UPPSALA UNIVERSITY
Fredrik Wahlberg

Dugga 1: Sets, functions and probability

Pass: at least 60% correct (i.e. 24p). Pass with distinction: at least 80% correct on the two tests added. 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.

General rules: Mobile phones must be switched off.

Tools: Pen and calculator.

Formulas:

Bayes' theorem:	$P(B_i\|A) = \frac{P(A\|B_i)P(B_i)}{P(A\|B_1)P(B_1)+P(A\|B_2)P(B_2)+...+P(A\|B_n)P(B_n)}$
Theorem of total probability:	$P(A\|B) = P(A\|B_1)P(B_1) + P(A\|B_2)P(B_2) + ... + P(A\|B_n)P(B_n)$
Expectation:	$E(X) = \sum_{i=1}^{n} P(X = x_i)x_i$
Variance:	$Var(X) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$, where $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

Set theory (10p)

1. For sets A={x∈ℕ| 1 ≤ x ≤ 3} and B={a, b, c}, show the resulting sets of the following statements. (6p)

   (a) A
   (b) A ∪ B
   (c) A ∩ B
   (d) |A| + |B| - |A ∩ B|
   (e) A × B
   (f) A × ∅

   > (a) {1, 2, 3}
   > (b) {1, 2, 3, a, b, c}
   > (c) ∅
   > (d) 3 + 3 - 0 = 6
   > (e) {(1, a), (1, b), (1, c), (2, a) … (3, c)}
   > (f) ∅

3. For *any finite sets* A and B, argue for each of the following statements that it is true for all A and B, false for all A and B, or true for some A and B. (4p)

   (a) A ⊆ (A ∪ B)
   (b) (A − B) ⊆ A
   (c) |A ∪ B| = |A| + |B| − |A ∩ B|
   (d) |A − B| = |B − A|

   > (a) Necessarily true for all A and B. A is always a subset of a union between itself and any B. A union never removes anything.
   > (b) Necessarily true for all A and B. (A-B) must include A or less by definition. One can not remove elements from A and get something that is larger than A.
   > (c) Necessarily true for all A and B. Can be shown with a Venn diagram or derived. Compare to the general addition rule.
   > (d) Can be true for disjunct A and B where |A|=|B|, including empty sets.

Functions (4p)

3. For the following recursive function f:ℕ→ℕ, enumerate the results all f(n) where n∈{x∈ℕ| x ≤ 10}. (4p)

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

{x∈ℕ| x ≤ 10} = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
f(n) is the Lucas series (related to the Fibonacci series)
f(1) = 2, f(2) = 1, f(3) = 3, f(4) = 4, f(5) = 7, f(6) = 11, f(7) = 18, f(8) = 29, f(9) = 47, f(10) = 76

Probability (26p)

4. In a game of dice, two players take turns in throwing two six-sided dice. The player throwing the dice wins the turn if the difference between the dice is an even number (including 0), otherwise the non-throwing player wins the turn. The game is played until one player wins 10 turns. (8p)
Give probabilities as percentages and answer with relevant calculations/reasoning.
   (a) Define a suitable sample space, Ω, for one turn in this game.
   (b) Define one (or more) suitable events relevant for the problem.
   (c) Show (using a table or event tree) the number of outcomes in Ω that favour the player throwing the dice.
   (d) Calculate the probability of the throwing player or the non-throwing player winning a turn, respectively.
   (e) After 10 turns (i.e. 5 turns throwing the dice for each player), what is the probability that the game has ended.

(a) T = {1, 2, 3, 4, 5, 6}, Ω = TxT = {(1, 1), (1, 2), (1,3), (1, 4) … (6,5), (6,6)}
(b) A = {(a, b)∈Ω | |a-b|∈{0, 2, 4}} (caster wins)
(c) Event A marked in bold

|a - b|	**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**

(d) P(A) = |A|/|Ω| = 18/36 = ½ = 50%
P(A)+P(A^C)=1 (from axioms) ⇒ P(A^C) = 1-P(A) = ½ = 50%
(e) One can think of a tree diagram with two outcomes per level. Since each turn is independent the probability of 10 wins requires five throwing wins and 5 non-throwing wins (due to the players switching roles between turns). The end criterion can be reached by either player winning. $2 \cdot P(A)^5 P(A^C)^5 = 2 \cdot (½)^5 (½)^5 \approx 0.2\%$

5. At an organisation with 7250 employees, the staff is divided into two overlapping work groups. For the year 2018, 4983 people work in production roles and 5239 people with administrative roles. (6p)
Give probabilities as percentages and answer with relevant calculations/reasoning.
   (a) How many people take on the roles of both work groups?

(b) Is the probability of a random employee belonging to one work group statistically independent of the probability of the employee belonging to the other?
(c) What is the probability of a random employee working in an administrative role given that we know they already work in a production role, and vice versa (*i.e.* $P(A|B)$ *and* $P(B|A)$)?

Def. A: in admin, B: in production
(a) (2p): $|\Omega| = |A \cup B| = |A| + |B| - |A \cap B| \Rightarrow 7250 = 5239 + 4983 - |A \cap B| \Rightarrow$ $\mathbf{|A \cap B| = 2972}$
(b) (2p): If statistically independent: $P(A \cap B) = P(A)P(B)$
$P(A \cap B) = P(A)P(B) \Rightarrow |A \cap B|/|\Omega| = (|A|/|\Omega|) * (|B|/|\Omega|)$
Left Hand Side: $|A \cap B|/|\Omega| \approx 41\%$
Right hand side:: $(|A|/|\Omega|) * (|B|/|\Omega|) \approx 50\%$
Not independent
(c) (2p): $P(A|B) = P(A \cap B)/P(B) = (|A \cap B|/|\Omega|) * (|\Omega|/|B|) = 2972/4983 \approx$ **60%**
$P(B|A) = P(A \cap B)/P(A) = (|A \cap B|/|\Omega|) * (|\Omega|/|A|) = 2972/5239 \approx$ **57%**

6. In the South African region KwaZulu-Natal, the prevalence of HIV is among the highest in the world at 30%. The ELISA HIV test is 99.7% accurate for those who carry HIV and 92.6% accurate for those who don't. Assuming a test is positive for a random person from KwaZulu-Natal, what is the probability that that person carries HIV? (12p)
Give probabilities as percentages and answer with relevant calculations/reasoning
(a) Define relevant variables, a sample space and events.
(b) Find the probabilities given by the text in terms of your chosen probability space from (a).
(c) Draw an event tree
(d) What is the probability of having the infection, given a positive ELISA result *(i.e. P(being sick | positive test)*?

Redo the calculations from (d) for:
(e) Sweden, with an HIV prevalence of 0.2%
(f) Kenya, with an HIV prevalence of 5%

Multiple tests are often performed to increase accuracy (assumed to be independent). This can be especially important for regions with a low prevalence due to the many false positives. For Sweden, calculate:
(g) The probability of having the infection, given two consecutive positive ELISA results?
(h) The probability of having the infection, given three consecutive positive ELISA results?

(a) (1p): A: Having HIV, B: Positive test
(b) (1p): $P(A) = 30\%$
$P(B|A) = 99.7\%$
$P(B^C|A^C) = 92.6\%$
(c) (2p): Draw a tree with 2x2 splits
Since $P(A)+P(A^C)=1$: $P(A^C) = 70\%$
Since $P(B|A)+P(B^C|A)=1$: $P(B^C|A) = 0.3\%$
Since $P(B|A^C)+P(B^C|A^C)=1$: $P(B|A^C) = 7.4\%$
(d) (2p): By Bayes' theorem (or from the tree):
$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^C)P(A^C)} = \frac{0.997 \cdot 0.3}{0.997 \cdot 0.3 + 0.074 \cdot 0.7} = 0.8524 \approx 85\%$$
(e) (1p): From (d): $P(A|B) = \frac{0.997 \cdot 0.002}{0.997 \cdot 0.002 + 0.074 \cdot 0.998} = 0.0263 \approx 2.6\%$
(f) (1p): From (d): $P(A|B) = \frac{0.997 \cdot 0.05}{0.997 \cdot 0.05 + 0.074 \cdot 0.95} = 0.4149 \approx 41\%$
(g) (2p): A tree illustrating this would first have a branching for A and A^C followed by two consecutive splits for the two tests (i.e. $|\Omega|=8$). The probability for a first positive test given A is $P(B|A)$ as before. Since the tests are independent, the probability $P(B|A)$ can also be used for the second test. The same argument can be used for $P(B|A^C)$. This leads to

a Bayes' theorem with the following probabilities (this can also be seen in a tree diagram):

$P(A|2xB) = \frac{P(B|A)^2 P(A)}{P(B|A)^2 P(A) + P(B|A^C)^2 P(A^C)}$

$= \frac{0.997^2 \cdot 0.002}{0.997^2 \cdot 0.002 + 0.074^2 \cdot 0.998} = 0.2667 \approx 27\%$

(h) (2p): Same reasoning as in (g):

$|\Omega|$=16

$P(A|3xB) = \frac{0.997^3 \cdot 0.002}{0.997^3 \cdot 0.002 + 0.074^3 \cdot 0.998} = 0.8305 \approx 83\%$