International Electronic Journal of

Mathematics Education

Volume 4, Number 3, October 2009 www.iejme.com

Special issue on "Research and Developments in Probability Education"
Manfred Borovcnik & Ramesh Kapadia (Eds)

ALL APPENDICES TO

ON CONDITIONAL PROBABILITY PROBLEM SOLVING RESEARCH – STRUCTURES AND CONTEXTS

M. Pedro Huerta

To article

Glossary of terms

ISSN: 1306-3030

TABLES – PARTIALLY EXTENDED

Top
To article

Events, probabilities, and relations between probabilities

Basic events: A, B. Absolute and intersection probabilities arranged in a two-way table

	B	$\overline{B}$	
A	$p(A \cap B)$	$p(A \cap \overline{B})$	$p(A)$
$\overline{A}$	$p(\overline{A} \cap B)$	$p(\overline{A} \cap \overline{B})$	$p(\overline{A})$
	$p(B)$	$p(\overline{B})$	**1**

Columns as probability distributions

Column table	B	$\overline{B}$	
A	$p(A \mid B) = \frac{p(A \cap B)}{p(B)}$	$p(A \mid \overline{B}) = \frac{p(A \cap \overline{B})}{p(\overline{B})}$	$p(A)$
$\overline{A}$	$p(\overline{A} \mid B) = \frac{p(\overline{A} \cap B)}{p(B)}$	$p(\overline{A} \mid \overline{B}) = \frac{p(\overline{A} \cap \overline{B})}{p(\overline{B})}$	$p(\overline{A})$
	$1 = \frac{p(B)}{p(B)}$	$1 = \frac{p(\overline{B})}{p(\overline{B})}$	**1**

Rows as probability distributions

Row table	B	$\overline{B}$	
A	$p(B \mid A) = \frac{p(A \cap B)}{p(A)}$	$p(\overline{B} \mid A) = \frac{p(A \cap \overline{B})}{p(A)}$	$1 = \frac{p(A)}{p(A)}$
$\overline{A}$	$p(B \mid \overline{A}) = \frac{p(\overline{A} \cap B)}{p(\overline{A})}$	$p(\overline{B} \mid \overline{A}) = \frac{p(\overline{A} \cap \overline{B})}{p(\overline{A})}$	$1 = \frac{p(\overline{A})}{p(\overline{A})}$
	$p(B)$	$p(\overline{B})$	**1**

Table 1 ***extended***. Events, probabilities, and relations between probabilities.

Events and probabilities

Basic events	4 events	4 intersection events	4 absolute probabilities	4 intersection probabilities	4 conditional probabilities each from columns	from rows
A	A	$A \cap B$	$p(A)$	$p(A \cap B)$	$p(A \mid B)$	$p(B \mid A)$
	$\overline{A}$	$\overline{A} \cap B$	$p(\overline{A})$	$p(\overline{A} \cap B)$	$p(\overline{A} \mid B)$	$p(\overline{B} \mid A)$
B	B	$A \cap \overline{B}$	$p(B)$	$p(A \cap \overline{B})$	$p(A \mid \overline{B})$	$p(B \mid \overline{A})$
	$\overline{B}$	$\overline{A} \cap \overline{B}$	$p(\overline{B})$	$p(\overline{A} \cap \overline{B})$	$p(\overline{A} \mid \overline{B})$	$p(\overline{B} \mid \overline{A})$

Probability relationships

Source	6 complementary relationships	4 intersection relationships	8 conditional relationships
Co-lumns	$p(A \mid B) + p(\overline{A} \mid B) = 1$	$p(B) = p(A \cap B) + p(\overline{A} \cap B)$	$p(B) \times p(A \mid B) = p(A \cap B)$ $p(B) \times p(\overline{A} \mid B) = p(\overline{A} \cap B)$
	$p(A \mid \overline{B}) + p(\overline{A} \mid \overline{B}) = 1$	$p(\overline{B}) = p(A \cap \overline{B}) + p(\overline{A} \cap \overline{B})$	$p(\overline{B}) \times p(A \mid \overline{B}) = p(A \cap \overline{B})$ $p(\overline{B}) \times p(\overline{A} \mid \overline{B}) = p(\overline{A} \cap \overline{B})$
Rows	$p(B \mid A) + p(\overline{B} \mid A) = 1$	$p(A) = p(A \cap B) + p(A \cap \overline{B})$	$p(A) \times p(B \mid A) = p(A \cap B)$ $p(A) \times p(\overline{B} \mid A) = p(A \cap \overline{B})$
	$p(B \mid \overline{A}) + p(\overline{B} \mid \overline{A}) = 1$	$p(\overline{A}) = p(\overline{A} \cap B) + p(\overline{A} \cap \overline{B})$	$p(\overline{A}) \times p(B \mid \overline{A}) = p(\overline{A} \cap B)$ $p(\overline{A}) \times p(\overline{B} \mid \overline{A}) = p(\overline{A} \cap \overline{B})$
Mar-gins	$p(A) + p(\overline{A}) = 1$ $p(B) + p(\overline{B}) = 1$		

Table 1 *extended.* Events, probabilities, and relations between probabilities – continued.

Top
To article

	Witness accuracy (80%)		
	Witness says Blue	Witness says Green	Total
Blue Cabs	120	30	150
Green Cabs	170	680	850
Total	290	710	1000

Table 2a. Data for the taxi problem in a two-way table.

	Witness accuracy		
	Witness says Blue	Witness says Green	Total
Blue Cabs	120		
Green Cabs			850
Total	290		1000

Table 2b. Data for the taxi problem reformulated – presented in terms of ternary problem.

	L_0			**L_1**			**L_2**			**L_3**		
C_0	C_0T_1	$\varnothing$	$\varnothing$	C_0T_1	C_0T_2	C_0T_3	C_0T_1	C_0T_2	C_0T_3	C_0T_1	C_0T_2	C_0T_3
C_1	C_1T_1	$\varnothing$	$\varnothing$	C_1T_1	C_1T_2	C_1T_3	C_1T_1	C_1T_2	C_1T_3	$\varnothing$	$\varnothing$	$\varnothing$
C_2	C_2T_1	$\varnothing$	$\varnothing$	C_2T_1	$\varnothing$	C_2T_3	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

Table 3. Classification of ternary problems of conditional probability into families and subfamilies.

Top

To article

Problem	Situation	Context	Phenomena referring to: events	Phenomena referring to: conditional probabilities	Specific terms	Classification and data format
P_1	Diagnostic test	Health context	• Be tubercular • Not be tubercular • Test gives a positive result	•If a person is tubercular, the test yields a positive result with a high probability (in % format) •If a person is not tubercular, the test yields a positive result with a small probability (in % format) •If the test is positive there is a probability (<1) that the person actually is tubercular	•Positive result in tests	$L_2C_1T_1$ family (1, 0, 2) form • Rate • Percentages • $p(D\|+)$
P_2			• Be diabetic • A person does not suffer from diabetes but is positive on test • A person suffers from diabetes but is negative on test	•FPC or FPR (False Positive Coefficient or Rate) •FPC or FNR (False negative Coefficient or Rate)	•FPC or FPR •FNC or FNR •Prevalence of diabetes •Test is positive •Test is negative •Diagnostic Test	$L_2C_1T_1$ family •(1,0,2) •a) $p(D\|+)$ •b) $p(\overline{D}\|-)$ •Percentages
P_3			• Not suffer from uterine cancer • Suffer from uterine cancer • Test positive in diagnostics without uterine cancer • Test negative in diagnostics with uterine cancer • A person tested negative who does not suffer from uterine cancer	•False positive coefficient or rate (FPC or FPR) •False negative coefficient or rate (FNC or FPR)	•False positive coefficient or rate •False negative coefficient or rate •Pre-test probability •Negative result in test •Diagnostic Test	$L_2C_1T_1$ family •(1,0,2) • $p(\overline{D}\|-)$ •Probability
P_4			• Be infected by tuberculosis. • NOT be infected by tuberculosis.	•Sensitivity •Specificity •False positive	•Sensitivity •Specificity •False positive •Prevalence of disease •Tuberculin Test	$L_3C_0T_2$ family •(0,0,3) • $p(D)$ •Probability

Table 4. Aspects of the phenomenological analysis of the problems 1 to 4.

Top To article

Pre-test				Post-test			
D	$\overline{D}$	$+$	$-$	$D\cap +$	$\overline{D}\cap +$	$D\cap -$	$\overline{D}\cap -$
Be ill Be infected	Not be ill Not be infected	Test positive	Test negative	Suffer	Not suffer	Suffer	Not suffer
				from a specific disease			
Suffer	Not suffer	in the diagnostics, regardless of the health status		**and** test positive	**and** test positive	**and** test negative	**and** test negative
from a specific disease				(or similar phrases referring to the conjunction)			

Table 5. Reference sets in the Diagnostic Test in Health Context.

Phenomena (referring to probabilities)	Specific Terms	Organizational means	Data Format
Mistakes produced by test	FPR	$p\,(+\vert\overline{D}\,)$	Necessarily in percentages and probabilities
	FNR	$p\,(-\vert D\,)$	
Success produced by test “VALIDITY”	Sensitivity	$p\,(+\vert D\,)$	
	Specificity	$p\,(-\vert\overline{D}\,)$	
Mistakes produced in the diagnostic procedure “DIAGNOSTIC ERRORS”	negative diagnosis is false (often called “false negative”)	$p\,(D\,\vert -)$	
	positive diagnosis is false (often called “false positive”)	$p\,(\overline{D}\,\vert +)$	
Success produced in the diagnostic procedure “PREDICTIVE VALUES”	PPV or Positive Predictive Value Positive diagnosis is correct	$p\,(D\,\vert +)$	
	NPV or Negative Predictive Value Negative diagnosis is correct	$p\,(\overline{D}\,\vert -)$	
to have the disease (pre-test)	Prevalence of the disease	$p\,(D\,)$	Reasonably in natural numbers or absolute or natural frequencies
not to have the disease (pre-test)	Prevalence of no disease	$p\,(\overline{D}\,)$	
Results from the diagnostic test	To test positive	$p\,(+)$	
	To test negative	$p\,(-)$	
to have the disease *and* to test positive in the diagnostic procedure	We didn’t find out or it does not exist	$p\,(D\cap +)$	
not to have the disease *and* to test positive in the diagnostic procedure		$p\,(\overline{D}\cap +)$	
to have the disease *and* to test negative in the diagnostic procedure		$p\,(D\cap -)$	
not to have the disease *and* to test negative in the diagnostic procedure		$p\,(\overline{D}\cap -)$	

Table 6. Results of the phenomenological analysis in Diagnostic Test in Health Context.

Top To article

Updating prevalence of *D* consists of solving two problems

Given prevalence of D, the probability of + and the sensitivity of the test, calculate:

a) the positive predictive value (PPV) b) the False negative rate (FNR).

Prevalence of D	Refers to	Context of terms: health	Context of terms: mathematics	Context of involved relationships: health	Context of involved relationships: mathematics
Prior to test	All people	Prevalence of D	Absolute probability: $p(D)$	—	—
Updated	A person who has been tested	Probability of D conditioned by test result:			
		a) Positive predictive value (PPV)	$p(D\mid +)$	$= PPV = \dfrac{\text{prevalence } D \times \text{sensitivity}}{\text{probability } (+)}$	$= \dfrac{p(D)\times p(+\mid D)}{p(+)}$
		b) False negative rate (FNR)	$p(D\mid -)$	$=$ False negative $= \dfrac{\text{prevalence } D \times (1-\text{sensitivity})}{1-\text{probability } (+)}$	$= \dfrac{p(D)\times[1-p(+\mid D)]}{1-p(+)}$

Table next to Figure 7 *extended*. Updating prevalence of D consists of solving two problems.

Top

To article

ENLARGED FIGURES

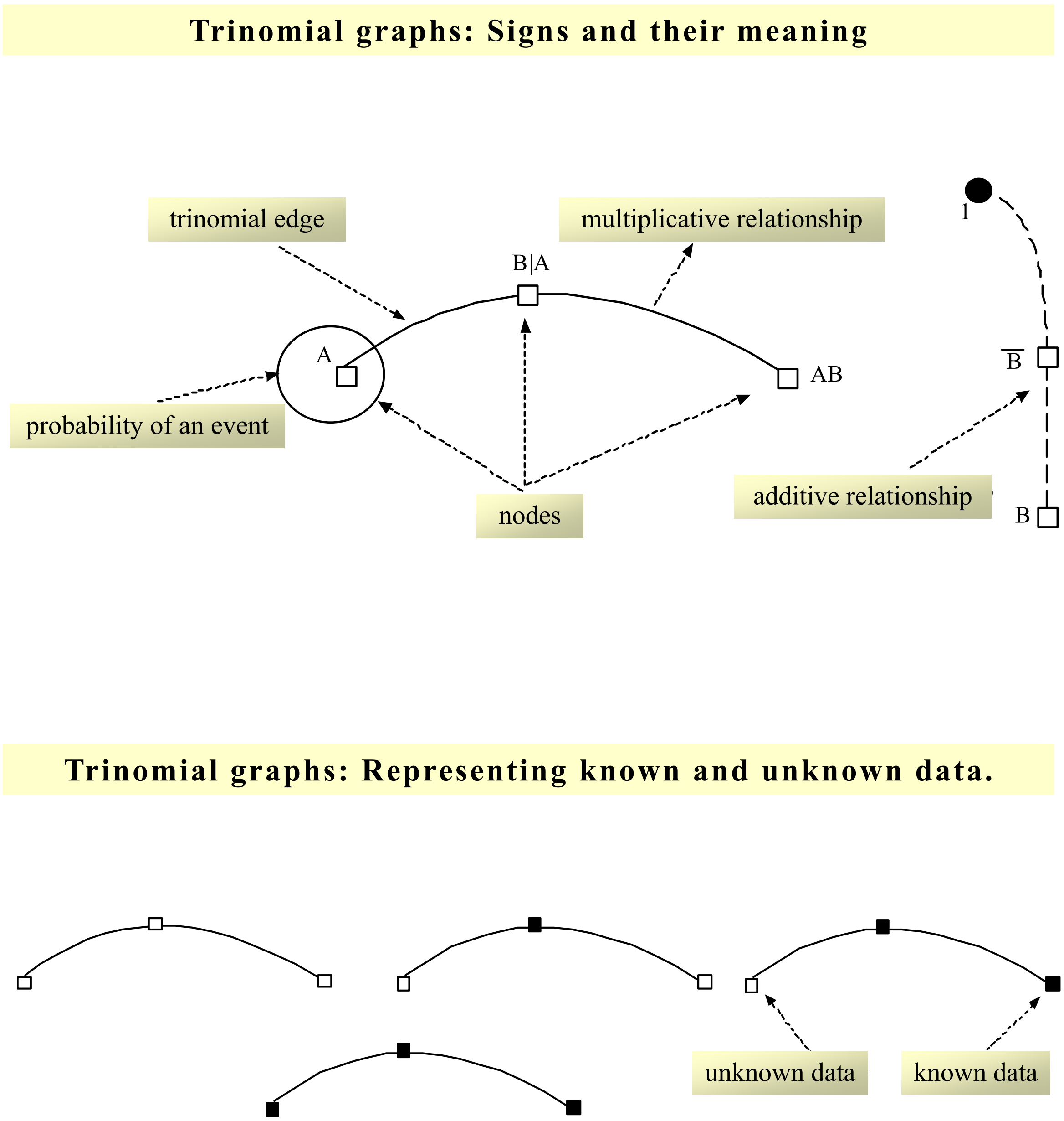

Figure 1. Signs and their meaning in trinomial graphs.

Top
To article

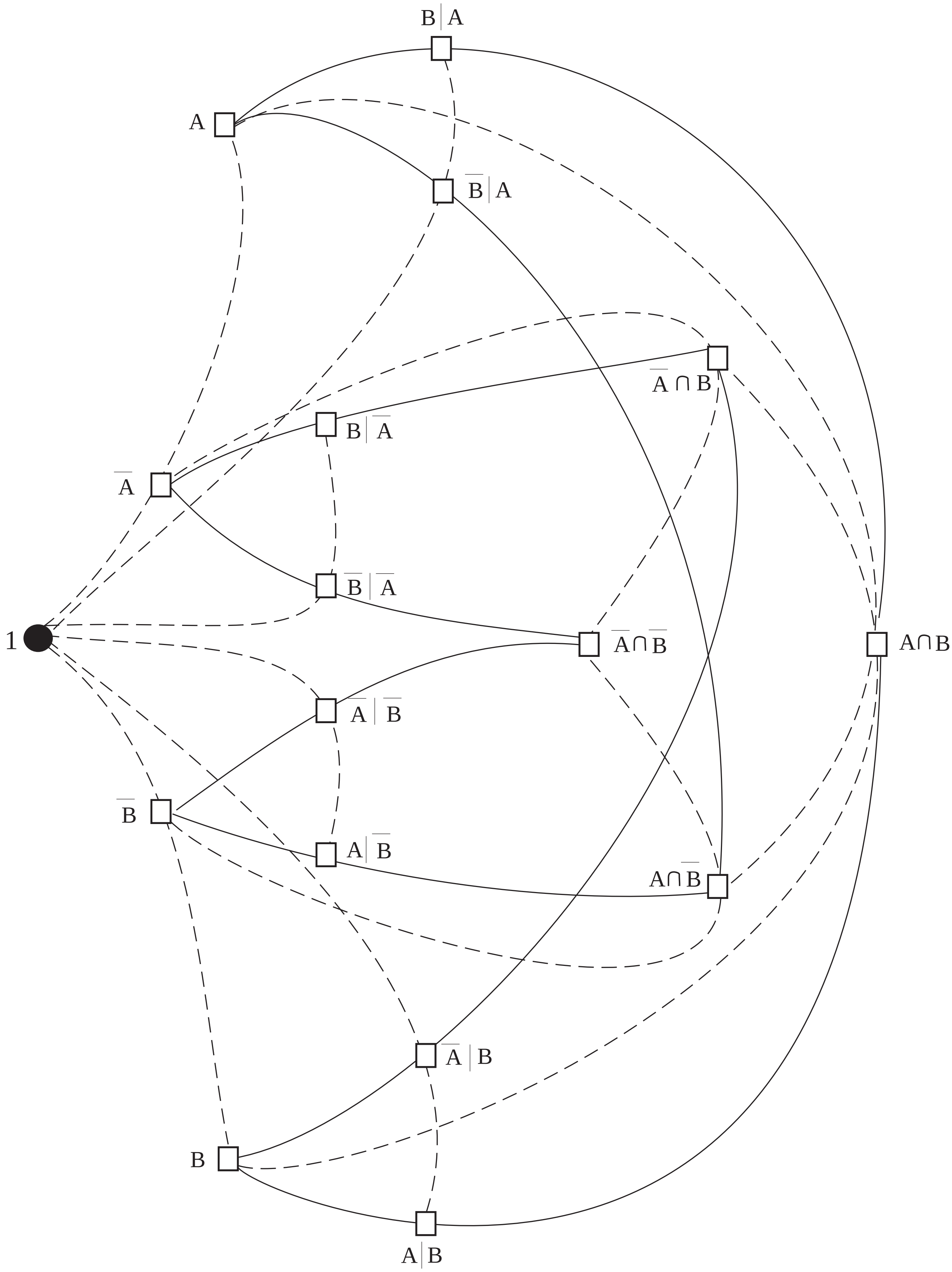

Figure 2. Trinomial Graph of ternary problems of conditional probability.

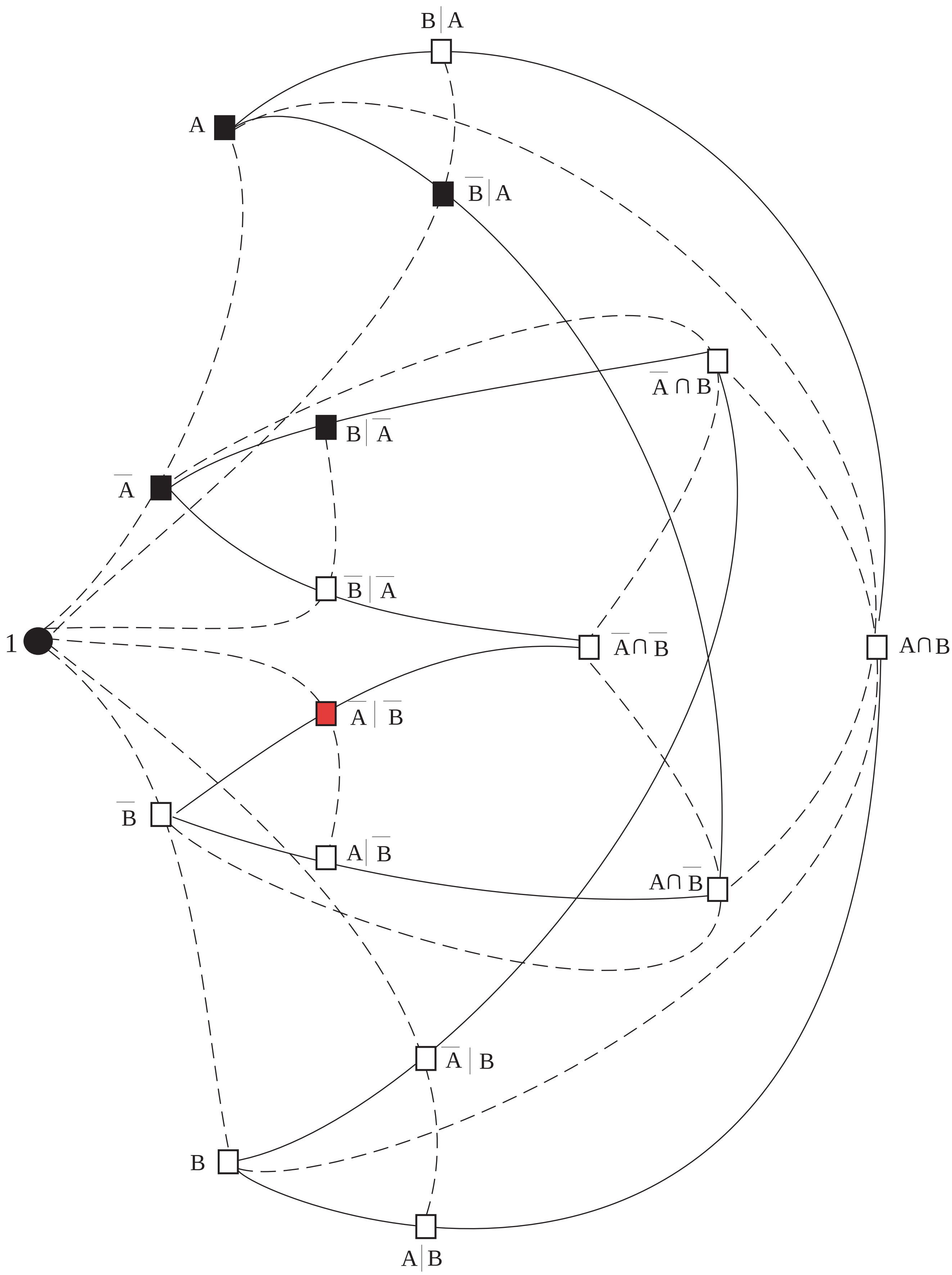

Figure 3. Trinomial graph for medical diagnosis task 2.

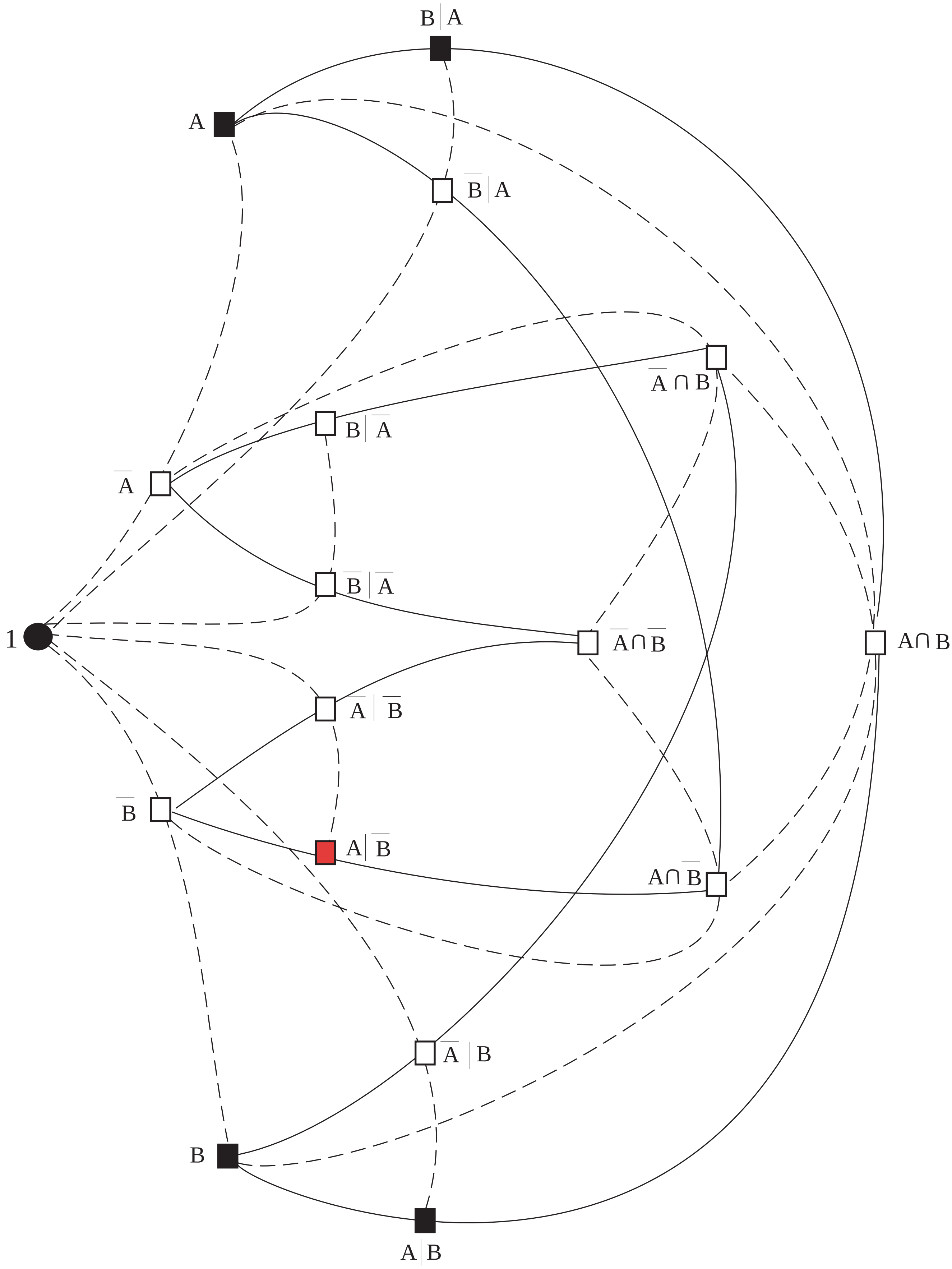

Figure 4a. Trinomial graph for redundant problem.

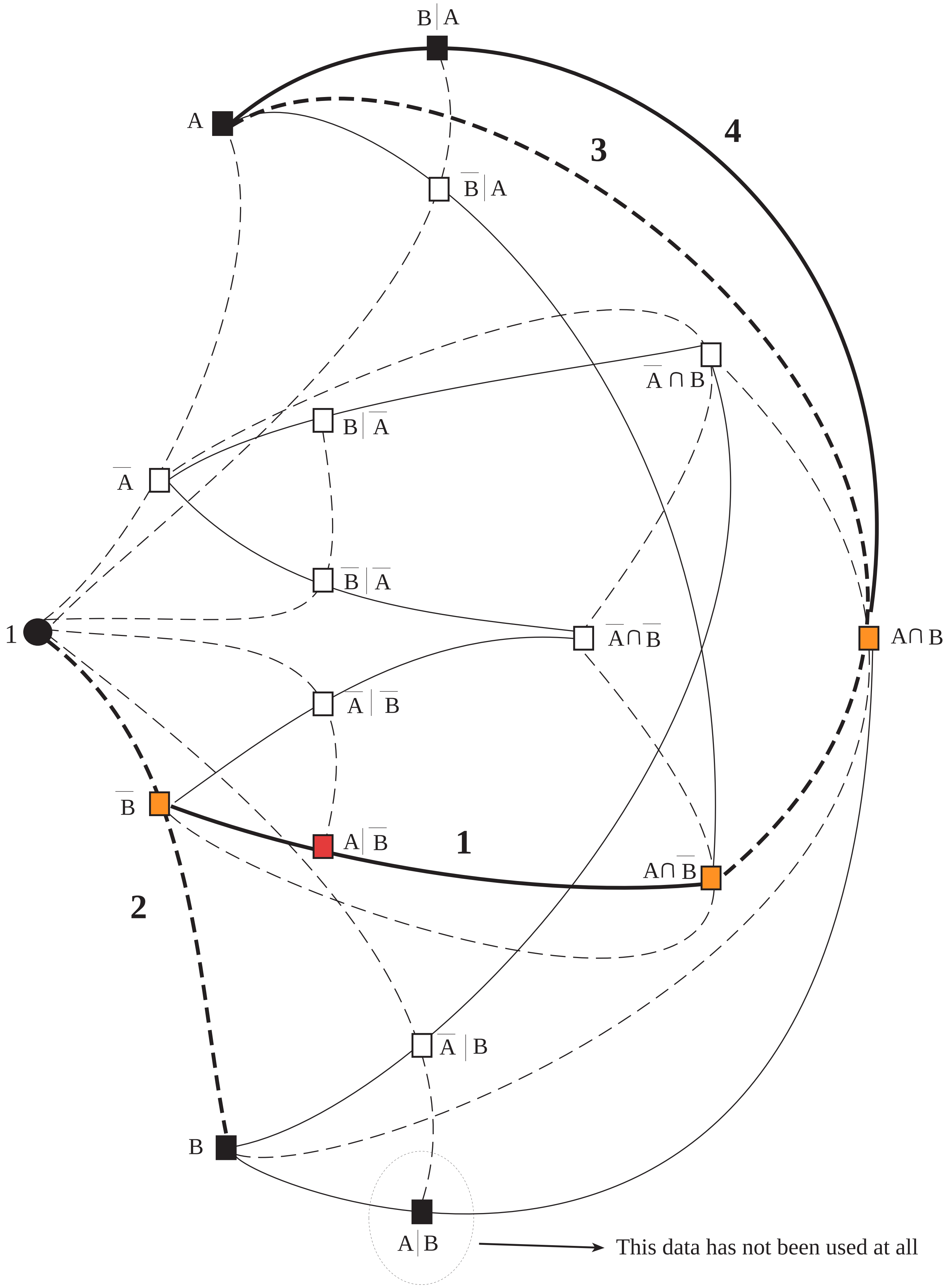

Figure 4b. Final state of algorithm.

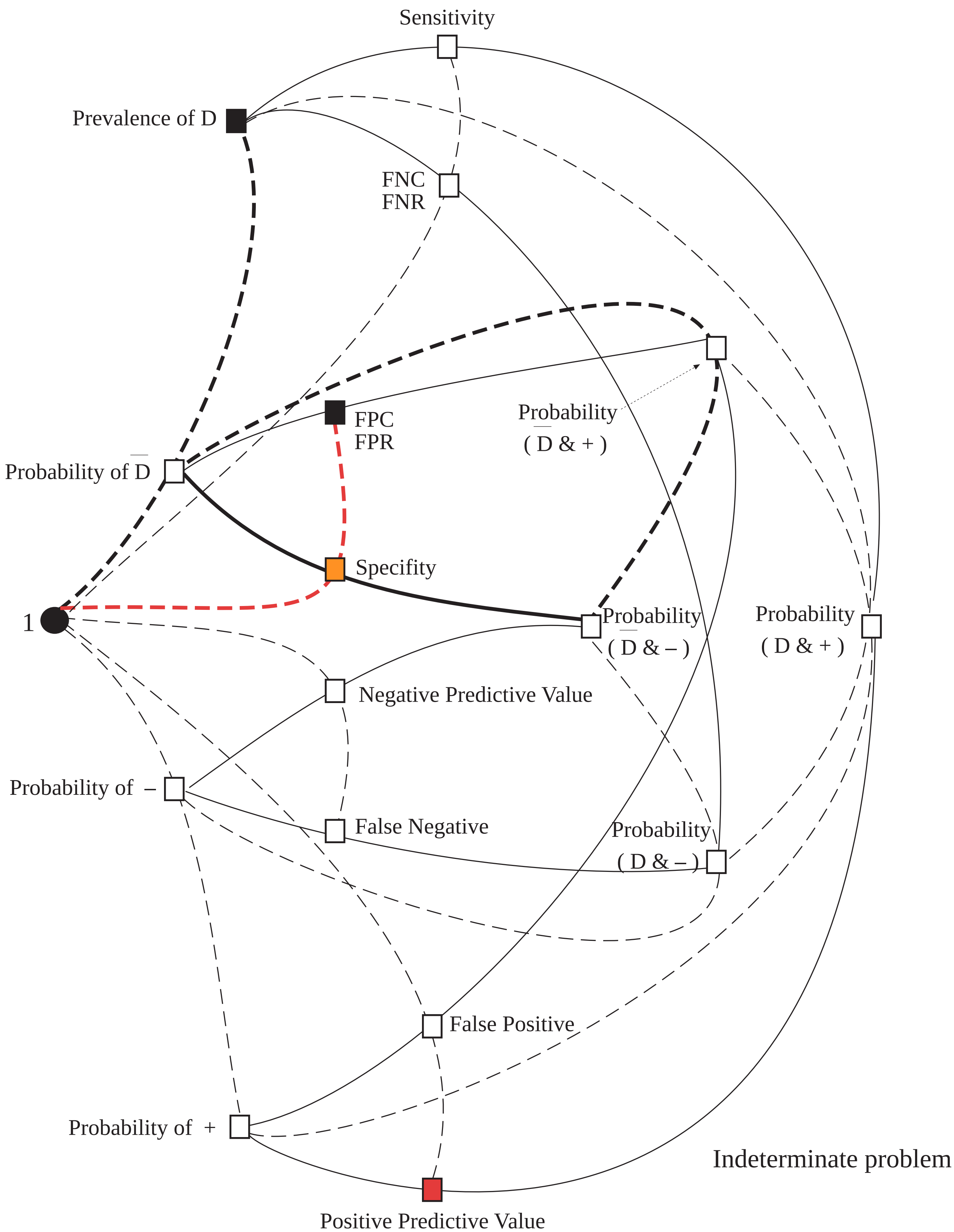

Figure 5. Indeterminate problem.

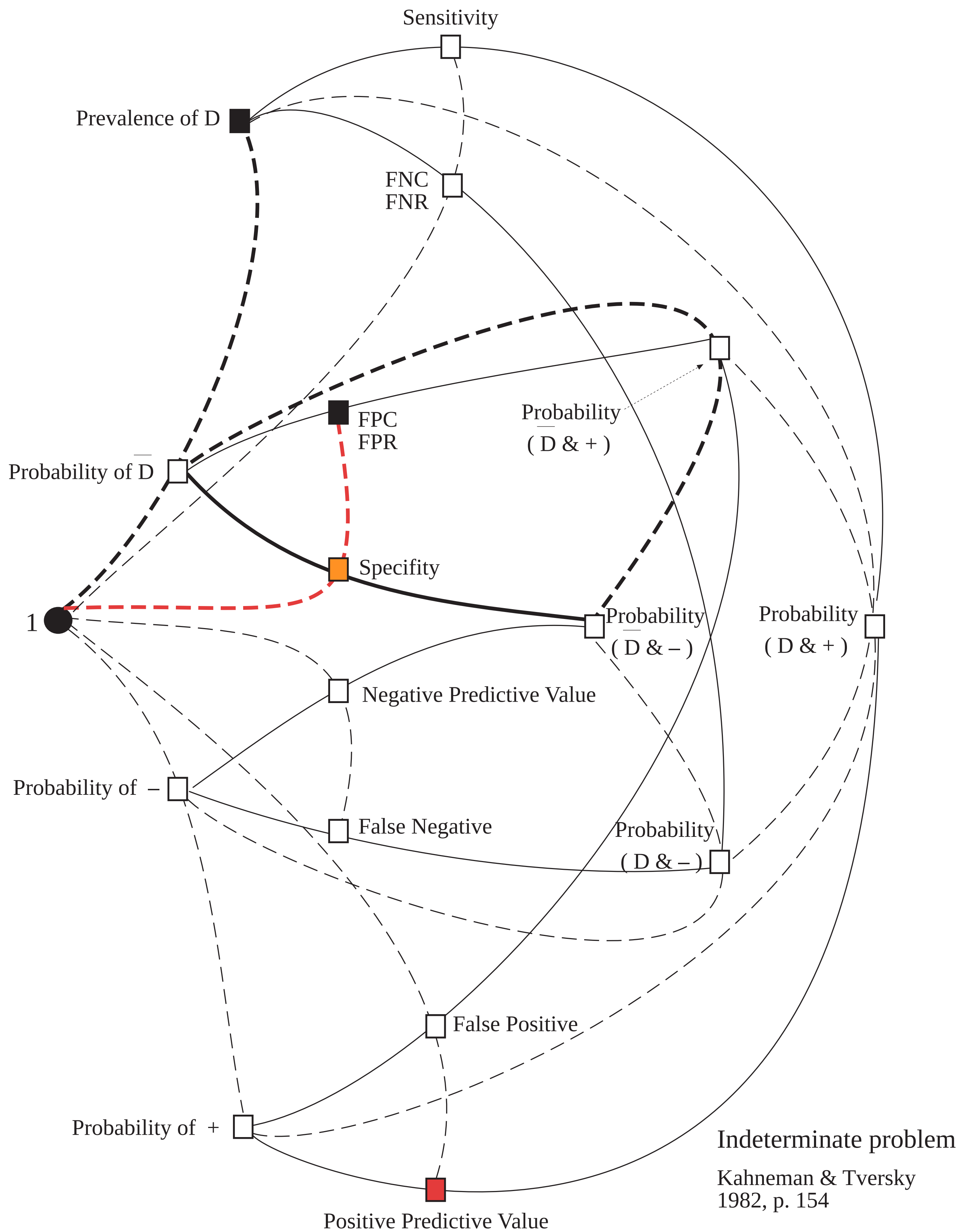

Figure 6a. Modified disease problem.

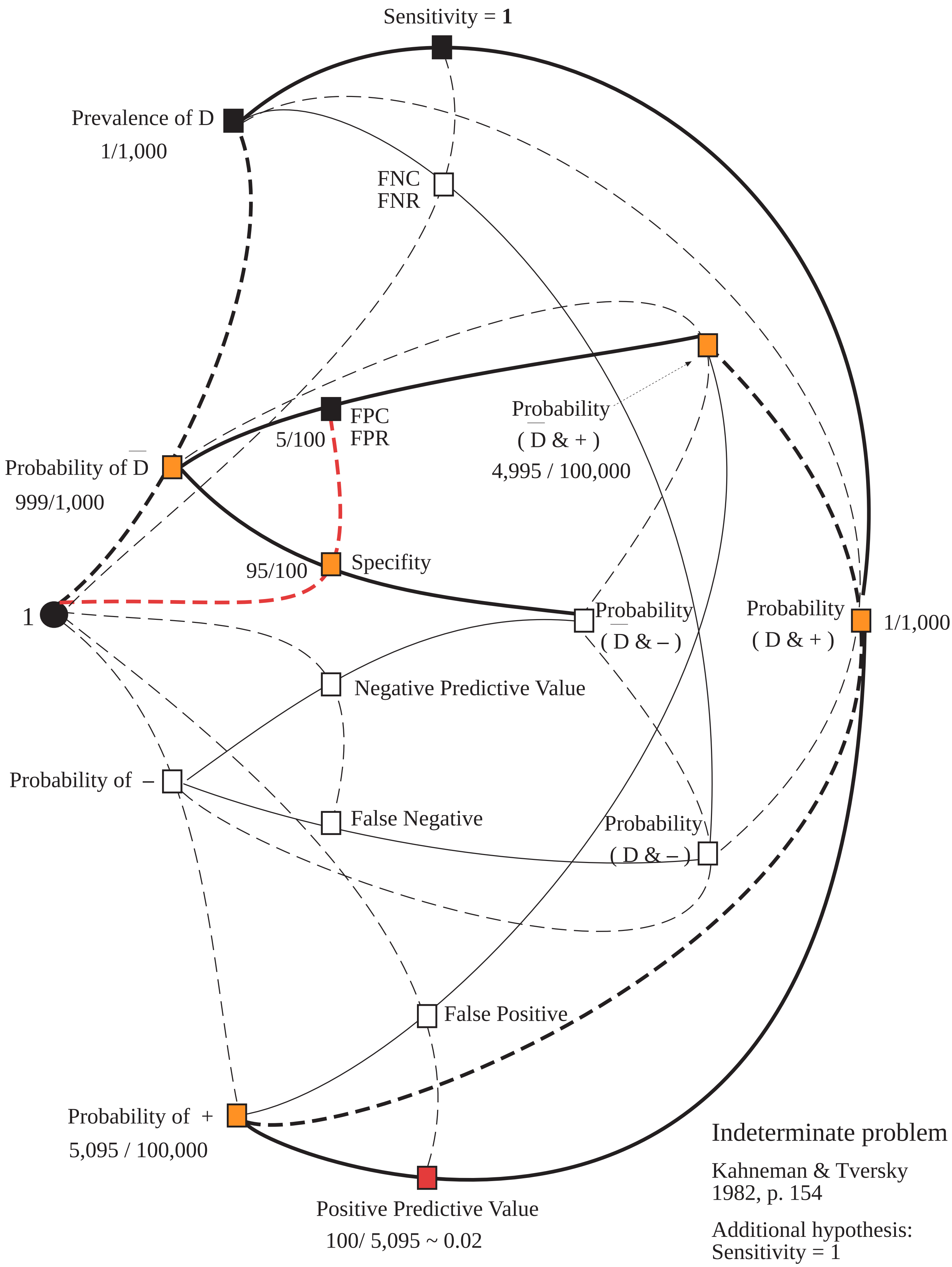

Figure 6b. Misinterpretation of false positive rate.

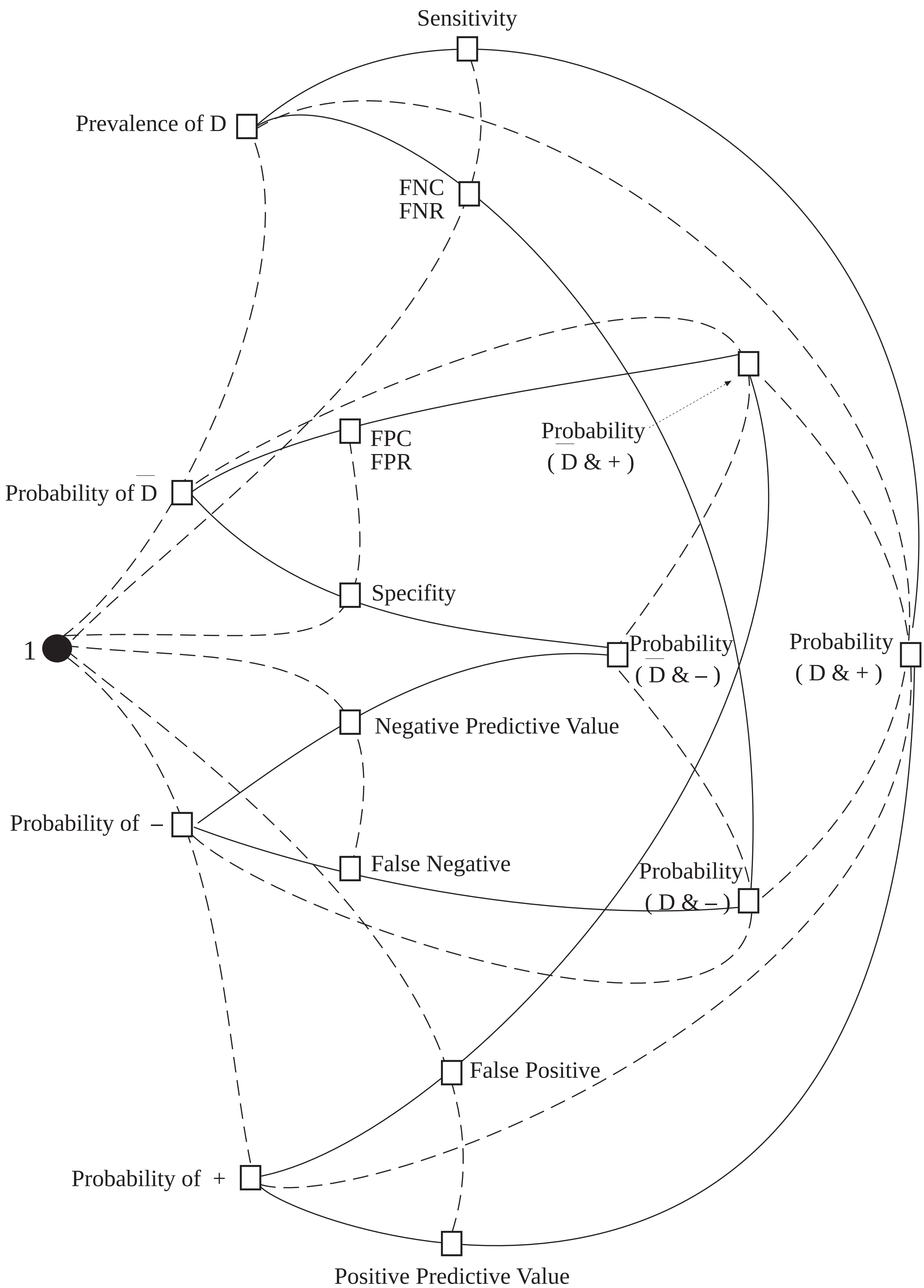

Figure 7. Graph of the world of ternary problems in the diagnostic test in health context.

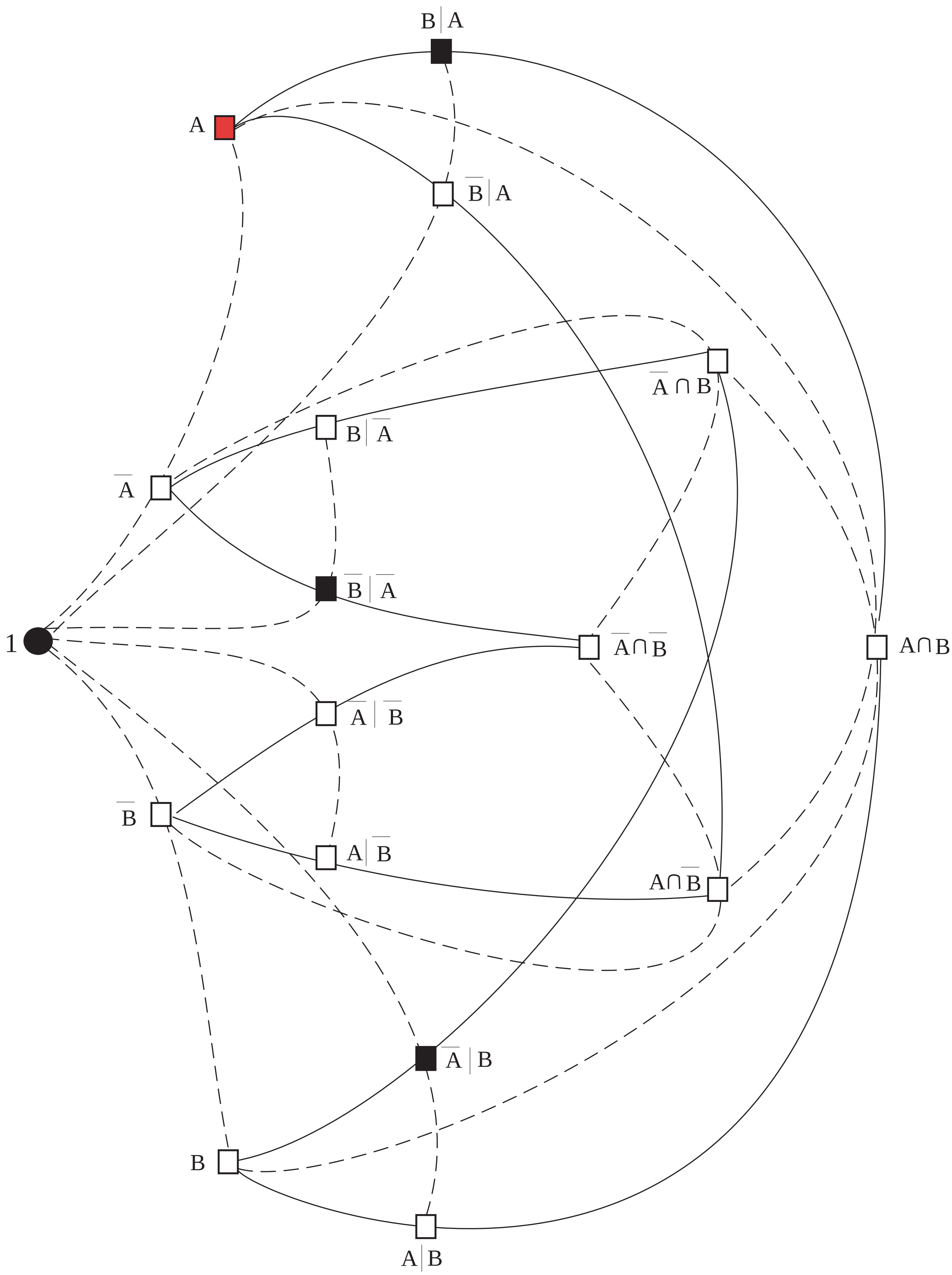

Figure 8. Trinomial graph for P_4

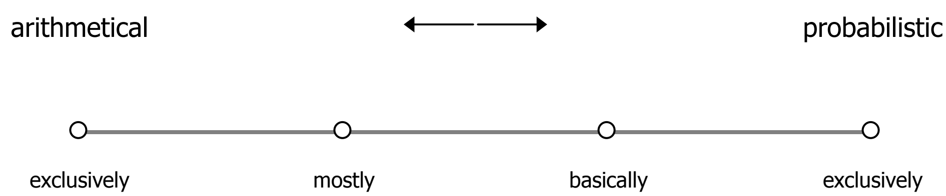

Figure 9. Four categories of thinking along the opposite pair of arithmetical and probabilistic.

Top
To article

Author : M. Pedro Huerta
E-mail : manuel.p.huerta@uv.es
Address : Departament de Didàctica de la Matemàtica. Universitat de València. Spain