UPPSALA UNIVERSITY
Fredrik Wahlberg and Mats Dahllöf

5LN445 Mathematics for Language Technologists
7 December 2018

# Exam

This exam is split up into two parts, each representing one half of the course. If you passed dugga 1, you should skip part 1 of this exam. If you passed dugga 2, you should skip part 2 of this exam. Pass: at least 60% correct (per part), point will be normalised to account for the respective exams difference in max points. Pass with distinction: at least 80% correct on the two parts (or duggas) added, again normalised. When in doubt about the interpretation of a question, make reasonable assumptions and motivate those. If you get stuck on a task, try to solve other tasks first, then go back. Please read the whole exam before beginning.
**Tools:** Calculator, probit table (included) and table of equations (included).

## Part 1: Logic, sets and probability

**Predicate calculus (10p)**

1. Show, using a truth table, the truth values of the basic logic operations. (6p)

    a) $A \land B$
    b) $A \lor B$
    c) A xor B
    d) $\neg A$
    e) $A \rightarrow B$
    f) $A \leftrightarrow B$

| A | B | $(A \land B)$ | $(A \lor B)$ | (A xor B) | $\neg A$ | $A \rightarrow B$ | $A \leftrightarrow B$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |

2. For some expressions, the order of evaluation matters e.g. “a-(b+c)” is not equal to “(a-b)+c”. Does the order matter when evaluating a sequence of xor operations? For the expression “A xor B xor C”, show with a truth table, whether the following two statements are logically equivalent (i.e. show that the two ways of evaluating “A xor B xor C” give the same outcomes). (4p)
    (A xor B) xor C
    A xor (B xor C)

| A | B | C | (A xor B)(1) | (1) xor C | (B xor C) (2) | A xor (2) |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | 0 | 1 |
| 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| 1 | 0 | 1 | 1 | 0 | 1 | 0 |
| 1 | 0 | 0 | 1 | 1 | 0 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 | 1 | 1 | 1 |
| 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Set theory (8p)**

3. For sets A={x∈ℕ| 1 ≤ x ≤ 3} and B={a, b, c}, show the results of the following statements. (8p)
   a) A
   b) A ∪ B
   c) |A ∪ B|
   d) |A| + |B| - |A ∩ B|
   e) A × B
   f) P(A), *i.e. the power set of A*
   g) A ∩ B
   h) |P(A)|

a) A = {1, 2, 3}
b) A ∪ B= {1, 2, 3, a, b, c}
c) |A ∪ B| = 6
d) |A| + |B| - |A ∩ B| = 6
e) A × B = {(1, a), (1, b) … (3, c)}
f) P(A) = {∅, {1}, [{2}, {3}, {1,2} … }
g) A ∩ B = ∅
h) $2^{|A|} = 2^3 = 8$

**Probability (15p)**

4. The Monty Hall problem is from a game show that is played as follows: You are shown three closed doors by the game show host, and told that behind two doors are goats and behind a third door is a car (winner's prize). Your task is to choose a door and at the end of the game you will win the car, assuming you chose the door with a car behind it. After you have chosen a door, the game show host opens a door you have not chosen, revealing a goat. You are asked if you want to switch door, you do the switch (or not) and then the game finishes. Analyse this problem using probability theory. (7p)
   a. Draw an event tree showing all possibles game progressions.

   From the tree in part a, calculate the following probabilities.
   b. P(win | change door)
   c. P(win | not change door)
   d. P(win)
   e. P(lose)

a. A tree with 3 splits followed by 3 splits and, lastly, 2 splits.
b. P(win | change door) = ⅔
c. P(win | not change door) = ⅓
d. P(win) = ½
e. P(lose) = 1 - P(win) = ½

5. In some population group, 10% have a certain disease. All are given a screening test. Of those who have the disease, 90% will get a positive screening result (i.e. the test say they have the disease). Of those who don't have the disease, 10% will get a positive screening result. When a person gets a positive screening result, what is the probability that they actually have the disease? *Hint: Use Bayes' rule and the theorem of total probability.* (8p)
   a. Derive Bayes' rule from the equation for conditional probability. *i.e. P(A|B) = P(AB)/P(B)*
   b. Define the events A and B.

c. Find numbers for the probabilities needed to use Bayes' rule (i.e. having the disease, getting a positive result, getting a positive result given that one has the disease etc).
d. What is the probability of having the disease, given a positive screening result?

a. $P(A|B) = P(AB)/P(B) \Leftrightarrow P(AB) = P(A|B)P(B) = P(B|A)P(A) \Rightarrow$ **$P(A|B) = P(B|A)P(A)/P(B)$**
b. A (or any arbitrary character) is defined as having the disease
   B (or any arbitrary character) is defined as getting a positive result
c. $P(A)=0.1$
   $P(B|A) = 0.9$
   $P(B|A^c) = 0.1$
   $P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) = 0.9*0.1 + 0.1(1-0.1) = 0.09 + 0.09 = 0.18$ (total probability theorem)
d. $P(A|B) = P(B|A)P(A)/P(B) = 0.9*0.1/0.18 = 50\%$ (Bayes' rule)

# Part 2: Statistics, Linear algebra and Graphs

**Statistics (23p)**

6. The function *f*(x), below, can be interpreted as a probability density function (continuous, $x \in \mathbb{R}$) or a probability mass function (discrete, $x \in \mathbb{Z}$). (6p)

$$f(x) = \begin{cases} 0, & x < -2 \\ C(2+x), & -2 \le x \le -1 \\ C, & -1 < x < 1 \\ C(2-x), & 1 \le x \le 2 \\ 0, & 2 < x \end{cases}$$

   For *both* the discrete and continuous case, find:
   a. The normalization constants C.
   b. $P(X \le 10)$.
   c. $P(-1 \le X \le 1)$.
   d. $P(X \le 0)$.

> a. Continuous: The area under the curve can been seen as two triangles and one rectangle as follows $\frac{1 \cdot C}{2} + 2 \cdot C + \frac{1 \cdot C}{2} = 3C = 1 \to C = \frac{1}{3}$
> Discrete: f(-2)+f(-1)+f(0)+f(1)+f(2) = C+C+C = 3C = 1, gives C = ⅓
> b. All probability mass is below 10, hence $P(X \le 10)=1$ in both cases.
> c. Continuous: The shape of the are is a rectangle, 2*C = ⅔
> Discrete: f(-1)+f(0)+f(1)= C+C+C = 3C = 1
> d. Continuous: f(x) is symmetric around 0 ⇒ $P(X \le 0) = .5$
> Discrete: $P(X \le 0)$ = f(-2) + f(-1) + f(0) = 0 + C + C = ⅔

7. In the INSARK dataset, the lengths of Swedish conscripts are reported. The histogram over heights is, approximately, shaped like a normal distribution. Assuming the distribution of heights is modelled, in centimeters, as $X \sim \mathcal{N}(179, (6.2)^2)$, find: (9p)
   a. E(X).
   b. Var(X).
   c. The three standard deviation span of heights, i.e. $[\mu-3\sigma, \mu+3\sigma]$

   For a random person in the dataset, find the probability of:
   d. Being able to reach the highest kitchen shelf, i.e. being longer than 190 cm.
   e. Being short enough for driving a small vehicle, i.e. shorter than 165 cm.
   f. Being very average, i.e. a height between 175 cm and 185 cm.

> a. $E(X) = \mu = 179$
> b. $Var(X) = \sigma^2 = (6.2)^2 \approx 38.4$
> c. $[\mu-3\sigma, \mu+3\sigma] \to [179-3*6.2, 179+3*6.2] \to [160.4, 197.4]$
> d. $P(190 < X) = 1-\phi((190-179)/6.2) \approx 1-\phi(1.77) \approx 3.8\%$
> e. $P(X < 165) = \phi((165-179)/6.2) \approx \phi(-2.26) = 1-\phi(2.26) \approx 1.2\%$
> f. $P(175 \le X \le 185) = \phi((185-179)/6.2) - \phi((175-179)/6.2) \approx \phi(0.97)-\phi(-0.65) \approx 57\%$

8. If a die is loaded is tricky to find out due to that the effect of the loading hardly shows for any single throw. You suspect that a six sided die is loaded, and have an afternoon free. After throwing the die 100 times, a six has come up 23 times. You would expect a six to come up 100/6≈16.7 times. Is this result *significantly* off from a fair die? To determine this, find: (8p)
   a. The distribution parameters for modelling the trows of the die as bernoulli trials, i.e. for a binomial distributions.

b. Not having the internet at hand, find the normal approximation of this binomial distribution, in order to simplify later calculations. *Note that 100 throws should be considered a small number of samples.*
c. To get an idea of what to expect for random variation of outcomes, find a 95% interval, i.e. [μ-1.96σ, μ+1.96σ].
d. What is the one sided p-value for refuting $H_0$ (fair die) in favour of $H_a$ (loaded die giving more sixes)?
e. Are the results from part d significant at a 90%, 95% or 99% confidence level?

> a. B ~ Binom(n=100, p=1/6)
> b. Normal approximation of the binomial distribution:
> $\mu = np \approx 16.7$, $\sigma = \sqrt{np(1-p)} \approx 3.73$
> $X \sim \mathcal{N}(16.7, (3.73)^2)$
> c. [μ-1.96σ, μ+1.96σ] → [9.4, 24]
> d. Since n is so small, we need to use continuity correction.
> $P(B \leq x|H_0) \approx 1 - \Phi\left(\frac{x-\frac{1}{2}-\mu}{\sigma}\right)$
> $= 1 - \Phi\left(\frac{22.5-16.7}{3.73}\right) \approx 1 - \Phi(1.55) \approx 0.0606$
> e. 90%: Yes
> 95%: No
> 99%: No

**Linear algebra (10p)**

9. Given the following vectors ($\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{u}_1$, $\mathbf{u}_2$), give the resulting vector from expressions a-f below. (6p)

$\mathbf{v}_1 = (-1, -5, -2)^T$ $\mathbf{u}_1 = (2, 6, 3)^T$
$\mathbf{v}_2 = (0, 3, -2)^T$ $\mathbf{u}_2 = (1, -2, 3)^T$

a. $-\mathbf{v}_1$
b. $\mathbf{v}_1 + \mathbf{v}_2$
c. $4(3\mathbf{v}_1 - 2\mathbf{v}_2)$
d. $||\mathbf{u}_1||$
e. $\mathbf{u}_1/||\mathbf{u}_1||$
f. $\mathbf{u}_1 \cdot \mathbf{u}_2$

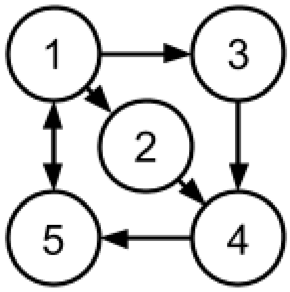
> a. $-\mathbf{v}_1 = (1, 5, 2)^T$
> b. $\mathbf{v}_1 + \mathbf{v}_2 = (-1, -2, -4)^T$
> c. $4(3\mathbf{v}_1 - 2\mathbf{v}_2) = (-12, -84, -8)^T$
> d. $||\mathbf{u}_1|| = 7$
> e. $\mathbf{u}_1/||\mathbf{u}_1|| = (2/7, 6/7, 3/7)^T$
> f. $\mathbf{u}_1 \cdot \mathbf{u}_2 = -1$

10. Given the lines (on parametric form) $\mathbf{p}=\mathbf{u}_1+t\mathbf{v}_1$ and $\mathbf{q}=\mathbf{u}_2+s\mathbf{v}_2$, where $t,s\in\mathbb{R}$ and with the vectors ($\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{u}_1$, $\mathbf{u}_2$) from above, find: (4p)
    a. The number pair (t, s) for where the lines intersect.
    b. The point where the lines intersect.

> a. $\mathbf{p} = \mathbf{u}_1 + t\mathbf{v}_1 = (2-t, 6-5t, 3-2t)^T$
> $\mathbf{q} = \mathbf{u}_2 + s\mathbf{v}_2 = (1, -2+3s, 3-2s)^T \Rightarrow$ Setting $p_i=q_i$ gives: 2-t=1, 6-5t=-2+3s, 3-2t=3-2s
> $p_1=q_1$: 2-t = 1 ⇒ t = 1
> $p_3=q_3$: 3-2t = 3-2s and t = 1 ⇒ s = 1
> b. Setting t=1 and s=1 in the expressions for **p** and **q** gives the point $(1, 1, 1)^T$

**Graph theory (6p)**

11. For the directed graph to the right: (6p)
    a. Make an adjacency matrix.
    b. Is this graph acyclic (i.e. are there no possible cycles)?
    c. Find all paths from vertex 1 to vertex 5.




1. A = (0 1 1 0 1)
   (0 0 0 1 0)
   (0 0 0 1 0)
   (0 0 0 0 1)
   (1 0 0 0 0)
2. No, there are possible cycles, e.g.1 3 4 5 1.
3. 1 (1 to 2) 2 (2 to 4) 4 (4 to 5) 5
   1 (1 to 3) 3 (3 to 4) 4 (4 to 5) 5
   1 (1 to 5) 5

# Table of equations for part 2

## Statistics

**Mass/density/distribution functions**

$$\sum_K f(k) = 1$$

$$\int f(x)dx = 1$$

PDF: f(x) = P(X = x)
CDF: F(x) = P(X ≤ x)

### Binomial distribution

$$B \sim Binom(n, p)$$

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

$$P(B \leq x) = \sum_{k=1}^{x} \binom{n}{k} p^k (1-p)^{n-k}$$

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

**Normal approximation**

$$\mu = np$$

$$\sigma^2 = np(1-p)$$

$$P(a \leq B \leq b) = \Phi\left(\frac{b + \frac{1}{2} - \mu}{\sigma}\right) - \Phi\left(\frac{a - \frac{1}{2} - \mu}{\sigma}\right)$$

### Normal distribution

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

$$f_{PDF}(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

**Linear combinations**

$aX_1 + bX_2 + c$, where: $a, b \in \mathbb{R} \ \wedge \ a, b \neq 0$

$$\mu_{new} = a\mu_1 + b\mu_2 + c$$

$$\sigma^2_{new} = (a\sigma_1)^2 + (b\sigma_2)^2$$

sample mean: $X^* \sim \mathcal{N}(\bar{X}, \sigma^2/\sqrt{n})$

**Maximum likelihood estimators (MLE)**

$$\mu = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

**Standard normal**

$$X \sim \mathcal{N}(0, 1^2)$$

$$Z = \frac{X - \mu}{\sigma}$$

$$P(Z \leq x) = \Phi(x)$$

$$\Phi(-x) = 1 - \Phi(x)$$

## Linear Algebra

$$\bar{p}, \bar{q} \in \mathbb{R}^n$$

$$\bar{p} \cdot \bar{q} = \sum_{i=1}^{n} p_i q_i = \|\bar{p}\| \, \|\bar{q}\| \, cos\theta$$

**Geometry**

Line: $\bar{p} = \bar{p_0} + t\bar{v}$

Plane: $\bar{n}(\bar{p} - \bar{p_0}) = 0$

## Graph theory

$$A_{ij} = \begin{cases} 1, & [v_i \rightarrow v_j] \in E(G) \\ 0, & otherwise \end{cases}$$