CSE 573 Spring 2014

Assignment 3: Uncertainty

Turn in the file named “answers.pdf” via Catalyst Dropbox by 11pm, May 21.

Problem 1. (20 Points) According to extremely reliable sources (Wikipedia), 78% of email is spam. According to experiments conducted on an inbox, 11% of spam email messages contain the word “Pills”. In comparison, only 1% of non-spam email messages contain this word.

(a) (5 Points) What is the probability that a message contains the word “Pills” and is spam?
(b) (5 Points) What is the probability that a message does not contain the word “Pills”?
(c) (5 Points) What is the probability that a message does not contain the word “Pills” or is spam?
(d) (5 Points) Given that a message contains the word “Pills”, what is the probability that it is spam?

Problem 2. (30 Points) Suppose you are a witness to a night-time hit-and-run accident involving a taxi in Athens. All taxis in Athens are blue or green, and green taxis are nine times more common that blue taxis. You immediately point your camera at the taxi and snap a burst of shots. However, the color filter in your camera is unstable; it sometimes incorrectly displays green for a blue object, or blue for a green object.

Suppose:
There is a 30% chance of your camera filter not working for any shot.
A green taxi appears blue in 80% of shots with non-working filter.
A blue taxi appears green in 20% of shots with non-working filter.
If the camera filter is working, you correctly observe blue taxi or green taxi.

(a) (15 points) In the photo, you *observe* a blue taxi. However, your camera filter is unstable sometimes, so you are not sure if it is actually a blue taxi or a green taxi. What’s is the probability that your photo is of a blue taxi? Draw a Bayesian network.

(b) (15 points) You find other photos of the same taxi from the burst. All four of your photos show a blue taxi. What is the probability that the taxi in these pictures was actually a blue taxi? Draw a Bayesian network.

Background of Genetics of Mendel's Peas

In his famous experiments, Mendel developed the foundations of hereditary genetics. The experiments involved breeding red and white flowered pea plants. There is a gene for the color of the flower, and there are two possible kinds of alleles (genetic values): R for red ones and r for white ones. Each plant has two alleles: one inherited from each parent. If both its alleles are r for white, then its flowers will be white, but if either or both its alleles are R for red, then its flowers will be red, since R for red is the dominant allele. The red allele is represented as R, and the white one as r. So a plant that is Rr has one red and one white allele, and therefore red flowers. If a plant which is Rr is mated with a plant which is rr, then the offspring will be Rr or rr, since it always gets an r from one parent, and may get an R or r from the other. If the parents are RR and rr, then all offspring will be Rr.

Node P1, P2 and Child represent their genetics respectively. Node Color_P1 represents the color of parent 1. The relation table of Color_P1 is derived strictly from the dominance information, and the relation tables of Color_P2 and Color_Child are the same. The relation table of P1 assumes that R and r occur in equal proportion, and the relation table of P2 is the same. The relation table of Child indicates that during mating there is an equal chance for the inheritance of either allele.

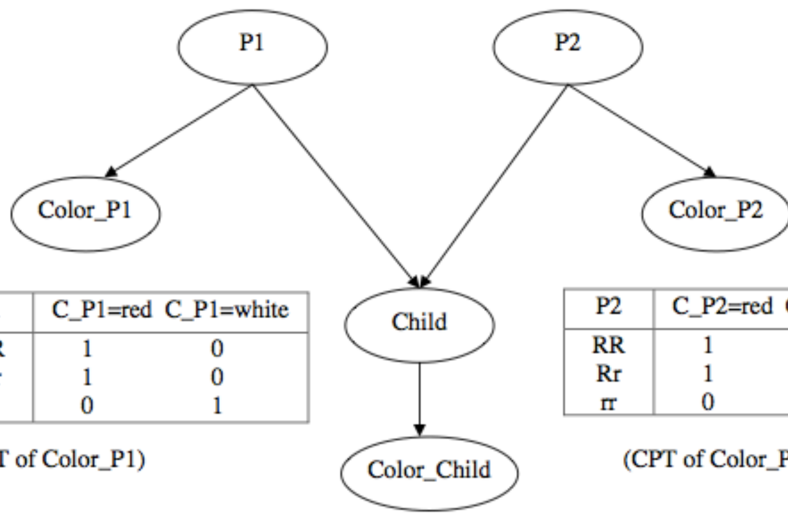
Figure 1. The original Bayesian network of "Genetics of Mendel's Peas"

P1=RR	P1=Rr	P1=rr
1/4	1/2	1/4

(CPT of P1)

P2=RR	P2=Rr	P2=rr
1/4	1/2	1/4

(CPT of P2)



P1	C_P1=red	C_P1=white
RR	1	0
Rr	1	0
rr	0	1

(CPT of Color_P1)

P2	C_P2=red	C_P2=white
RR	1	0
Rr	1	0
rr	0	1

(CPT of Color_P2)

(CPT of Color_Child)

Child	C_Child=red	C_Child=white
RR	1	0
Rr	1	0
rr	0	1

(CPT of Child)

P1	P2	Child=RR	Child=Rr	Child=rr
RR	RR	1	0	0
RR	Rr	1/2	1/2	0
RR	rr	0	1	0
Rr	RR	1/2	1/2	0
Rr	Rr	1/4	1/2	1/4
Rr	rr	0	1/2	1/2
rr	RR	0	1	0
rr	Rr	0	1/2	1/2
rr	rr	0	0	1

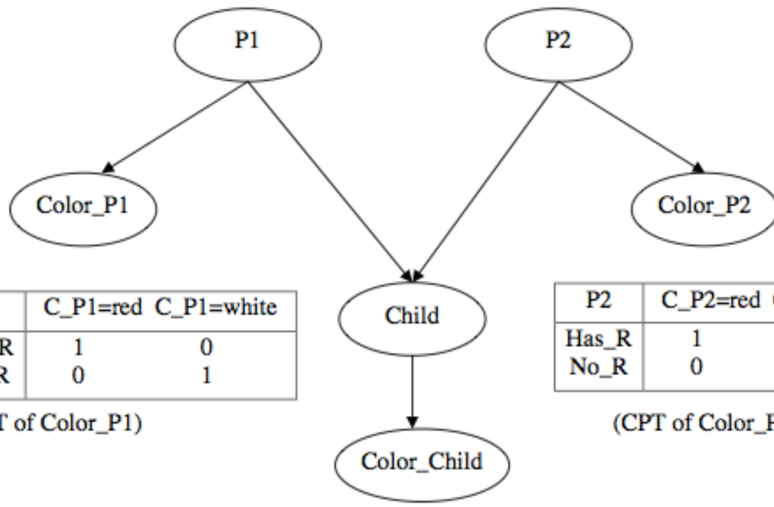
Figure 2. The compressed Bayesian network of "Genetics of Mendel's Peas"

P1=Has_R	P1=No_R
3/4	1/4

(CPT of P1)

P2=Has_R	P2=No_R
3/4	1/4

(CPT of P2)



P1	C_P1=red	C_P1=white
Has_R	1	0
No_R	0	1

(CPT of Color_P1)

P2	C_P2=red	C_P2=white
Has_R	1	0
No_R	0	1

(CPT of Color_P2)

Child	C_Child=red	C_Child=white
Has_R	1	0
No_R	0	1

(CPT of Color_Child)

P1	P2	Child=Has_R	Child=No_R
Has_R	Has_R	8/9	1/9
Has_R	No_R	2/3	1/3
No_R	Has_R	2/3	1/3
No_R	No_R	0	1

(CPT of Child)

Problem 3. (15 points) In this assignment, to make the representation more compact, we consider a compressed Genetics of Mendel's Peas that uses two values to represent three alleles: Has_R for Rr and RR, and No_R for rr. Furthermore, CPT values must be adjusted to match this change. We can prove that the two Bayesian networks are equivalent if we can observe only the color variables. Answer the following questions.

(a) (3 points) How is the CPT of P1 adjusted?

(b) (8 points) In the CPT of Child, the number of values is reduced from 27 to just 8 by compression. Show how to derive those numbers, such as 8/9 and 2/3 etc.

(c) (4 points) If the original network is compressed in a different way, using Has_r for Rr and rr and No_r for RR, is it possible to make proper changes in CPTs such that it still works equivalently when we can observe only the color variables? Why?

Problem 4. (35 points) Consider the compressed Genetics of Mendel's Peas, where all variables are Boolean, though they haves special alias for 0 and 1.

(a) (3 points) What is the Markov blanket of Child?

(b) (4 points) Is Color_P1 independent of Color_P2? Why, or why not?

(c) (4 points) Is Color_P1 independent of Color_P2 given Child? Why, or why not?

(d) (4 points) Suppose you are using likelihood weighting to compute the probability distribution of P1, Color_P1, Child and Color_Child, given the Color_P2 is white. What weight would you give to the sample (P1=has_R, P2=no_R, Child=has_R, Color_P1=red, Color_P2=white, Color_Child=red)?

(e) (10 points) What is the probability that Child has a color different from both parents? Show how to compute it.

(f) (10 points) What is the probability that Child and both parents have the same color? Show how to compute it.