Continuous random variable: Outputs are *not* separate and distinct.

Probability <u>density</u> function (P<u>D</u>F)
"What is the <u>likelihood</u> of getting this realisation <u>x</u>?"
PDF ⇔ Probability measure P from earlier, P(X=x)
(Note: The probability must sum to 1, i.e. ∫P(X=x)dx = 1)

Cumulative distribution function (CDF)
"What is the probability of getting x or something below x?"
Continuous CDF ⇔ P(X < x), i.e. F(k) = Σ f(k) *(sum up to, and including, k)*



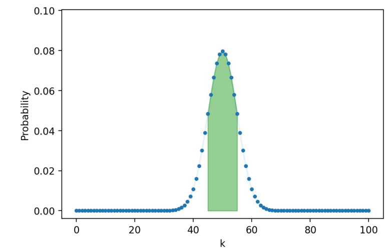
Note that both the CDF and the PDF requires an ordering of the "outcomes". Why?

Normal Approximation with Continuity Correction

We toss a fair coin n = 100 times. Letting s denote the number of heads obtained, find the normal approximation to P(45 ≤ X ≤ 55).

p = ?, n = ?, k = ?
μ = ?, σ = ?

$$P(a \leq X \leq b) \approx \Phi(\frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}) - \Phi(\frac{a-\frac{1}{2}-np}{\sqrt{np(1-p)}})$$



http://onlinestatbook.com/2/calculators/binomial_dist.html

Normal Approximation with Continuity Correction

We toss a fair coin n = 100 times. Letting s denote the number of heads obtained, find the normal approximation to P(45 ≤ X ≤ 55).

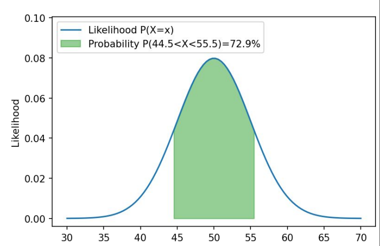
p = .5, n = 100, k = 45 to 55, X ~ Binom(n, p)
μ = np = 50, σ = sqrt(np(1-p)) = 5

$$P(a \leq X \leq b) \approx \Phi(\frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}) - \Phi(\frac{a-\frac{1}{2}-np}{\sqrt{np(1-p)}})$$

$$P(45 \leq X \leq 55) \approx \Phi(\frac{55+\frac{1}{2}-50}{5}) - \Phi(\frac{45-\frac{1}{2}-50}{5})$$

= Φ(1.1) - Φ(-1.1) = Φ(1.1) - (1 - Φ(1.1)) = ?

Binom CDF: $\sum_{k=45}^{55} \binom{100}{k} 0.5^{100} \approx 0.7287$



http://onlinestatbook.com/2/calculators/binomial_dist.html

Old exams

Note that if you have:

1. Read
   a. Chapters 3.1-3.5, 4.1, 4.3
   b. The slides (mostly L7, i.e. today's)
2. Completed the exercises
   a. Chapters 3.1-3.5, 4.1, 4.3
   b. Exercise 1
   c. Last year's exams (dugga 1)

…then nothing on the exam will be new to you.