This list contains exercises of the type you will find in an exam for the course *Natuurlijke Taalmodellen en Interfaces*.

Contents

Points

Question:	1	2	3	4	Total
Points:	1	1	3	1	6

1 Random variables and rules of probabilities

1. (1 point) Let X be a random variable whose sample space is the English vocabulary Σ and whose mapping to $\mathbb{R}$ is realised by an arbitrary *enumeration* of Σ. Given the partial definitions of X below, mark those that are definitely **invalid**?

◯ $X(\omega) = \begin{cases} 1 & \text{if } \omega = \{\text{the}\} \\ 2 & \text{if } \omega = \{\text{the}\} \\ 3 & \text{if } \omega = \{\text{cat}\} \\ 4 & \text{if } \omega = \{\text{dog}\} \\ \dots \end{cases}$

◯ $X(\omega) = \begin{cases} 1 & \text{if } \omega = \{\text{the}\} \\ 1 & \text{if } \omega = \{\text{a}\} \\ 2 & \text{if } \omega = \{\text{cat}\} \\ 3 & \text{if } \omega = \{\text{dog}\} \\ \dots \end{cases}$

◯ $X(\omega) = \begin{cases} 1 & \text{if } \omega = \{\text{the, a}\} \\ 2 & \text{if } \omega = \{\text{cat}\} \\ 3 & \text{if } \omega = \{\text{dog}\} \\ \dots \end{cases}$

◯ $X(\omega) = \begin{cases} 1 & \text{if } \omega = \{\text{the}\} \\ 2 & \text{if } \omega = \{\text{a}\} \\ 3 & \text{if } \omega = \{\text{cat}\} \\ 4 & \text{if } \omega = \{\text{dog}\} \\ \dots \end{cases}$

---

2. (1 point) Number the identities on the right according to the concepts on the left.

1. Chain rule	(_____)	$P_{A\|B}(a\|b) = \frac{P_{AB}(a,b)}{P_B(b)}$
2. Conditional probability	(_____)	$P_A(a) = \sum_{b \in \mathcal{B}} P_{AB}(a,b)$
3. Bayes rule	(_____)	$P_{AB}(a,b) = P_B(b) P_{A\|B}(a\|b)$
4. Marginal probability	(_____)	$P_{B\|A}(b\|a) = \frac{P_B(b) P_{A\|B}(a\|b)}{P_A(a)}$

2 Categorical distributions

3. Let X be a Categorical random variable:

$$X \sim \text{Cat}(\theta_1, \ldots, \theta_v)$$

(a) (½ point) What is the support $\mathcal{X}$ of the random variable?

(b) (½ point) What is the value of $P_X(x)$?

(c) (1 point) What conditions apply to valid parameters $\langle \theta_1, \ldots, \theta_v \rangle$?

(d) (1 point) Given a data set of n i.i.d. observations, what is the maximum likelihood estimate of θ_x?

Total for Question 3: 3

4. (1 point) Select, out of the list below, vector(s) that constitute(s) **valid** categorical parameters for a categorical random variable that may take on one out of 7 classes.

- ◯ $\langle 0.1, 0.1, 0.1, 0.1, 0.1, 0.2, 0.3 \rangle$
- ◯ $\langle 0.1, 0.1, 0.1, 0.1, 0.1, 0.5 \rangle$
- ◯ $\langle 0.2, 0.2, 0.1, 0.1, 0.1, 0.2, 0.2 \rangle$

Assessment

Question	Points	Score
1	1	
2	1	
3	3	
4	1	
Total:	6	