

Assignment 1: Probability Theory Warm-Up

(1 P.)

- (a) Calculate the answers to the following questions. Hint: You should restrict yourself to the specific probability distributions mentioned in the lecture slides and investigate their applicability in terms of sufficient parameters given in the statements.
- (i) A dice is rolled 10 times. What is the probability that the face “4” turns up for 3 times?
 - (ii) How many times do we need to roll a dice until a “4” or a “2” turns up?
 - (iii) Suppose we have a document collection where the term “computer” occurs 5 times per document on average. What is the probability that the term “computer” occurs exactly 8 times in a document?
- (b) Consider the following joint distribution of Boolean random variables X_1 , X_2 , and X_3 :

X_1	X_2	X_3	Pr
0	0	0	0.06
0	0	1	0.08
0	1	0	0.18
0	1	1	0.20
1	0	0	0.06
1	0	1	0.10
1	1	0	0.12
1	1	1	0.20

(i) What are the marginal distributions of X_1 and X_2 ?

(ii) What is the conditional distribution $\Pr(X_1 | X_2)$?

(iii) What is the variance and the expectation of X_2 ?

Assignment 2: BIM

(1 P.)

Consider the query $\{q := \text{Michael Jordan computer science}\}$ with the four terms $t_1 = \text{Michael}$, $t_2 = \text{Jordan}$, $t_3 = \text{computer}$, $t_4 = \text{science}$. An initial query evaluation returns the documents d_1, \dots, d_8 that are intellectually evaluated by a human user for relevance. The occurrences of the terms t_1, \dots, t_4 in the documents as well as the relevance feedback of the user are depicted in the following table, where “1” points out a relevant document and “0” points out a non-relevant document.

	t_1	t_2	t_3	t_4	Relevant
d_1	1	0	1	0	0
d_2	1	1	0	0	0
d_3	1	0	0	0	0
d_4	0	1	1	1	1
d_5	1	1	1	1	1
d_6	0	1	0	1	1
d_7	0	1	1	0	0
d_8	1	0	1	1	1

Compute the similarities of the documents d_9 and d_{10} , given below, to the given query using the probabilistic retrieval model with relevance feedback according to the formula by Robertson & Spärck-Jones with Lidstone smoothing ($\lambda = 0.5$).

	t_1	t_2	t_3	t_4
d_9	1	0	1	0
d_{10}	0	1	0	1

Assignment 3: Bayesian Network Inference

(1 P.)

- (a) Consider the Bayesian network with random variables **Cloudy**, **Sprinkler**, **Rain** and **Wet** as given in the lecture. Compute the probability that the ground is wet given the information that it is cloudy.
- (b) Consider the following binary random variables:
- S : a student solves the assignments regularly and successfully
 - M : a student had a good mathematical training from earlier course
 - T : a student passes the written test
 - B : a student achieves bonus points
 - F : a student is female

Now consider that following dependencies among these variables exist:

- A student solves the assignments regularly and successfully when the student had a good mathematical training from earlier course or the student is female.
 - A student passes the written test when the student solves the assignments regularly and successfully.
 - A student achieves bonus points when the student passes the written test or the student solves the assignments regularly and successfully.
- (i) Model the bayesian network based on the mentioned dependencies.
- (ii) Discuss which parameters (conditional probabilities) need to be estimated.
- (iii) Show how to derive the joint probability distribution of S , M , and B .