Bayesian Concepts in Software Testing: An Initial Review

Daniel Rodríguez
Universidad de Alcalá

Javier Dolado
U. País Vasco/Euskal Herriko Unibertsitatea

Javier Tuya
Universidad de Oviedo

PRESI TIN2013-46928-C3-1-R, TIN2013-46928-C3-2-R

Outline

- Bayes-Laplace Rule
  - Bayes Theorem
  - Frequentism vs Bayesianism
  - Example: Breast Cancer
  - Example: Drunk Driver
- Bayesian Networks
  - Definition
  - Basic Example: Defects insertion
  - Inferences
- Literature Review (2010 onwards)

Bayes' rule or **Bayes-Laplace rule**

- Bayes' Rule: most probably due to **Reverend Thomas Bayes** (1702-1761, Kent, England) (or to *Nicholas Saunderson* 1682-1739)
- Only published one mathematical paper in his entire life.
- After his death published *'An essay towards solving a problem in the doctrine of chances'*, 1763, submitted by Richard Price
- **Pierre-Simon Laplace** (1749-1827) independently discovered Bayes' rule in 1812 and made it operational
- Bayesian methods were set aside in favour of non-Bayesian (frequentist) methods in the 20th century

Bayes' Rule

Two definitions of Probability:

- ***Frequentism***: the definition of probability is related to the *frequency of an event*. The parameters of interest are fixed but the data are a repeatable random sample, hence there is a frequency. No prior information is used. In a strict frequentist view, it does not make sense to talk about the true value of the parameter θ under study. The true value of θ is fixed, by definition. Frequentists compute **P(data|θ),** which is the probability of observing the data given the null hypothesis.
- ***Bayesianism***: the definition of probability is related to the *level of knowledge* about an event. The value of knowledge about an event is based on **prior** information and the available data. The parameters of interest are unknown and the data are fixed. From a Bayesian viewpoint we can talk about the probability that the true value of the parameter θ lies in an interval. Bayesians compute **P(θ|data)**, which is the probability of a given outcome, given the data.

Bayes' Rule

- It is all about *conditional probabilities*
- The rule allows to make inferences from *causes to effects (symptoms)* and from *effects to causes*
- Given two events, H and E,

$$P(H \cap E) = P(H|E) \cdot P(E)$$

$$P(E \cap H) = P(E|H) \cdot P(H)$$

$$P(H \cap E) = P(E \cap H)$$

$$P(H|E) \cdot P(E) = P(E|H) \cdot P(H)$$

Bayes' Rule

- The conditional probability of H, given E, is

POSTERIOR

LIKELIHOOD

PRIOR

$$P(H|E)=\frac{P(E|H)\cdot P(H)}{P(E)}$$

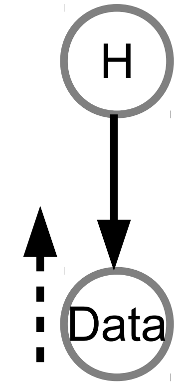
MARGINAL LIKELIHOOD
or EVIDENCE

P(E) is $P(E)=P(E|H)\cdot P(H)+P(E|\neg H)\cdot P(\neg H)$

P(E) acts as a constant

$$\boxed{Posterior \propto Likelihood \cdot Prior}$$

Example: Breast Cancer



$$P(Hypothesis|Data)=\frac{P(Data|Hypothesis)P(Hypothesis)}{P(Data)}$$

- It is a textbook example that shows how to infer causes from the symptoms. The data below is for the sake of example. Factual data can be found at http://www.breastcancer.org

1% of women between age forty-fifty who participate in routine screening have breast cancer.

90% of women with breast cancer will get positive mammographies.

9.6% of women without breast cancer will also get positive mammographies.

A woman in this age group had a positive mammography in a routine screening.

What is the probability that she actually has breast cancer?

$$P(Breast\ Cancer=present|\mathrm{Test+}=positive)\,?$$

Example: Breast Cancer

- The random variables that we define are

Variable	Value	When the Variable Takes This Value
– **Breast Cancer**	positive	breast cancer is present
	negative	breast cancer is not present
– **Test+**	positive	the test result is positive
	negative	the test result is negative

$$P(Breast\ Cancer \mid \mathrm{Test+}) = \frac{P(\mathrm{Test+} \mid BreastCancer)\,P(BreastCancer)}{P(\mathrm{Test+})}$$

$$P(present \mid positive) = \frac{P(positive \mid present)\,P(present)}{P(positive)}$$

Example: Breast Cancer

$$P(present|positive)=\frac{P(positive|present)P(present)}{P(positive)}$$

1% of women between age forty-fifty who participate in routine screening have breast cancer.

90% of women with breast cancer will get positive mammographies. (TRUE POSITIVES)

9.6% of women without breast cancer will also get positive mammographies (FALSE POSITIVES)

$$P(present|positive)=\frac{P(positive|present)P(present)}{P(positive)}=$$

$$=\frac{(0.90)\cdot(0.01)}{P(positive)}=\frac{0.009}{0.10404}=0.08650519\approx 8.6\%$$

$$P(positive)=P(positive|present)P(present)+P(positive|absent)P(absent)=$$

$$=(0.90)\cdot(0.01)+(0.096)\cdot(0.99)=(0.009)+(0.09504)=0.10404$$

Bayes' rule Graphical Illustration 1

1% of women between age forty-fifty who participate in routine screening have breast cancer.

90% of women with breast cancer will get positive mammographies. (TRUE POSITIVES)

9.6% of women without breast cancer will also get positive mammographies (FALSE POSITIVES)

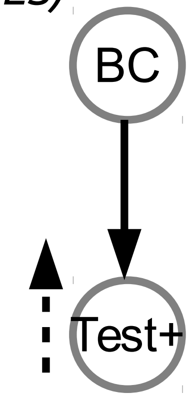


Table with Conditional Probabilities

Breast Cancer (is the Hypothesis)

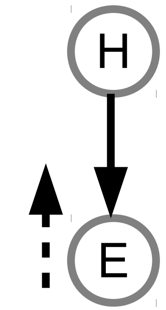
Test+ (is the Evidence)	*present* **1%**	*absent* **99%**
positive	**90%** (True Positive Rate)	**9.6%** (False Positive Rate)
negative	**10%** (False Negative Rate)	**90.4%** (True Negative Rate)

Sensitivity = recall = 90%
Specificity = 90.4%

Type I Error, α, = 9.6%
Type II Error, β, = 10%

$$P(present|positive)=\frac{(0.01)(0.9)}{(0.01)(0.9)+(0.096)(0.99)}\approx 8.6\,\%$$

Bayes' rule Graphical Illustration 2



1% out of 10000 = 100 with breast cancer
90% show up positive test = 90
9.6% show up false positive test = 960

Table with Frequencies

Breast Cancer

	present	absent	
Test+ positive	90 (true positives)	960 (false positives)	1050
negative	10 (false negatives)	8940 (true negatives)	8950
	100	9900	10000

90/(90+960) = 90/1050 = **8.6%**

We could also change the prior

Test+++

1% of women between age forty-fifty who participate in routine screening have breast cancer.

We have availabe a new test procedure Test+++ such that

90% of women with breast cancer will get positive mammographies. (TRUE POSITIVES)

***0.01% of women** without breast cancer will also get positive mammographies (FALSE POSITIVES)*

Table with Conditional Probabilities

Test+++	Breast Cancer present **1%**	Breast Cancer absent **99%**
positive	**90%** (True Positive R)	**0.01%** (False Positive Rate)
negative	**10%** (False Negative Rate)	**99.99%** (True Negative Rate)

$$P(present|positive) = \frac{(0.9)(0.01)}{(0.9)(0.01)+(0.0001)(0.99)} = 98.91\%$$

Bayes' rule

- Bayesianism was used to prove that mass screening for a large population for a rare disease such as AIDS or other was unprofitable.
- Another case is prostate cancer. The disease is so rare that the tests for identifying the specific agent result in most cases returning positive results for a patient that does not have cancer.

Example 2: Drunk Driver

- A group of policemen have breathalyzers displaying false drunkenness in 5% of the cases in which the driver is sober.
- However, the breathalyzers never fail to detect a truly drunk person.
- One in a thousand drivers are driving drunk.
- Suppose the policemen then stop a driver at random, and force the driver to take a breathalyzer test. It indicates that the driver is drunk.
- We assume you don't know anything else about him or her.
- **How high is the probability he or she really is drunk?**

Example 2: Drunk Driver

- We must find the probability that the driver is drunk given that the breathalyzer indicated they are drunk.

Drunk Driver

Breathalyzer	*Drunk* 0.001	*Sober* 0.999
positive	**100%** (True Positive Rate)	**5%** (False Positive Rate)
negative	**0%** (False Negative Rate)	**95%** (True Negative Rate)

$$P(Drunk|positive)=\frac{P(positive|Drunk)P(Drunk)}{P(positive)}=\frac{(0.001)(1.0)}{(0.001)(1.0)+(0.05)(0.999)}$$

$$=\frac{(0.001)}{(0.001)(1.0)+(0.04995)}=\frac{(0.001)}{0.05095}=0.019627085\approx 2\%$$

Bayes' Rule

$$P(Hypothesis|Data) = \frac{P(Data|Hypothesis)P(Hypothesis)}{P(Data)}$$

If instead of having H and ¬H, only one hypothesis and its negation, there are **n** Hypotheses *mutually exclusive and exhaustive*, the rule becomes

$$P(Hypothesis_j|Data) = \frac{P(Data|Hypothesis_j)P(Hypothesis_j)}{\sum_{1\ldots n}^{n} P(Data|Hypothesis_1)P(Hypothesis_1)+\ldots+P(Data|Hypothesis_n)P(Hypothesis_n)}$$

- then Bayesian Networks and their graphical representation come into play in order to ease the computations when there are many variables involved

Bayesian model comparison

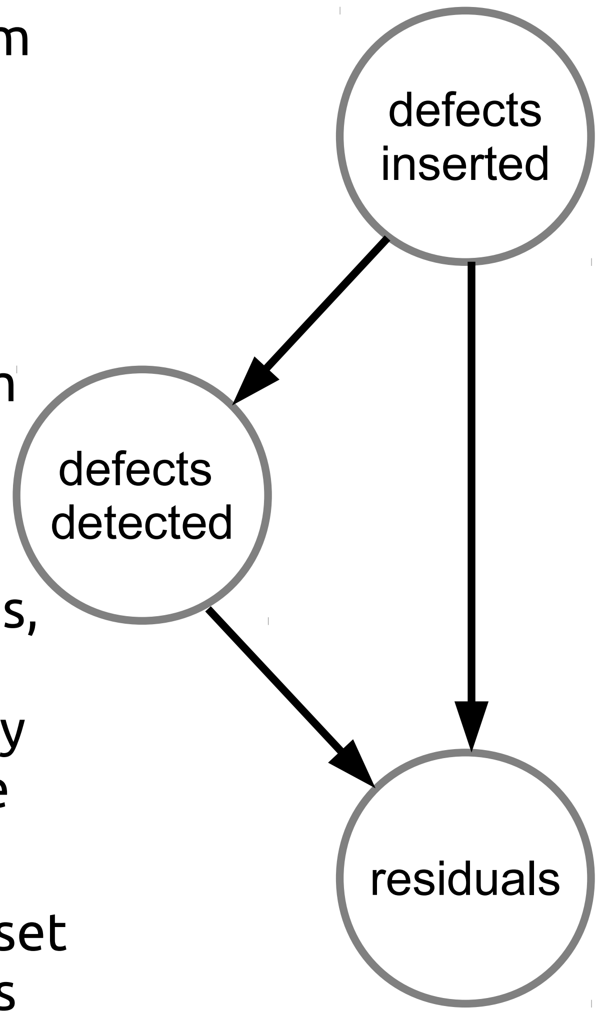
Bayesian model comparison is performed by means of *posterior odds*. The posterior odds ratio for a model M1 against another model M2 involves a ratio of marginal likelihoods, the so-called *Bayes factor*

$$\frac{P(M_1|x)}{P(M_2|x)} = \frac{P(x|M_1)}{P(x|M_2)} \frac{P(M_1)}{P(M_2)}$$

$$\boxed{Posterior\ odds = Bayes\ factor \cdot Prior\ odds}$$

Bayesian Networks

- A Bayesian network, Bayes network or belief network is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG).
- It allows to make inferences from causes to symptoms and from symptoms to causes
- Bayesian networks are DAGs whose nodes represent random variables, unknown parameters or hypotheses.
- Edges represent conditional dependencies.
- Each node is associated with a probability function that takes, as input, a particular set of values for the node's parent variables, and gives (as output) the probability (or probability distribution, if applicable) of the variable represented by the node.
- Let G=(V,E) a Directed Acyclic Digraph (DAG) composed of a set of vertices V and a set of edges E among the pairs of vertices of V. Let P be a joint probability distribution of the random variables in the set of vertices V. We call (G,P) a Bayesian Network if (G,P) satisfies the Markov condition.



Example: Basic bayesian network for defect estimation

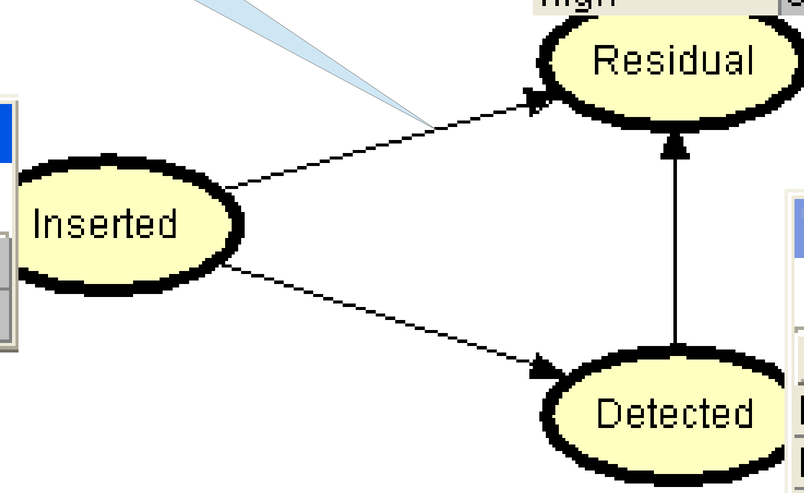
- A basic example involving three variables.
- The expert has decided that in order to model their predictive problem
  - Defects inserted: by the development group with different probabilities.
  - Defects detected: Probability of detecting defects depending on the defects inserted
  - Residual defects: Probability of remaining defects depending on the inserted and detected

1) Create the network with the conditional probabilities tables

Residual

Edit Functions View

Inserted	low		high	
Detected	low	high	low	high
low	0.4	0.8	0.05	0.7
high	0.6	0.2	0.95	0.3



Inserted

Edit Functions View

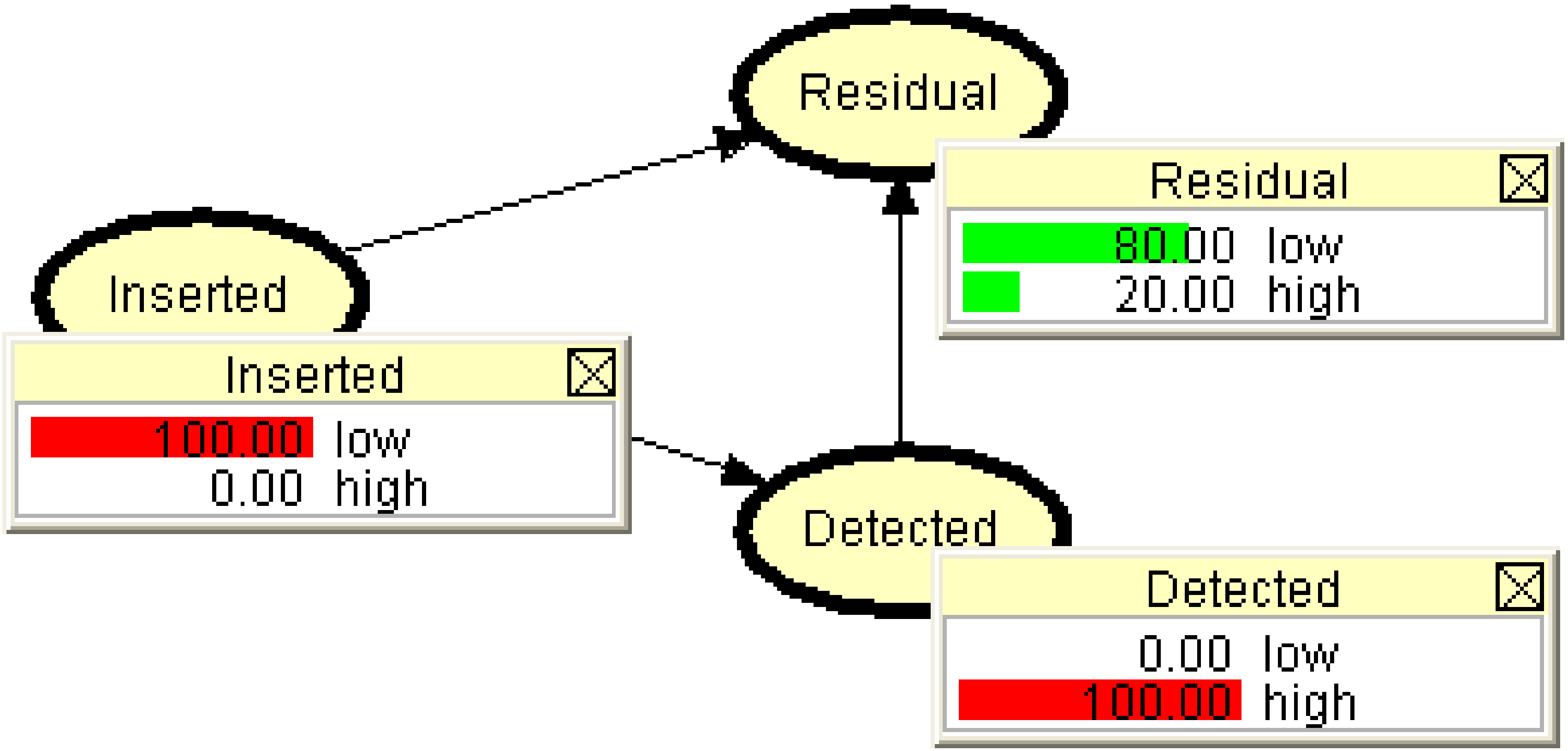
low	0.5
high	0.5

Detected

Edit Functions View

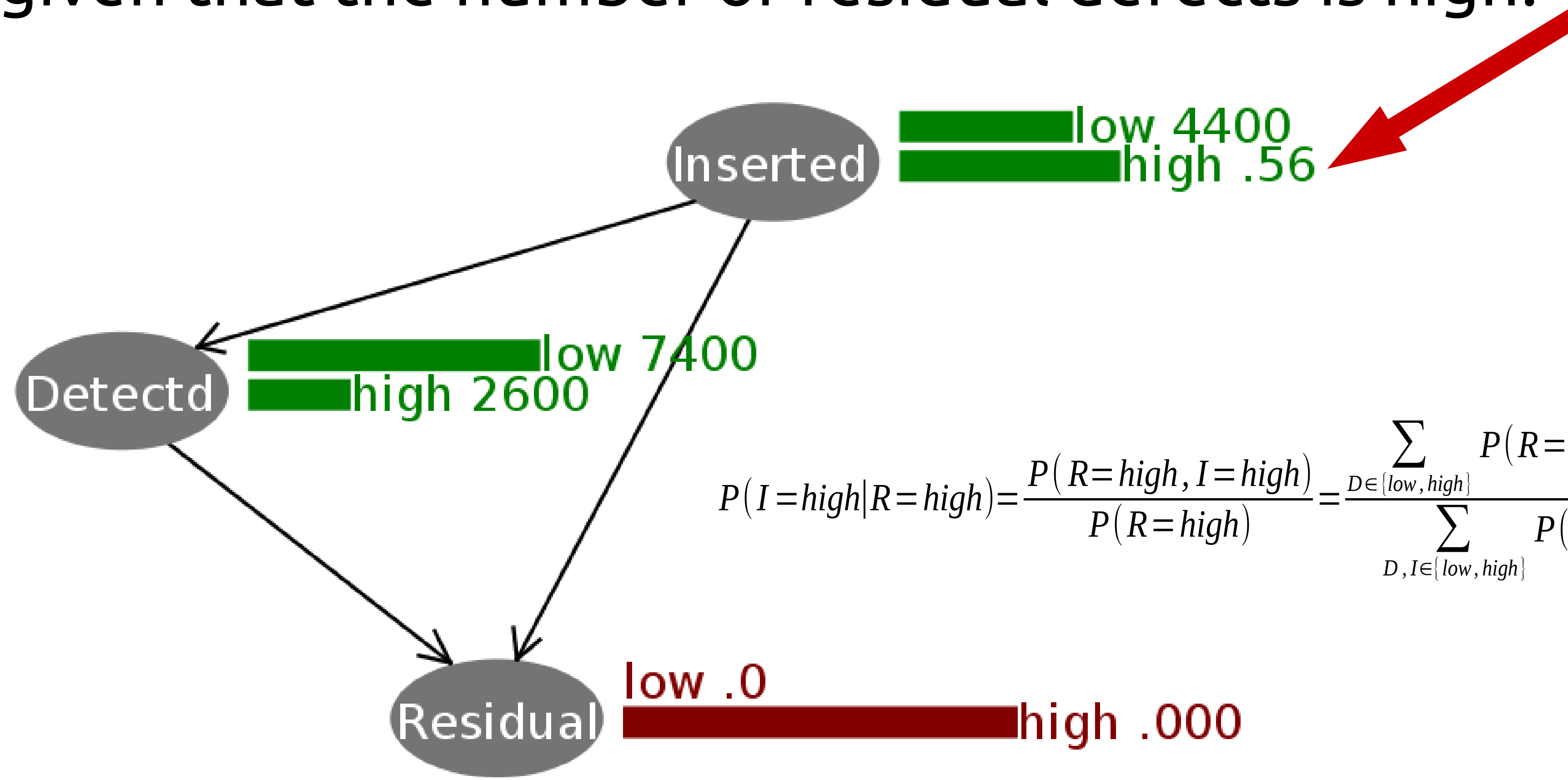
Inserted	low	high
low	0.6	0.4
high	0.4	0.6

- If we introduce "the evidences" we get the output probabilities (estimates)



Evidences (in red) are introduced in the network

- From symptoms we can get to the causes: “What is the probability that the number of inserted defects was high, given that the number of residual defects is high?“ → 56%



$$P(I=high|R=high)=\frac{P(R=high,I=high)}{P(R=high)}=\frac{\sum_{D\in\{low,high\}}P(R=high,D,I=high)}{\sum_{D,I\in\{low,high\}}P(R=high,D,I)}$$

If we fix the evidence that we had a high number of residual defects, we can deduce, by means of the conditional probability formula, the probability (high or low) of the defects inserted.

- For this small probabilistic network the computations can be done by hand, but they are tedious to perform.
- Computations performed manually (check the wikipedia entry 'Bayesian network' for the explanation of the procedure)
- It is impractical to compute the joint probability distribution when there are many variables

$$P(I=high \mid R=high) = \frac{P(R=high, I=high)}{P(R=high)} = \frac{\sum_{D \in \{low, high\}} P(R=high, D, I=high)}{\sum_{D, I \in \{low, high\}} P(R=high, D, I)}$$

Manual Computations for the Defects BN:

$$P(Residuals, Detected, Inserted) = P(R|D, I) \cdot P(D|I) \cdot P(I)$$

$$P(I=high|R=high) = \frac{P(R=high, I=high)}{P(R=high)} = \frac{\sum_{D \in \{low, high\}} P(R=high, D, I=high)}{\sum_{D, I \in \{low, high\}} P(R=high, D, I)} = \frac{0.09+0.19}{0.09+0.19+0.18+.04} = \frac{0.28}{0.5} = 0.56$$

$$P(R=high, D=high, I=high) = P(R=high|D=high, I=high)\,P(D=high|I=high)\,P(I=high) = 0.3 \times 0.6 \times 0.5 = 0.09$$

$$P(R=high, D=low, I=high) = P(R=high|D=low, I=high)\,P(D=low|I=high)\,P(I=high) = 0.95 \times 0.4 \times 0.5 = 0.19$$

$$P(R=high, D=low, I=low) = P(R=high|D=low, I=low)\,P(D=low|I=low)\,P(I=low) = 0.6 \times 0.6 \times 0.5 = 0.18$$

$$P(R=high, D=high, I=low) = P(R=high|D=high, I=low)\,P(D=high|I=low)\,P(I=low) = 0.2 \times 0.4 \times 0.5 = 0.04$$

- Bayesian networks exploit the graphical properties of the DAG (d-separation) in order to decrease the amount of computations by providing message passing algorithms.

Inference

- Bayesian networks allow several types of inference:
  - **Probabilistic inference**: given the *evidence* variables compute the *posterior* distribution of other variables.
  - **Parameter learning**: in order to specify a BN we need to specify conditional distributions that include parameters which are unknown and must be estimated from data.
  - **Structure learning**: BNs are specified by experts in the field and then are used to perform inference, in their simplest form. But in other situations the definition of the network must be learned from data.

Literature Review

- Aim: identify, classify and analyse the available literature since 2010 related to different aspects of software testing and quality that apply Bayesian concepts
- Currently, from 2010 onwards. In 2010 position paper by Namin and Sridharan related to the potential of Bayesian reasoning methods. Obstacles detected:
  - Generalization of the conclusions
  - Sensitivity to prior probabilities
  - Difficulties for software engineers to understand bayesian concepts
- We use the protocol for a Systematic Literature Review (EBSE website and Kitchenham recommendations.)

- Data Sources: ISI Web of Science, Scopus, Elsevier Science Direct, IEEE Xplore, SpringerLink, ACM Digital Library, Wiley Interscience, Google Scholar and The Collection of Computer Science Bibliographies.
- Keyword search: *Bayesian* & *networks* & *software testing*. Other combinations of keywords did not generate additional results.

Table 1. Repositories and papers found and selected

Digital Library	*Domain*	*Returned*	*Relevant*
ISI Web of Science	CS, Eng	10	6
Scopus	CS	45	16
Elsevier ScienceDirect	CS	40	6
SpringerLink	-	133	10
IEEEExplore	-	24	3
ACM DL	-	12	-
Wiley Interscience	-	62	-
Google Scholar	-	2,390	-
Total			41

- Software Testing Effort Prediction and Productivity Estimates
  - 4 references
  - This topic is concerned with the estimation of the test costs in terms of person-days. Few works have recently applied Bayesian models for testing effort estimation
- Fault and Defect Prediction. Software Reliability
  - 23 references
  - The topic of reliability is another area where Bayesian approaches have been explored by multiple researchers, specially for real-time systems.
- Quality Models
  - 2 references
  - A quality model describes in a structured way the concept of quality in a software system.

- Test Data Generation, Test Case Selection and Test Plan Generation
  - 7 references
  - Test data generation and test case priorization are important areas within software testing.
- Graphical User Interface (GUI) Testing
  - 2 references
  - Two works have built a BN that uses the prior knowledge of testers and the BN updates the values depending on the results of the test cases.
- Philosophy of Technology
  - 1 reference
  - Bayes concepts and the software testing field have been used as the substrate for defining the software engineering area as a “scientifically attested technology”

- 60% of the references lie on the "software reliability" area. The next areas of applications are "test data generation" and "test effort estimation", with 11% and 10% of the references, respectively .
- we may highlight the following issues after reviewing the literature:
  - Generalization of the conclusions: every work builds its BN starting from scratch and the BN is adapted to its specific problem. A "meta study" or meta-analysis of the results obtained by different researchers would uncover potential similarities in the results and in the graphical structure of the BN.
  - Sensitivity to priors: an essential characteristic of BNs is the need to provide prior probabilities to variables. One way to avoid discrepancies is to set standard priors in the field, which could be agreed upon in case of parameters such as productivity, etc. The fact that BNs allow us to update the variable probabilities can moderate the results obtained with different priors, provided a robust BN.

- Probabilistic graphical models can help in testing activities (and decision making in general) as supervised (prediction) and unsupervised (clustering) techniques from the data mining point of view as well as optimisation approaches.
- In prediction, we can consider classifiers such as Naïve Bayes and more complex structures such as TAN (Tree Augmented Naïve Bayes) to generic networks such as Bayesian Networks or Markov Models and their extensions (e.g. Dynamic BNs, Influence Diagrams). These latter Bayesian approches have not yet been fully exploited (in comparison with the former simpler Bayesian classifiers).

Acknowledgements

PROJECTS

“Testing of data persistence and user perspective under new paradigms“

“Gamificación y prototipado de procesos para la detección temprana de oportunidades en la producción del software“

PRESI TIN2013-46928-C3-1-R, TIN2013-46928-C3-2-R

Ministerio de Economía y Competitividad, Spain