Internet Analytics (COM-308)

Homework Set 7

Exercise 1

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 $E, F \in [0, 1]$.

(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.

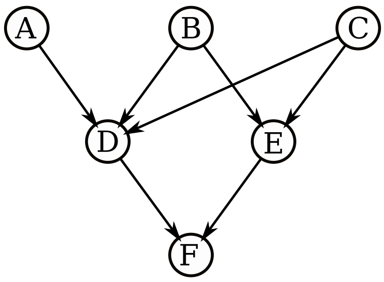


Figure 1: Bayesian network of 6 variables.

(a) If all the variables are independent the number of parameters is $9 + 9 + 9 + 5 + 1 + 1 = 34$.

(b) In case of general distribution we need to consider all the possible combinations of these variables thus the number of parameters is $10^3 * 6 * 2 * 2 - 1 = 23999$.

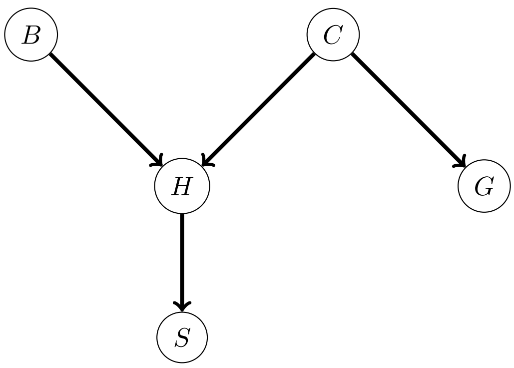
(c) For the given BN the number of parameters is $3 * 9 + 10^3 * 5 + 10^2 * 1 + 6 * 2 * 1 = 5139$.

(d) We can remove one out of three edges coming into D; in such a way we reduce the number of parameters to $3 * 9 + 10^2 * 5 + 10^2 * 1 + 6 * 2 * 1 = 639$.

Exercise 2

We have seen Bayesian networks (BN) in the class. In this exercise we study the inference problem over a BN.

(a) Given the network below calculate the conditional and marginal probabilities $P(G|C, H)$, $P(S|C, G)$, $P(S|C, G, H)$, $P(H)$ and $P(S)$.



$P(B)$

$B=0$	$B=1$
0.6	0.4

$P(C)$

$C=0$	$C=1$
0.8	0.2

$P(G|C)$

C	$G=0$	$G=1$
0	0.6	0.4
1	0.3	0.7

$P(H|B,C)$

(B,C)	$H=0$	$H=1$
$(0,0)$	0.8	0.2
$(0,1)$	0.5	0.5
$(1,0)$	0.4	0.6
$(1,1)$	0.1	0.9

$P(S|H)$

H	$S=0$	$S=1$
0	0.5	0.5
1	0.1	0.9

- G given C is independent of H. Therefore, $P(G|C,H)=P(G|C)$ which is given above.
- S given C is independent of G. We have:

$$P(S|C)=\sum_{i,j}P(S|C,B=i,H=j)P(B=i,H=j|C).$$

We know that S given H is independent of B and C. Therefore, we have

$$P(S|C)=\sum_{i,j}P(S|H=j)P(B=i,H=j|C).$$

We have

$$P(B=i,H=j|C)=P(H=j|B=i,C)P(B=i|C)=P(H=j|B=i,C)P(B=i).$$

Wrap up above,

$P(B,H|C)$

(B,H)	$C=0$	$C=1$
$(0,0)$	0.48	0.3
$(0,1)$	0.12	0.3
$(1,0)$	0.16	0.04
$(1,1)$	0.24	0.36

$P(S|C)$

S	$C=0$	$C=1$
0	0.356	0.236
1	0.644	0.764

- S given H is independent of C and G. Therefore, $P(S|C,G,H)=P(S|H)$ which is given above.
- $P(H)=\sum_{i,j}P(H,B=i,C=j)=\sum_{i,j}P(H|B=i,C=j)P(B=i,C=j)$. Therefore, we have

$P(H)$

$H=0$	$H=1$
0.58	0.42

- $P(S)=\sum_i P(S,H=i)=\sum_i P(S|H=i)P(H=i)$. Therefore, we have

$P(S)$

$S=0$	$S=1$
0.332	0.668

(b) Compare the following probabilities, and determine what type of inference they represent (prediction, explanation, intercausal):

- $P(G = 1|H = 1)$ *and* $P(G = 1|H = 1, B = 1)$
- $P(S = 1|H = 1)$ *and* $P(S = 1|H = 1, B = 1)$
- $P(C = 1)$ *and* $P(C = 1|S = 1)$

We have:

- $P(G = 1|H = 1) > P(G = 1|H = 1, B = 1)$: intercausal reasoning (explaining away).
- $P(S = 1|H = 1) = P(S = 1|H = 1, B = 1)$: prediction (they are equal because S given H is independent of B)
- $P(C = 1) < P(C = 1|S = 1)$: explanation.

Exercise 3

(a) The three plots in Figure 2 show the densities of three Dirichlet distributions over the 2-simplex with parameters $\alpha_1 = (0.3, 0.3, 0.3)$, $\alpha_2 = (1.1, 1.1, 1.1)$ and $\alpha_3 = (10, 10, 10)$; determine which plot corresponds to which α.

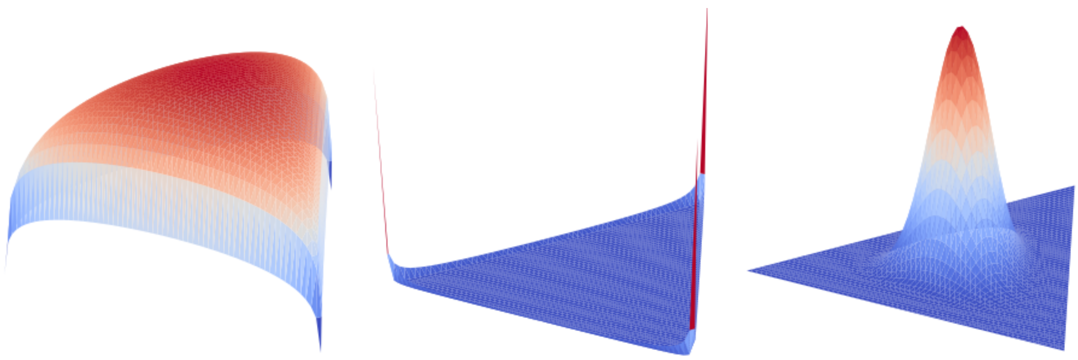
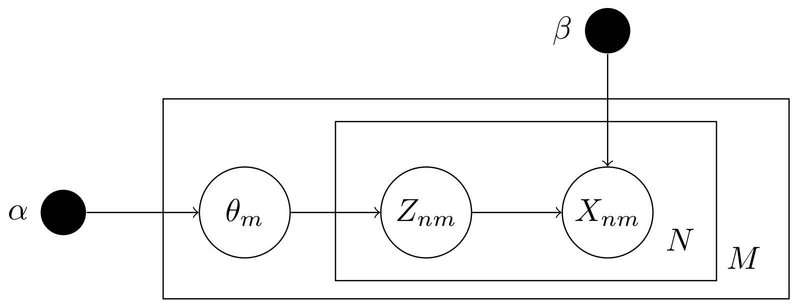


Figure 2: Three Dirichlet densities on the 2-simplex.

By definition the PDF of Dirichlet distribution is $f(\theta_1, \theta_2, \theta_3) \propto \theta_1^{\alpha_1 - 1}\theta_2^{\alpha_2 - 1}\theta_3^{\alpha_3 - 1}$.

The first plot corresponds to $\alpha_2 = (1.1, 1.1, 1.1)$, it is close to uniform distribution of theta's over the 2-simplex. The second plot corresponds to $\alpha_1 = (0.3, 0.3, 0.3)$, it has peaks at the corners of the simplex. The third plot corresponds to $\alpha_3 = (10, 10, 10)$, giving a peak inside the simplex.

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



Assume we have three topics ($K = 3$) sports, politics, and shopping. The prior on word distribution per topic, β, is given in the following table:

WordTopic	*Sport*	*Politics*	*Shopping*
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 below are generated with three different values of $\alpha = 0.01, 1$ *and* 100*. Identify the corresponding value of* α *for each document.*

Hint: Guess 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 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 party policy democracy party policy election election democracy party party democracy election policy democracy party election party democracy democracy party election election party policy party election election democracy election party 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 election 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

In class, we saw that α controls uniformity/sparsity of topic vectors:

- small $\alpha < 1$: prior on topic distribution is concentrated at the edge/corners of the simplex;
- large $\alpha > 1$: prior on topic distribution has most probability mass inside the simplex.

From word per topic distributions we can see that each word appears in only one topic. Therefore, we can count the number of words form each topic in the documents.

- Document 1: topic frequency is $(0.30, 0.38, 0.32)$ and probably this document is the one with the largest value of $\alpha = 100$.
- Document 2: topic frequency is $(0, 1.0, 0)$. This document has only one topic. We can conclude this is probably the document with the smallest $\alpha = 0.01$.
- Document 3: topic frequency is $(0.59, 0.35, 0.06)$ and $\alpha = 1$.