

# Exercices

Serafin Moral

July 28, 2004

1. Assume that we throw a thumbtack ( $X_1$ ) with results: heads (1) and tails (2) and that depending of the result, we pick up a ball from one of two urns with balls of different colors.

Let  $X_2$  the color of the ball, which can be Red (R), Blue (B) and Green (G).

Assume the network:



- (a) Determine appropriate 'a priori' Dirichlet distributions.  
 (b) Obtain Bayesian estimations under the following sequence of observations:

$X_1$	1	2	1	1	2	1	1	1	2	2
$X_2$	R	G	R	B	B	R	G	R	G	G

- (c) Obtain a global imprecise estimation.  
 (d) Obtain local imprecise estimations under appropriate sample sizes.
2. Assume that we have three urns with 10 balls each.
- The first one,  $U_1$ , has 4 red, 4 blue, and 2 of unknown colour, the second,  $U_2$ , has 3 red, 5 blue, and 2 unknown, and the third,  $U_3$ , 6 red, 2 blue, and 2 of unknown colour.
  - We also have that the balls with unknown colour are blue or red and that they have the same composition of colours in the three urns: either are both red, or blue, or one red and the other blue.
  - We consider the following experiment: a ball is chosen at random from the first urn,  $U_1$ , (its colour is variable  $Z$ ). Then an urn ( $U_2$  or  $U_3$ ) is chosen and two balls are drawn at random and with replacement from it (variables  $X$  and  $Y$  represent the colours of these two balls). If  $Z$  is red then both balls are from  $U_2$  and if  $Z$  is blue then the balls are from  $U_3$ .
- (a) Compute the marginal credal set for variables  $X, Z$ .  
 (b) Compute the conditional set of  $Y$  for the different values of  $Z$ .

- (c) Show that there is not strong conditional independence.
- (d) Compute the least informative credal set for which there is strong conditional independence and the computed marginal and conditional credal sets are the same than the above computed.
- (e) Give an interpretation of this credal set in terms of urns and balls.
3. Use Elvira to design a Bayesian network with 10 or more independent binary variables with uniform distribution. Obtain a sample of size 4. Learn networks from the sample with different procedures. Describe the obtained networks (number of arcs, most extreme probabilities).
4. Build a classification tree (using ID3 and the imprecise probability method) for the problem of classifying variable *Cancer* from the following database:

<i>Age</i>	<i>Genetic</i>	<i>Cough</i>	<i>X – Ray</i>	<i>Cancer</i>
Young	Yes	No	No	Yes
Young	Yes	Yes	Yes	Yes
Young	No	Yes	No	No
Interm.	Yes	No	No	No
Interm.	No	Yes	Yes	Yes
Interm.	No	No	Yes	No
Interm.	No	Yes	No	No
Old	No	Yes	Yes	Yes
Old	No	No	Yes	No
Old	No	Yes	Yes	Yes
Old	No	Yes	No	Yes
Old	Yes	Yes	Yes	Yes

5. Given the following frequencies for *X* and *Y*, determine whether we should use *X* to classify *Y* by using your intuition, Bayesian score, upper entropy score, and imprecise score in the following situations:

(a) 

	<i>y</i> <sub>1</sub>	<i>y</i> <sub>2</sub>	<i>y</i> <sub>3</sub>
<i>x</i> <sub>1</sub>	1	2	0
<i>x</i> <sub>2</sub>	0	1	2

(b) 

	<i>y</i> <sub>1</sub>	<i>y</i> <sub>2</sub>	<i>y</i> <sub>3</sub>
<i>x</i> <sub>1</sub>	2	2	1
<i>x</i> <sub>2</sub>	1	2	2

(c) 

	<i>y</i> <sub>1</sub>	<i>y</i> <sub>2</sub>
<i>x</i> <sub>1</sub>	3	5
<i>x</i> <sub>2</sub>	0	1

(d) 

	<i>y</i> <sub>1</sub>	<i>y</i> <sub>2</sub>
<i>x</i> <sub>1</sub>	3	5
<i>x</i> <sub>2</sub>	4	3
<i>x</i> <sub>3</sub>	1	0

		$y_1$	$y_2$
(e)	$x_1$	2	0
	$x_2$	3	1
	$x_3$	0	1

6. It is possible to propagate imprecise probability intervals in Elvira with the following command:

[-out outputfile]

Test the procedure with the following networks:

(with and without file),

(only for variable ).

Compare the intervals obtained with the two versions of network. Compare the times in relation with the network size. Compare your intervals with the ones obtained by other students.

7. Consider two binary variables  $X, Y$  with a global credal set, which determines a

system of probability intervals given by:

		$y_1$	$y_2$
$x_1$	$[\underline{P}(x_1, y_1), \overline{P}(x_1, y_1)]$	$[\underline{P}(x_1, y_2), \overline{P}(x_1, y_2)]$	$[\underline{P}(x_1, y_2), \overline{P}(x_1, y_2)]$
$x_2$	$[\underline{P}(x_2, y_1), \overline{P}(x_2, y_1)]$	$[\underline{P}(x_2, y_2), \overline{P}(x_2, y_2)]$	$[\underline{P}(x_2, y_2), \overline{P}(x_2, y_2)]$

Determine the conditions under which we can have epistemic independence or strong independence.

8. Compute the Bayesian and imprecise scores for the two models (dependence and independence) with the following data:

$Case$	1	2	3	4	5	6	7	8
$X_1$	1	1	1	1	2	2	2	2
$X_2$	1	1	1	1	2	2	2	2

And with the following data:

$Case$	1	2	3	4	5	6	7	8
$X_1$	1	1	1	1	2	2	2	2
$X_2$	1	1	2	2	1	1	2	2

To compute the Gamma function you can use the following values of the logarithm of this function and take into account that  $\Gamma(x) = (x - 1) * \Gamma(x - 1)$ .

	0.05	0.1	0.2	0.25	0.3
Loggamma	2.96887920105176	2.25271265173425	1.52406382243085	1.28802252469814	1.09579799481814
	1/3	0.4	0.5	0.6	2/3
Loggamma	0.98542064692783	0.796677817701839	0.572364942924743	0.398233858069267	0.303150275147547
	0.7	0.8	0.9	0.95	1.0
Loggamma	0.260867246531687	0.152059678399849	0.0663762397347472	0.0309687952379747	0