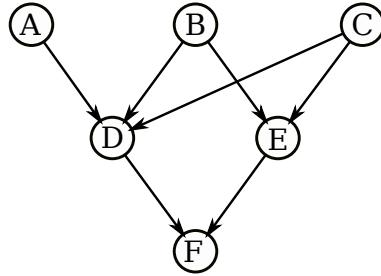


# Internet Analytics (COM-308)

## Problem Set 5

### Problem 1: Bayesian network: parameters

Assume you are given 6 random variables:  $A, B, C, D, E$  and  $F$ , which take values  $A, B, C \in [0, \dots, 9]$ ,  $D \in [0, \dots, 5]$  and  $E, F \in [0, 1]$ .



Bayesian network of 6 variables.

- (a) How many parameters does the model have if you assume full independence of these variables?
- (b) How many parameters does the model have if you assume a general distribution (no constraints)?
- (c) How many parameters does the model have if you assume the distribution is defined by the Bayesian network in Fig. 1?
- (d) Assume you want to increase a sparsity of the system as much as possible (reduce the number of parameters). Which edge is the best candidate to remove.

### Problem 2: Dirichlet distribution

Let  $X \sim \text{Dir}(\alpha_1, \alpha_2, \dots, \alpha_K)$ . Show that  $E[X_i] = \alpha_i / \sum_{1 \leq k \leq K} \alpha_k$ .

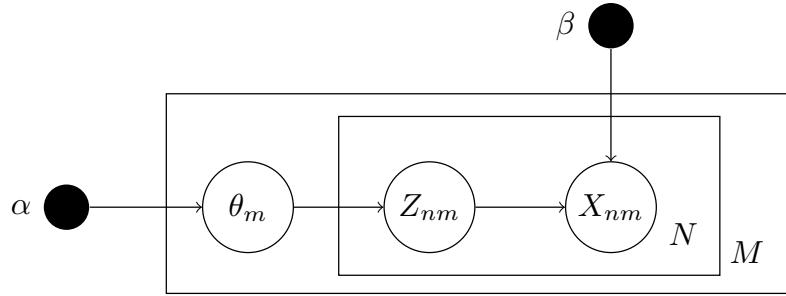
One way to do this is by writing the expectation as an integral that marginalizes out  $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_K$ , and then recognizing that the resulting integrand is again a Dirichlet density (with different parameters).

Hint: for the  $\Gamma$  function, the following property will be useful:  $\Gamma(x+1) = x\Gamma(x)$ .

### Problem 3: LDA-generated text

In this exercise we look at the generative model of LDA. The plate notation of LDA is shown below.

Assume we have three topics: “sports”, “politics” and “shopping”. The prior on word distribution per topic,  $\beta$ , is given in the following table



Word \ Topic	Sports	Politics	Shopping
Word			
soccer	$\frac{1}{4}$	0	0
champion	$\frac{1}{4}$	0	0
contest	$\frac{1}{4}$	0	0
score	$\frac{1}{4}$	0	0
party	0	$\frac{1}{4}$	0
election	0	$\frac{1}{4}$	0
policy	0	$\frac{1}{4}$	0
democracy	0	$\frac{1}{4}$	0
customer	0	0	$\frac{1}{4}$
cashier	0	0	$\frac{1}{4}$
market	0	0	$\frac{1}{4}$
shop	0	0	$\frac{1}{4}$

The three documents (given below) are generated with three different values of  $\alpha = (0.01, 0.01, 0.01)$ ,  $(1, 1, 1)$ , or  $(100, 100, 100)$ . Identify the corresponding value of  $\alpha$  for each of the three documents.

**Hint:** Look at the frequency of topics in each document.

### Document 1

election score democracy score election shop contest champion score market democracy policy market election election democracy champion election contest cashier cashier policy party election market champion score soccer policy policy champion customer customer champion democracy party soccer score market cashier party market cashier market shop shop contest champion cashier customer cashier score soccer soccer market democracy election democracy election party champion market soccer contest democracy election contest customer party customer party soccer democracy score policy party election cashier cashier policy policy contest cashier election contest customer score election democracy shop party party champion market shop cashier market election champion

### Document 2

policy party party election party party policy policy party party party party policy democracy party policy election election democracy party party democracy election policy democracy party election party democracy democracy party election election party party democracy democracy policy democracy party policy policy party election policy party party party election policy party election democracy democracy election election party party policy party policy election party democracy party election democracy election party democracy democracy policy party democracy democracy

election democracy policy election election party party party election policy policy  
policy party party policy

### Document 3

election policy contest score score party democracy democracy champion score shop policy  
champion champion champion soccer contest election policy score cashier policy party policy  
score party champion score contest election score party contest party champion soccer party  
shop democracy contest score contest election election contest policy election party policy soccer  
policy champion contest cashier contest soccer cashier champion champion champion soccer  
score score soccer policy contest soccer party democracy policy election policy democracy score  
score party champion score champion score champion score score soccer soccer democracy  
contest party champion market score contest contest contest champion contest contest contest  
election contest