*Orsolya Putz, Zoltán Varjú*

# Text Analytics with Python for Cognitive and Social Scientists

# *Contents*

# *Foreword*

I am delighted to introduce the first book on Multimedia Data Mining. When
I came to know about this book project undertaken by two of the most active
young researchers in the field, I was pleased that this book is coming in early
stage of a field that will need it more than most fields do. In most emerging
research fields, a book can play a significant role in bringing some maturity to
the field. Research fields advance through research papers. In research papers,
however, only a limited perspective could be provided about the field, its
application potential, and the techniques required and already developed in
the field. A book gives such a chance. I liked the idea that there will be a book
that will try to unify the field by bringing in disparate topics already available
in several papers that are not easy to find and understand. I was supportive
of this book project even before I had seen any material on it. The project
was a brilliant and a bold idea by two active researchers. Now that I have it
on my screen, it appears to be even a better idea.

Multimedia started gaining recognition in 1990s as a field. Processing,
storage, communication, and capture and display technologies had advanced
enough that researchers and technologists started building approaches to com-
bine information in multiple types of signals such as audio, images, video, and
text. Multimedia computing and communication techniques recognize corre-
lated information in multiple sources as well as insufficiency of information in
any individual source. By properly selecting sources to provide complemen-
tary information, such systems aspire, much like human perception system,
to create a holistic picture of a situation using only partial information from
separate sources.

Data mining is a direct outgrowth of progress in data storage and process-
ing speeds. When it became possible to store large volume of data and run
different statistical computations to explore all possible and even unlikely cor-
relations among data, the field of data mining was born. Data mining allowed
people to hypothesize relationships among data entities and explore support
for those. This field has been put to applications in many diverse domains and
keeps getting more applications. In fact many new fields are direct outgrowth
of data mining and it is likely to become a powerful computational tool.

# *Preface*

Approximately 17 million people in the USA (6% of the population) and 140 million people worldwide (this number is expected to rise to almost 300 million by the year 2025) suffer from *diabetes mellitus*. Currently, there a few dozens of commercialised devices for detecting blood glucose levels [1]. However, most of them are invasive. The development of a noninvasive method would considerably improve the quality of life for diabetic patients, facilitate their compliance for glucose monitoring, and reduce complications and mortality associated with this disease. Noninvasive and continuous monitoring of glucose concentration in blood and tissues is one of the most challenging and exciting applications of optics in medicine. The major difficulty in development and clinical application of optical noninvasive blood glucose sensors is associated with very low signal produced by glucose molecules. This results in low sensitivity and specificity of glucose monitoring by optical methods and needs a lot of efforts to overcome this difficulty.

A wide range of optical technologies have been designed in attempts to develop robust noninvasive methods for glucose sensing. The methods include infrared absorption, near-infrared scattering, Raman, fluorescent, and thermal gradient spectroscopies, as well as polarimetric, polarization heterodyning, photonic crystal, optoacoustic, optothermal, and optical coherence tomography (OCT) techniques [1-31].

For example, the polarimetric quantification of glucose is based on the phenomenon of optical rotatory dispersion, whereby a chiral molecule in an aqueous solution rotates the plane of linearly polarized light passing through the solution. The angle of rotation depends linearly on the concentration of the chiral species, the pathlength through the sample, and the molecule specific rotation. However, polarization sensitive optical technique makes it difficult to measure *in vivo* glucose concentration in blood through the skin because of the strong light scattering which causes light depolarization. For this reason, the anterior chamber of the eye has been suggested as a sight well suited for polarimetric measurements, since scattering in the eye is generally very low compared to that in other tissues, and a high correlation exists between the glucose in the blood and in the aqueous humor. The high accuracy of anterior eye chamber measurements is also due to the low concentration of optically active aqueous proteins within the aqueous humor.

On the other hand, the concept of noninvasive blood glucose sensing using the scattering properties of blood and tissues as an alternative to spectral absorption and polarization methods for monitoring of physiological glucose

concentrations in diabetic patients has been under intensive discussion for the last decade. Many of the considered effects, such as changing of the size, refractive index, packing, and aggregation of RBC under glucose variation, are important for glucose monitoring in diabetic patients. Indeed, at physiological concentrations of glucose, ranging from 40 to 400 mg/dl, the role of some of the effects may be modified, and some other effects, such as glucose penetration inside the RBC and the followed hemoglobin glycation, may be important [30-32].

Noninvasive determination of glucose was attempted using light scattering of skin tissue components measured by a spatially-resolved diffuse reflectance or NIR frequency-domain reflectance techniques. Both approaches are based on change in glucose concentration, which affects the refractive index mismatch between the interstitial fluid and tissue fibers, and hence reduces scattering coefficient. A glucose clamp experiment showed that reduced scattering coefficient measured in the visible range qualitatively tracked changes in blood glucose concentration for the volunteer with diabetes studied.

# *List of Figures*

# *List of Tables*

# *Symbols*

## Symbol Description

$\alpha$     To solve the generator maintenance scheduling, in the past, several mathematical techniques have been applied.

$\sigma^2$     These include integer programming, integer linear programming, dynamic programming, branch and bound etc.

$\sum$     Several heuristic search algorithms have also been developed. In recent years expert systems,

$abc$     fuzzy approaches, simulated annealing and genetic algorithms have also been tested.

$\theta\sqrt{abc}$     This paper presents a survey of the literature

$\zeta$     over the past fifteen years in the generator

$\partial$     maintenance scheduling. The objective is to

sdf     present a clear picture of the available recent literature

ewq     of the problem, the constraints and the other aspects of

bvcn     the generator maintenance schedule.

# Part I

# Basic Concepts

# 1

## *Text Analytics*

**CONTENTS**

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

## 1.1 Introduction

The term reliability usually refers to the probability that a component or system will operate satisfactorily either at any particular instant at which it is required or for a certain length of time. Fundamental to quantifying reliability s a knowledge of how to define, assess and combine probabilities [**?**]. This may hinge on identifying the form of the variability which is nherent n most processes. If all components had a fixed known lifetime there would be no need to model reliability.

1. A component part for an electronic item is manufactured at one of three different factories.

2. A component part $x$ for an electronic item is manufactured at one of three different factories.

3

    (a) A component part for an electronic item is manufactured at one of three different factories.

    (b) A component part $x$ for an electronic item is manufactured at one of three different factories.

       i. A component part for an electronic item is manufactured at one of three different factories.
      ii. A component part $x$ for an electronic item is manufactured at one of three different factories.
     iii. A component part for an electronic item is manufactured at one of three different factories.
     iv. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.
      v. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

    (c) A component part for an electronic item is manufactured at one of three different factories.

    (d) A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

    (e) A component part for enumerate list of an electronic item is manufactured at one of three different factories.

3. A component part for an electronic item is manufactured at one of three different factories.

4. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

5. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

### 1.1.1 A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty [**?**]? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main

**TABLE 1.1**

Now we are engaged $\left(a_g^a\right)$ $\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi[a] | 46, 281 | 6, 898 | 5, 400 |
| Keybrd[b] | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

"A Process is a structured, measured set of activities designed to produce a specific output for a particular customer or market— A process is thus a specific ordering of work activities across time and space, with a beginning, an end. and clearly defined inputs and outputs: a structure for action."

Thomas Davenport
Senior Adjutant to the Junior Marketing VP

**MultiRelational $k$-Anonymity.** Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables. This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k-1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables.

This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

In most literature on PPDP, they [?] consider a more relaxed, yet more practical, notion of privacy protection by assuming limited attacker's background knowledge. Below, the term "victim" refers to the record owner being linked. We can broadly classify linking models to two families.

A component part for an electronic item is [?] manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively.

One family considers a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call them *record linkage*, *attribute linkage*, and *table linkage*, respectively. In all types of linkages, we assume that the attacker knows the $QID$ of the victim. In record and attribute linkages, we further assume that the attacker knows the presence of the victim's record in the released table, and seeks to identify the victim's record and/or sensitive information from the table [?]. In table linkage, the attack seeks to determine the present or absent of the victim's record in the released table. A data table is considered to privacy preserved

**TABLE 1.2**
Now we are engaged $\left(a_g^a\right)$ $\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Table Head | | | |
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi | 46, 281 | 6, 898 | 5, 400 |
| Keybrd | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

if the table can effectively prevent the attacker from successfully performing these types of linkages on the table [**?**]. Sections 12.1-12.2 study this family of privacy models.

$$\text{var}\widehat{\Delta} = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \text{var}\,(\hat{\alpha}_j - \hat{\alpha}_k) = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \sigma^2(1/n_j + 1/n_k). \qquad (1.1)$$

An obvious measure of imbalance is just the difference in the number of times the two treatments are allocated

$$D_n = \mathcal{M}|n_A - n_B|. \qquad (1.2)$$

For rules such as deterministic allocation, for which the expected value of this difference can be calculated, we obtain the population value $\mathcal{D}_n$.

---

**Box Title Here**

Another family aims at achieving the *uninformative principle*: The published table should provide the attacker with little additional information beyond the background knowledge. There should not be a large difference between the prior and posterior beliefs; otherwise, there is a privacy threat [**?, ?**]. Many privacy models in this family are designed for statistical database and do not distinguish attributes in $T$ into $QID$, but some of them could also thwart record, attribute, and table linkages. Section 12.1 studies this family of privacy models.

Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.

---

**Theorem 1** *Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.*

**Proof 1** *Most of the proof of this theorem is routine. It is clear that $0 \in Z_m$ and $1 \in Z_m$ are the zero element and identity element. If $a \in Z_m$ and $a \neq 0$, then $m - a$ is the additive inverse of $a$. If $a \in Z_m$ and $a \neq 0$, then the greatest common divisor of $a$ and $m$ is 1, and hence there exist integers $s$ and $t$ such that $sa + tm = 1$. Thus $sa = 1 - tm$ is congruent to 1 modulo $m$. Let $s^*$ be the integer in $Z_m$ congruent to $s$ modulo $m$. Then we also have $s^*a \equiv 1 \mod m$. Hence $s^*$ is the multiplicative inverse of $a$ modulo $m$. Verification of the rest of the field properties is now routine.*

## 1.2 Record Linkage Model

In the privacy attack of *record linkage*, some value *qid* on $QID$ identifies a small number of records in the released table $T$, called a *group*. If the victim's
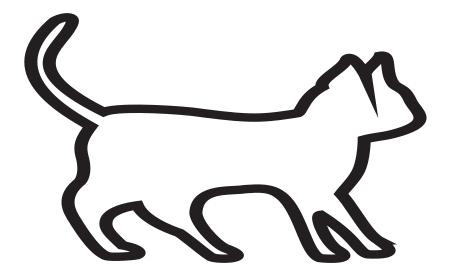
**FIGURE 1.1**
Figure caption goes here.

$QID$ matches the value $qid$, the victim is vulnerable to being linked to the small number of records in the group [**?**]. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group.

### 1.2.1 A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

#### 1.2.1.1 H3 A component part

A component part for an electronic item is manufactured at one of three [**?**] different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A
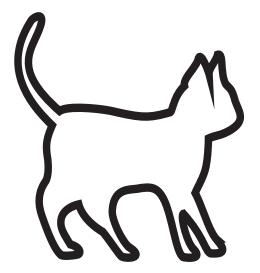
**FIGURE 1.2**
Figure caption goes here. Figure caption goes here. Figure caption goes here.
Figure caption goes here. Figure caption goes here. Figure caption goes here.

component is picked at random from the assembly line. What is the probability
that it is faulty?

A fundamental notion [**?**] is that of a subspace of $F^n$. Let $V$ be a nonempty
subset of $F^n$. Then $V$ is a *subspace* of $F^n$ provided $V$ is closed under vector
addition and scalar multiplication, that is,

(a)  For all $u$ and $v$ in $V$, $u + v$ is also in $V$.

(b)  For all $u$ in $V$ and $c$ in $F$, $cu$ is in $V$.

Let $u$ be in the subspace $V$. Because $0u = 0$, it follows that the zero vector
is in $V$. Similarly, $-u$ is in $V$ for all $u$ in $V$. A simple example of a subspace
of $F^n$ is the set of all vectors $(0, a_2, \ldots, a_n)$ with first coordinate equal to 0.
The zero vector itself is a subspace.

**Definition 1** Let $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ be vectors in $F^n$, and let $c_1, c_2, \ldots, c_m$
be scalars. Then the vector

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)}$$

is called a *linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. If $V$ is a subspace of
$F^n$, then $V$ is closed under vector addition and scalar multiplication, and
it follows easily by induction that a linear combination of vectors in $V$ is
also a vector in $V$. Thus *subspaces are closed under linear combinations*; in
fact, this can be taken as the defining property of subspaces. The vectors

(a)



(b)
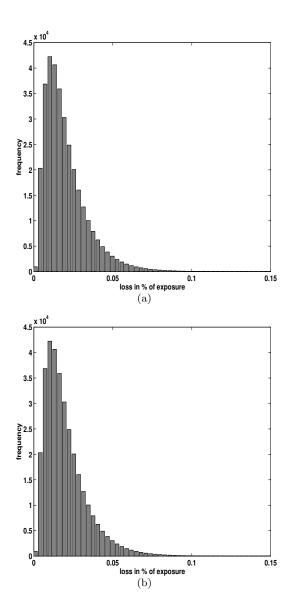
**FIGURE 1.3**
The bar charts depict the different risk contributions (top: 99% quantile, bottom: 99.9% quantile) of the business areas of a bank. The black bars are based on a Var/Covar approach, the white ones correspond to shortfall risk.

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ *span* $V$ (equivalently, form a *spanning set* of $V$) provided every vector in $V$ is a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. The zero vector can be written as a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ with all scalars equal to 0; this is a *trivial linear combination*. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are *linearly dependent* provided there are scalars $c_1, c_2, \ldots, c_m$, not all of which are zero, such that

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)} = 0,$$

that is, the zero vector can be written as a *nontrivial linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. For example, the vectors $(1, 4), (3, -1)$, and $(3, 5)$ in $\Re^2$ are linearly dependent since

$$3(1, 4) + 1(3, -2) - 2(3, 5) = (0, 0).$$

Vectors are *linearly independent* provided they are not linearly dependent. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are a *basis* of $V$ provided they are linearly independent and span $V$. By an *ordered basis* we mean a basis in which the vectors of the basis are listed in a specified order; to indicate that we have an ordered basis we write $(u^{(1)}, u^{(2)}, \ldots, u^{(m)})$. A spanning set $S$ of $V$ is a *minimal spanning set of $V$* provided that each set of vectors obtained from $S$ by removing a vector is not a spanning set for $V$. A linearly independent set $S$ of vectors of $V$ is a *maximal linearly independent set of vectors of $V$* provided that for each vector $w$ of $V$ that is not in $S$, $S \cup \{w\}$ is linearly dependent (when this happens, $w$ must be a linear combination of the vectors in $S$).  □

In addition to matrix addition, subtraction, and multiplication, there is one additional operation that we define now. It's perhaps the simplest of them all. Let $A = [a_{ij}]$ be an $m$ by $n$ matrix and let $c$ be a number [**?**]. Then the matrix $c \cdot A$, or simply $cA$, is the $m$ by $n$ matrix obtained by multiplying each entry of $A$ by $c$:

$$cA = [ca_{ij}].$$

The matrix $cA$ is called a *scalar multiple* of $A$.

---

## *Think About It...*

Commonly thought of as the first modern computer, ENTAC was built in 1944. It took up more space than an 18-wheeler's tractor trailer and weighed more than 17 Chevrolet Camaros. It consumed 140,000 watts of electricity while executing up to 5,000 basic arithmetic operations per second. One of today's popular microprocessors, the 486, is built on a tiny piece of silicon about the size of a dime.

With the continual expansion of capabilities, computing power

will eventually exceed the capacity for human comprehension or human control.

The Information Revolution
*Business Week*

## 1.3   Glossary

**360 Degree Review:** Performance review that includes feedback from superiors, peers, subordinates, and clients.

**Abnormal Variation:** Changes in process performance that cannot be accounted for by typical day-to-day variation. Also referred to as non-random variation.

**Acceptable Quality Level (AQL):** The minimum number of parts that must comply with quality standards, usually stated as a percentage.

**Activity:** The tasks performed to change inputs into outputs.

**Adaptable:** An adaptable process is designed to maintain effectiveness and efficiency as requirements change. The process is deemed adaptable when there is agreement among suppliers, owners, and customers that the process will meet requirements throughout the strategic period.

# 2

## Counting words and Zipf's Law

**CONTENTS**

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

## 2.1 Introduction

The term reliability usually refers to the probability that a component or system will operate satisfactorily either at any particular instant at which it is required or for a certain length of time. Fundamental to quantifying reliability s a knowledge of how to define, assess and combine probabilities [**?**]. This may hinge on identifying the form of the variability which is nherent n most processes. If all components had a fixed known lifetime there would be no need to model reliability.

1. A component part for an electronic item is manufactured at one of three different factories.

2. A component part $x$ for an electronic item is manufactured at one of three different factories.

(a) A component part for an electronic item is manufactured at one of three different factories.

(b) A component part $x$ for an electronic item is manufactured at one of three different factories.

    i. A component part for an electronic item is manufactured at one of three different factories.

    ii. A component part $x$ for an electronic item is manufactured at one of three different factories.

    iii. A component part for an electronic item is manufactured at one of three different factories.

    iv. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

    v. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

(c) A component part for an electronic item is manufactured at one of three different factories.

(d) A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

(e) A component part for enumerate list of an electronic item is manufactured at one of three different factories.

3. A component part for an electronic item is manufactured at one of three different factories.

4. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

5. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

### 2.1.1   A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty [**?**]? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main

**TABLE 2.1**
Now we are engaged $(a_g^a)$ $(a_g^a)$ in a great civil
war, testing whether that nation, or any
nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi[a] | 46, 281 | 6, 898 | 5, 400 |
| Keybrd[b] | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

"A Process is a structured, measured set of activities designed to produce a specific output for a particular customer or market— A process is thus a specific ordering of work activities across time and space, with a beginning, an end. and clearly defined inputs and outputs: a structure for action."

Thomas Davenport
Senior Adjutant to the Junior Marketing VP

**MultiRelational $k$-Anonymity.** Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables. This has proposed a privacy model called *MultiR $k$-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k-1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X,Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables.

This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \le i \le n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

In most literature on PPDP, they [?] consider a more relaxed, yet more practical, notion of privacy protection by assuming limited attacker's background knowledge. Below, the term "victim" refers to the record owner being linked. We can broadly classify linking models to two families.

A component part for an electronic item is [?] manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively.

One family considers a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call them *record linkage*, *attribute linkage*, and *table linkage*, respectively. In all types of linkages, we assume that the attacker knows the $QID$ of the victim. In record and attribute linkages, we further assume that the attacker knows the presence of the victim's record in the released table, and seeks to identify the victim's record and/or sensitive information from the table [?]. In table linkage, the attack seeks to determine the present or absent of the victim's record in the released table. A data table is considered to privacy preserved

**TABLE 2.2**
Now we are engaged $\left(a_g^a\right) \left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Table Head | | | |
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi | 46, 281 | 6, 898 | 5, 400 |
| Keybrd | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

if the table can effectively prevent the attacker from successfully performing these types of linkages on the table [**?**]. Sections 12.1-12.2 study this family of privacy models.

$$\text{var}\widehat{\Delta} = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \text{var}\,(\hat{\alpha}_j - \hat{\alpha}_k) = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \sigma^2(1/n_j + 1/n_k). \qquad (2.1)$$

An obvious measure of imbalance is just the difference in the number of times the two treatments are allocated

$$D_n = \mathcal{M}|n_A - n_B|. \qquad (2.2)$$

For rules such as deterministic allocation, for which the expected value of this difference can be calculated, we obtain the population value $\mathcal{D}_n$.

---

**Box Title Here**

Another family aims at achieving the *uninformative principle*: The published table should provide the attacker with little additional information beyond the background knowledge. There should not be a large difference between the prior and posterior beliefs; otherwise, there is a privacy threat [**?**, **?**]. Many privacy models in this family are designed for statistical database and do not distinguish attributes in $T$ into $QID$, but some of them could also thwart record, attribute, and table linkages. Section 12.1 studies this family of privacy models.

Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.

---

**Theorem 2** *Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.*

**Proof 2** *Most of the proof of this theorem is routine. It is clear that $0 \in Z_m$ and $1 \in Z_m$ are the zero element and identity element. If $a \in Z_m$ and $a \neq 0$, then $m - a$ is the additive inverse of $a$. If $a \in Z_m$ and $a \neq 0$, then the greatest common divisor of $a$ and $m$ is 1, and hence there exist integers $s$ and $t$ such that $sa + tm = 1$. Thus $sa = 1 - tm$ is congruent to 1 modulo $m$. Let $s^*$ be the integer in $Z_m$ congruent to $s$ modulo $m$. Then we also have $s^*a \equiv 1 \mod m$. Hence $s^*$ is the multiplicative inverse of $a$ modulo $m$. Verification of the rest of the field properties is now routine.*

## 2.2 Record Linkage Model

In the privacy attack of *record linkage*, some value $qid$ on $QID$ identifies a small number of records in the released table $T$, called a *group*. If the victim's

**FIGURE 2.1**
Figure caption goes here.

$QID$ matches the value $qid$, the victim is vulnerable to being linked to the small number of records in the group [**?**]. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group.

### 2.2.1    A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

#### 2.2.1.1    H3 A component part

A component part for an electronic item is manufactured at one of three [**?**] different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A
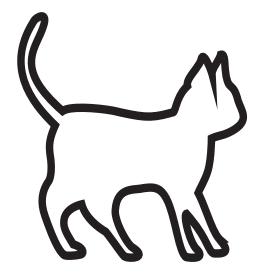
**FIGURE 2.2**
Figure caption goes here. Figure caption goes here. Figure caption goes here.
Figure caption goes here. Figure caption goes here. Figure caption goes here.

component is picked at random from the assembly line. What is the probability
that it is faulty?

A fundamental notion [**?**] is that of a subspace of $F^n$. Let $V$ be a nonempty
subset of $F^n$. Then $V$ is a *subspace* of $F^n$ provided $V$ is closed under vector
addition and scalar multiplication, that is,

(a)   For all $u$ and $v$ in $V$, $u + v$ is also in $V$.

(b)   For all $u$ in $V$ and $c$ in $F$, $cu$ is in $V$.

Let $u$ be in the subspace $V$. Because $0u = 0$, it follows that the zero vector
is in $V$. Similarly, $-u$ is in $V$ for all $u$ in $V$. A simple example of a subspace
of $F^n$ is the set of all vectors $(0, a_2, \ldots, a_n)$ with first coordinate equal to 0.
The zero vector itself is a subspace.

**Definition 2** Let $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ be vectors in $F^n$, and let $c_1, c_2, \ldots, c_m$
be scalars. Then the vector

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)}$$

is called a *linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. If $V$ is a subspace of
$F^n$, then $V$ is closed under vector addition and scalar multiplication, and
it follows easily by induction that a linear combination of vectors in $V$ is
also a vector in $V$. Thus *subspaces are closed under linear combinations*; in
fact, this can be taken as the defining property of subspaces. The vectors

(a)



(b)
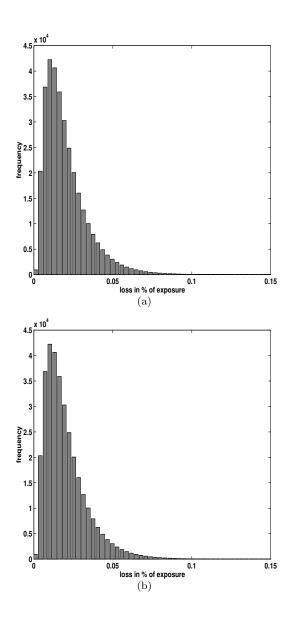
**FIGURE 2.3**
The bar charts depict the different risk contributions (top: 99% quantile, bottom: 99.9% quantile) of the business areas of a bank. The black bars are based on a Var/Covar approach, the white ones correspond to shortfall risk.

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ *span* $V$ (equivalently, form a *spanning set* of $V$) provided every vector in $V$ is a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. The zero vector can be written as a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ with all scalars equal to 0; this is a *trivial linear combination*. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are *linearly dependent* provided there are scalars $c_1, c_2, \ldots, c_m$, not all of which are zero, such that

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)} = 0,$$

that is, the zero vector can be written as a *nontrivial linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. For example, the vectors $(1, 4), (3, -1)$, and $(3, 5)$ in $\Re^2$ are linearly dependent since

$$3(1, 4) + 1(3, -2) - 2(3, 5) = (0, 0).$$

Vectors are *linearly independent* provided they are not linearly dependent. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are a *basis* of $V$ provided they are linearly independent and span $V$. By an *ordered basis* we mean a basis in which the vectors of the basis are listed in a specified order; to indicate that we have an ordered basis we write $(u^{(1)}, u^{(2)}, \ldots, u^{(m)})$. A spanning set $S$ of $V$ is a *minimal spanning set of $V$* provided that each set of vectors obtained from $S$ by removing a vector is not a spanning set for $V$. A linearly independent set $S$ of vectors of $V$ is a *maximal linearly independent set of vectors of $V$* provided that for each vector $w$ of $V$ that is not in $S$, $S \cup \{w\}$ is linearly dependent (when this happens, $w$ must be a linear combination of the vectors in $S$). $\square$

In addition to matrix addition, subtraction, and multiplication, there is one additional operation that we define now. It's perhaps the simplest of them all. Let $A = [a_{ij}]$ be an $m$ by $n$ matrix and let $c$ be a number [**?**]. Then the matrix $c \cdot A$, or simply $cA$, is the $m$ by $n$ matrix obtained by multiplying each entry of $A$ by $c$:

$$cA = [ca_{ij}].$$

The matrix $cA$ is called a *scalar multiple* of $A$.

---

## *Think About It...*

Commonly thought of as the first modern computer, ENTAC was built in 1944. It took up more space than an 18-wheeler's tractor trailer and weighed more than 17 Chevrolet Camaros. It consumed 140,000 watts of electricity while executing up to 5,000 basic arithmetic operations per second. One of today's popular microprocessors, the 486, is built on a tiny piece of silicon about the size of a dime.

With the continual expansion of capabilities, computing power

will eventually exceed the capacity for human comprehension or human control.

<div align="right">

The Information Revolution
*Business Week*

</div>

## 2.3   Glossary

**360 Degree Review:** Performance review that includes feedback from superiors, peers, subordinates, and clients.

**Abnormal Variation:** Changes in process performance that cannot be accounted for by typical day-to-day variation. Also referred to as non-random variation.

**Acceptable Quality Level (AQL):** The minimum number of parts that must comply with quality standards, usually stated as a percentage.

**Activity:** The tasks performed to change inputs into outputs.

**Adaptable:** An adaptable process is designed to maintain effectiveness and efficiency as requirements change. The process is deemed adaptable when there is agreement among suppliers, owners, and customers that the process will meet requirements throughout the strategic period.

# 3

## *Concordances, Collocations and Keywords*

**CONTENTS**

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

## 3.1 Introduction

The term reliability usually refers to the probability that a component or system will operate satisfactorily either at any particular instant at which it is required or for a certain length of time. Fundamental to quantifying reliability s a knowledge of how to define, assess and combine probabilities [**?**]. This may hinge on identifying the form of the variability which is nherent n most processes. If all components had a fixed known lifetime there would be no need to model reliability.

1. A component part for an electronic item is manufactured at one of three different factories.

2. A component part $x$ for an electronic item is manufactured at one of three different factories.

(a) A component part for an electronic item is manufactured at one of three different factories.

(b) A component part $x$ for an electronic item is manufactured at one of three different factories.

    i. A component part for an electronic item is manufactured at one of three different factories.

    ii. A component part $x$ for an electronic item is manufactured at one of three different factories.

    iii. A component part for an electronic item is manufactured at one of three different factories.

    iv. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

    v. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

(c) A component part for an electronic item is manufactured at one of three different factories.

(d) A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

(e) A component part for enumerate list of an electronic item is manufactured at one of three different factories.

3. A component part for an electronic item is manufactured at one of three different factories.

4. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

5. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

### 3.1.1   A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty [**?**]? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main

**TABLE 3.1**

Now we are engaged $\left(a_g^a\right)$ $\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi[a] | 46, 281 | 6, 898 | 5, 400 |
| Keybrd[b] | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

"A Process is a structured, measured set of activities designed to produce a specific output for a particular customer or market— A process is thus a specific ordering of work activities across time and space, with a beginning, an end. and clearly defined inputs and outputs: a structure for action."

Thomas Davenport
Senior Adjutant to the Junior Marketing VP

**MultiRelational $k$-Anonymity.** Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables. This has proposed a privacy model called *MultiR $k$-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables.

This has proposed a privacy model called *MultiR k-anonymity* to ensure *k*-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the *k*-anonymization is applied at the *record owner* level, not at the *record* level in traditional *k*-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

In most literature on PPDP, they [**?**] consider a more relaxed, yet more practical, notion of privacy protection by assuming limited attacker's background knowledge. Below, the term "victim" refers to the record owner being linked. We can broadly classify linking models to two families.

A component part for an electronic item is [**?**] manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively.

One family considers a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call them *record linkage*, *attribute linkage*, and *table linkage*, respectively. In all types of linkages, we assume that the attacker knows the $QID$ of the victim. In record and attribute linkages, we further assume that the attacker knows the presence of the victim's record in the released table, and seeks to identify the victim's record and/or sensitive information from the table [**?**]. In table linkage, the attack seeks to determine the present or absent of the victim's record in the released table. A data table is considered to privacy preserved

**TABLE 3.2**

Now we are engaged $\left(a_g^a\right)\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Table Head | | | |
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi | 46, 281 | 6, 898 | 5, 400 |
| Keybrd | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

if the table can effectively prevent the attacker from successfully performing these types of linkages on the table [**?**]. Sections 12.1-12.2 study this family of privacy models.

$$\text{var}\widehat{\Delta} = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \text{var}\,(\hat{\alpha}_j - \hat{\alpha}_k) = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \sigma^2(1/n_j + 1/n_k). \tag{3.1}$$

An obvious measure of imbalance is just the difference in the number of times the two treatments are allocated

$$D_n = \mathcal{M}|n_A - n_B|. \tag{3.2}$$

For rules such as deterministic allocation, for which the expected value of this difference can be calculated, we obtain the population value $\mathcal{D}_n$.

---

**Box Title Here**

Another family aims at achieving the *uninformative principle*: The published table should provide the attacker with little additional information beyond the background knowledge. There should not be a large difference between the prior and posterior beliefs; otherwise, there is a privacy threat [**?**, **?**]. Many privacy models in this family are designed for statistical database and do not distinguish attributes in $T$ into $QID$, but some of them could also thwart record, attribute, and table linkages. Section 12.1 studies this family of privacy models.

Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.

---

**Theorem 3** *Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.*

**Proof 3** *Most of the proof of this theorem is routine. It is clear that $0 \in Z_m$ and $1 \in Z_m$ are the zero element and identity element. If $a \in Z_m$ and $a \neq 0$, then $m - a$ is the additive inverse of $a$. If $a \in Z_m$ and $a \neq 0$, then the greatest common divisor of $a$ and $m$ is 1, and hence there exist integers $s$ and $t$ such that $sa + tm = 1$. Thus $sa = 1 - tm$ is congruent to 1 modulo $m$. Let $s^*$ be the integer in $Z_m$ congruent to $s$ modulo $m$. Then we also have $s^*a \equiv 1 \mod m$. Hence $s^*$ is the multiplicative inverse of $a$ modulo $m$. Verification of the rest of the field properties is now routine.*

## 3.2   Record Linkage Model

In the privacy attack of *record linkage*, some value *qid* on $QID$ identifies a small number of records in the released table $T$, called a *group*. If the victim's

**FIGURE 3.1**
Figure caption goes here.

*QID* matches the value *qid*, the victim is vulnerable to being linked to the small number of records in the group [**?**]. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group.

### 3.2.1   A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

#### 3.2.1.1   H3 A component part

A component part for an electronic item is manufactured at one of three [**?**] different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A

**FIGURE 3.2**
Figure caption goes here. Figure caption goes here. Figure caption goes here.
Figure caption goes here. Figure caption goes here. Figure caption goes here.

component is picked at random from the assembly line. What is the probability
that it is faulty?

A fundamental notion [**?**] is that of a subspace of $F^n$. Let $V$ be a nonempty
subset of $F^n$. Then $V$ is a *subspace* of $F^n$ provided $V$ is closed under vector
addition and scalar multiplication, that is,

(a)   For all $u$ and $v$ in $V$, $u + v$ is also in $V$.

(b)   For all $u$ in $V$ and $c$ in $F$, $cu$ is in $V$.

Let $u$ be in the subspace $V$. Because $0u = 0$, it follows that the zero vector
is in $V$. Similarly, $-u$ is in $V$ for all $u$ in $V$. A simple example of a subspace
of $F^n$ is the set of all vectors $(0, a_2, \ldots, a_n)$ with first coordinate equal to 0.
The zero vector itself is a subspace.

**Definition 3** Let $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ be vectors in $F^n$, and let $c_1, c_2, \ldots, c_m$
be scalars. Then the vector

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)}$$

is called a *linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. If $V$ is a subspace of
$F^n$, then $V$ is closed under vector addition and scalar multiplication, and
it follows easily by induction that a linear combination of vectors in $V$ is
also a vector in $V$. Thus *subspaces are closed under linear combinations*; in
fact, this can be taken as the defining property of subspaces. The vectors

**FIGURE 3.3**
The bar charts depict the different risk contributions (top: 99% quantile, bottom: 99.9% quantile) of the business areas of a bank. The black bars are based on a Var/Covar approach, the white ones correspond to shortfall risk.

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ *span* $V$ (equivalently, form a *spanning set* of $V$) provided every vector in $V$ is a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. The zero vector can be written as a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ with all scalars equal to 0; this is a *trivial linear combination*. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are *linearly dependent* provided there are scalars $c_1, c_2, \ldots, c_m$, not all of which are zero, such that

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)} = 0,$$

that is, the zero vector can be written as a *nontrivial linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. For example, the vectors $(1, 4), (3, -1)$, and $(3, 5)$ in $\Re^2$ are linearly dependent since

$$3(1, 4) + 1(3, -2) - 2(3, 5) = (0, 0).$$

Vectors are *linearly independent* provided they are not linearly dependent. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are a *basis* of $V$ provided they are linearly independent and span $V$. By an *ordered basis* we mean a basis in which the vectors of the basis are listed in a specified order; to indicate that we have an ordered basis we write $(u^{(1)}, u^{(2)}, \ldots, u^{(m)})$. A spanning set $S$ of $V$ is a *minimal spanning set of $V$* provided that each set of vectors obtained from $S$ by removing a vector is not a spanning set for $V$. A linearly independent set $S$ of vectors of $V$ is a *maximal linearly independent set of vectors of $V$* provided that for each vector $w$ of $V$ that is not in $S$, $S \cup \{w\}$ is linearly dependent (when this happens, $w$ must be a linear combination of the vectors in $S$). $\square$

In addition to matrix addition, subtraction, and multiplication, there is one additional operation that we define now. It's perhaps the simplest of them all. Let $A = [a_{ij}]$ be an $m$ by $n$ matrix and let $c$ be a number [**?**]. Then the matrix $c \cdot A$, or simply $cA$, is the $m$ by $n$ matrix obtained by multiplying each entry of $A$ by $c$:

$$cA = [ca_{ij}].$$

The matrix $cA$ is called a *scalar multiple* of $A$.

---

## *Think About It...*

Commonly thought of as the first modern computer, ENTAC was built in 1944. It took up more space than an 18-wheeler's tractor trailer and weighed more than 17 Chevrolet Camaros. It consumed 140,000 watts of electricity while executing up to 5,000 basic arithmetic operations per second. One of today's popular microprocessors, the 486, is built on a tiny piece of silicon about the size of a dime.

With the continual expansion of capabilities, computing power

will eventually exceed the capacity for human comprehension or human control.

The Information Revolution
*Business Week*

## 3.3  Glossary

**360 Degree Review:** Performance review that includes feedback from superiors, peers, subordinates, and clients.

**Abnormal Variation:** Changes in process performance that cannot be accounted for by typical day-to-day variation. Also referred to as non-random variation.

**Acceptable Quality Level (AQL):** The minimum number of parts that must comply with quality standards, usually stated as a percentage.

**Activity:** The tasks performed to change inputs into outputs.

**Adaptable:** An adaptable process is designed to maintain effectiveness and efficiency as requirements change. The process is deemed adaptable when there is agreement among suppliers, owners, and customers that the process will meet requirements throughout the strategic period.

# 4

## Sentiment and Emotion Analysis

**CONTENTS**

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

## 4.1 Introduction

The term reliability usually refers to the probability that a component or system will operate satisfactorily either at any particular instant at which it is required or for a certain length of time. Fundamental to quantifying reliability s a knowledge of how to define, assess and combine probabilities [**?**]. This may hinge on identifying the form of the variability which is nherent n most processes. If all components had a fixed known lifetime there would be no need to model reliability.

1. A component part for an electronic item is manufactured at one of three different factories.

2. A component part $x$ for an electronic item is manufactured at one of three different factories.

    (a) A component part for an electronic item is manufactured at one of three different factories.

    (b) A component part $x$ for an electronic item is manufactured at one of three different factories.

       i. A component part for an electronic item is manufactured at one of three different factories.

      ii. A component part $x$ for an electronic item is manufactured at one of three different factories.

     iii. A component part for an electronic item is manufactured at one of three different factories.

     iv. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

      v. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

    (c) A component part for an electronic item is manufactured at one of three different factories.

    (d) A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

    (e) A component part for enumerate list of an electronic item is manufactured at one of three different factories.

3. A component part for an electronic item is manufactured at one of three different factories.

4. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

5. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

### 4.1.1 A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty [**?**]? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main

**TABLE 4.1**

Now we are engaged $\left(a_g^a\right)$ $\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi[a] | 46, 281 | 6, 898 | 5, 400 |
| Keybrd[b] | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

"A Process is a structured, measured set of activities designed to produce a specific output for a particular customer or market— A process is thus a specific ordering of work activities across time and space, with a beginning, an end. and clearly defined inputs and outputs: a structure for action."

Thomas Davenport
Senior Adjutant to the Junior Marketing VP

**MultiRelational $k$-Anonymity.** Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables. This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables.

This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \le i \le n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

In most literature on PPDP, they [?] consider a more relaxed, yet more practical, notion of privacy protection by assuming limited attacker's background knowledge. Below, the term "victim" refers to the record owner being linked. We can broadly classify linking models to two families.

> A component part for an electronic item is [?] manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively.

One family considers a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call them *record linkage*, *attribute linkage*, and *table linkage*, respectively. In all types of linkages, we assume that the attacker knows the $QID$ of the victim. In record and attribute linkages, we further assume that the attacker knows the presence of the victim's record in the released table, and seeks to identify the victim's record and/or sensitive information from the table [?]. In table linkage, the attack seeks to determine the present or absent of the victim's record in the released table. A data table is considered to privacy preserved

**TABLE 4.2**
Now we are engaged $\left(a_g^a\right)\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Table Head | | | |
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi | 46, 281 | 6, 898 | 5, 400 |
| Keybrd | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

if the table can effectively prevent the attacker from successfully performing these types of linkages on the table [**?**]. Sections 12.1-12.2 study this family of privacy models.

$$\text{var}\widehat{\Delta} = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \text{var}\,(\hat{\alpha}_j - \hat{\alpha}_k) = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \sigma^2(1/n_j + 1/n_k). \qquad (4.1)$$

An obvious measure of imbalance is just the difference in the number of times the two treatments are allocated

$$D_n = \mathcal{M}|n_A - n_B|. \qquad (4.2)$$

For rules such as deterministic allocation, for which the expected value of this difference can be calculated, we obtain the population value $\mathcal{D}_n$.

---

**Box Title Here**

Another family aims at achieving the *uninformative principle*: The published table should provide the attacker with little additional information beyond the background knowledge. There should not be a large difference between the prior and posterior beliefs; otherwise, there is a privacy threat [**?**, **?**]. Many privacy models in this family are designed for statistical database and do not distinguish attributes in $T$ into $QID$, but some of them could also thwart record, attribute, and table linkages. Section 12.1 studies this family of privacy models.

Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.

---

**Theorem 4** *Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.*

**Proof 4** *Most of the proof of this theorem is routine. It is clear that $0 \in Z_m$ and $1 \in Z_m$ are the zero element and identity element. If $a \in Z_m$ and $a \neq 0$, then $m - a$ is the additive inverse of $a$. If $a \in Z_m$ and $a \neq 0$, then the greatest common divisor of $a$ and $m$ is 1, and hence there exist integers $s$ and $t$ such that $sa + tm = 1$. Thus $sa = 1 - tm$ is congruent to 1 modulo $m$. Let $s^*$ be the integer in $Z_m$ congruent to $s$ modulo $m$. Then we also have $s^*a \equiv 1 \mod m$. Hence $s^*$ is the multiplicative inverse of $a$ modulo $m$. Verification of the rest of the field properties is now routine.*

## 4.2   Record Linkage Model

In the privacy attack of *record linkage*, some value $qid$ on $QID$ identifies a small number of records in the released table $T$, called a *group*. If the victim's

**FIGURE 4.1**
Figure caption goes here.

*QID* matches the value *qid*, the victim is vulnerable to being linked to the small number of records in the group [**?**]. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group.

### 4.2.1 A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

#### 4.2.1.1 H3 A component part

A component part for an electronic item is manufactured at one of three [**?**] different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A

**FIGURE 4.2**
Figure caption goes here. Figure caption goes here. Figure caption goes here. Figure caption goes here. Figure caption goes here. Figure caption goes here.

component is picked at random from the assembly line. What is the probability that it is faulty?

A fundamental notion [**?**] is that of a subspace of $F^n$. Let $V$ be a nonempty subset of $F^n$. Then $V$ is a *subspace* of $F^n$ provided $V$ is closed under vector addition and scalar multiplication, that is,

(a)   For all $u$ and $v$ in $V$, $u + v$ is also in $V$.

(b)   For all $u$ in $V$ and $c$ in $F$, $cu$ is in $V$.

Let $u$ be in the subspace $V$. Because $0u = 0$, it follows that the zero vector is in $V$. Similarly, $-u$ is in $V$ for all $u$ in $V$. A simple example of a subspace of $F^n$ is the set of all vectors $(0, a_2, \ldots, a_n)$ with first coordinate equal to 0. The zero vector itself is a subspace.

**Definition 4** Let $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ be vectors in $F^n$, and let $c_1, c_2, \ldots, c_m$ be scalars. Then the vector

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)}$$

is called a *linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. If $V$ is a subspace of $F^n$, then $V$ is closed under vector addition and scalar multiplication, and it follows easily by induction that a linear combination of vectors in $V$ is also a vector in $V$. Thus *subspaces are closed under linear combinations*; in fact, this can be taken as the defining property of subspaces. The vectors
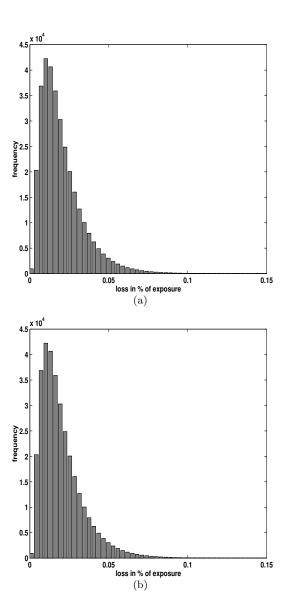
**FIGURE 4.3**

The bar charts depict the different risk contributions (top: 99% quantile, bottom: 99.9% quantile) of the business areas of a bank. The black bars are based on a Var/Covar approach, the white ones correspond to shortfall risk.

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ *span* $V$ (equivalently, form a *spanning set* of $V$) provided every vector in $V$ is a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. The zero vector can be written as a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ with all scalars equal to 0; this is a *trivial linear combination.* The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are *linearly dependent* provided there are scalars $c_1, c_2, \ldots, c_m$, not all of which are zero, such that

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)} = 0,$$

that is, the zero vector can be written as a *nontrivial linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. For example, the vectors $(1, 4), (3, -1)$, and $(3, 5)$ in $\Re^2$ are linearly dependent since

$$3(1, 4) + 1(3, -2) - 2(3, 5) = (0, 0).$$

Vectors are *linearly independent* provided they are not linearly dependent. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are a *basis* of $V$ provided they are linearly independent and span $V$. By an *ordered basis* we mean a basis in which the vectors of the basis are listed in a specified order; to indicate that we have an ordered basis we write $(u^{(1)}, u^{(2)}, \ldots, u^{(m)})$. A spanning set $S$ of $V$ is a *minimal spanning set of $V$* provided that each set of vectors obtained from $S$ by removing a vector is not a spanning set for $V$. A linearly independent set $S$ of vectors of $V$ is a *maximal linearly independent set of vectors of $V$* provided that for each vector $w$ of $V$ that is not in $S$, $S \cup \{w\}$ is linearly dependent (when this happens, $w$ must be a linear combination of the vectors in $S$). $\square$

In addition to matrix addition, subtraction, and multiplication, there is one additional operation that we define now. It's perhaps the simplest of them all. Let $A = [a_{ij}]$ be an $m$ by $n$ matrix and let $c$ be a number [**?**]. Then the matrix $c \cdot A$, or simply $cA$, is the $m$ by $n$ matrix obtained by multiplying each entry of $A$ by $c$:

$$cA = [ca_{ij}].$$

The matrix $cA$ is called a *scalar multiple* of $A$.

---

### *Think About It...*

Commonly thought of as the first modern computer, ENTAC was built in 1944. It took up more space than an 18-wheeler's tractor trailer and weighed more than 17 Chevrolet Camaros. It consumed 140,000 watts of electricity while executing up to 5,000 basic arithmetic operations per second. One of today's popular microprocessors, the 486, is built on a tiny piece of silicon about the size of a dime.

With the continual expansion of capabilities, computing power

will eventually exceed the capacity for human comprehension or human control.

The Information Revolution
*Business Week*

## 4.3   Glossary

**360 Degree Review:** Performance review that includes feedback from superiors, peers, subordinates, and clients.

**Abnormal Variation:** Changes in process performance that cannot be accounted for by typical day-to-day variation. Also referred to as non-random variation.

**Acceptable Quality Level (AQL):** The minimum number of parts that must comply with quality standards, usually stated as a percentage.

**Activity:** The tasks performed to change inputs into outputs.

**Adaptable:** An adaptable process is designed to maintain effectiveness and efficiency as requirements change. The process is deemed adaptable when there is agreement among suppliers, owners, and customers that the process will meet requirements throughout the strategic period.

# 5

## *Personality Traits*

**CONTENTS**

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

## 5.1 Introduction

The term reliability usually refers to the probability that a component or system will operate satisfactorily either at any particular instant at which it is required or for a certain length of time. Fundamental to quantifying reliability s a knowledge of how to define, assess and combine probabilities [**?**]. This may hinge on identifying the form of the variability which is nherent n most processes. If all components had a fixed known lifetime there would be no need to model reliability.

1. A component part for an electronic item is manufactured at one of three different factories.

2. A component part $x$ for an electronic item is manufactured at one of three different factories.

    (a) A component part for an electronic item is manufactured at one of three different factories.

    (b) A component part $x$ for an electronic item is manufactured at one of three different factories.

        i. A component part for an electronic item is manufactured at one of three different factories.
        ii. A component part $x$ for an electronic item is manufactured at one of three different factories.
        iii. A component part for an electronic item is manufactured at one of three different factories.
        iv. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.
        v. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

    (c) A component part for an electronic item is manufactured at one of three different factories.

    (d) A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

    (e) A component part for enumerate list of an electronic item is manufactured at one of three different factories.

3. A component part for an electronic item is manufactured at one of three different factories.

4. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

5. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

### 5.1.1   A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty [**?**]? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main

**TABLE 5.1**

Now we are engaged $\left(a_g^a\right)$ $\left(a_g^a\right)$ in a great civil
war, testing whether that nation, or any
nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi[a] | 46, 281 | 6, 898 | 5, 400 |
| Keybrd[b] | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

"A Process is a structured, measured set of activities designed to produce a specific output for a particular customer or market— A process is thus a specific ordering of work activities across time and space, with a beginning, an end. and clearly defined inputs and outputs: a structure for action."

Thomas Davenport
Senior Adjutant to the Junior Marketing VP

**MultiRelational $k$-Anonymity.** Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [?] database usually contains multiple relational tables. This has proposed a privacy model called *MultiR $k$-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k-1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [?] database usually contains multiple relational tables.

This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k-1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

In most literature on PPDP, they [?] consider a more relaxed, yet more practical, notion of privacy protection by assuming limited attacker's background knowledge. Below, the term "victim" refers to the record owner being linked. We can broadly classify linking models to two families.

A component part for an electronic item is [?] manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively.

One family considers a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call them *record linkage*, *attribute linkage*, and *table linkage*, respectively. In all types of linkages, we assume that the attacker knows the $QID$ of the victim. In record and attribute linkages, we further assume that the attacker knows the presence of the victim's record in the released table, and seeks to identify the victim's record and/or sensitive information from the table [?]. In table linkage, the attack seeks to determine the present or absent of the victim's record in the released table. A data table is considered to privacy preserved

**TABLE 5.2**
Now we are engaged $\left(a_g^a\right) \left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|-------|-----------|-----------|-----------|
| Table Head | | | |
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi | 46, 281 | 6, 898 | 5, 400 |
| Keybrd | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

if the table can effectively prevent the attacker from successfully performing these types of linkages on the table [**?**]. Sections 12.1-12.2 study this family of privacy models.

$$\mathrm{var}\widehat{\Delta} = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \mathrm{var}\,(\hat{\alpha}_j - \hat{\alpha}_k) = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \sigma^2(1/n_j + 1/n_k). \qquad (5.1)$$

An obvious measure of imbalance is just the difference in the number of times the two treatments are allocated

$$D_n = \mathcal{M}|n_A - n_B|. \qquad (5.2)$$

For rules such as deterministic allocation, for which the expected value of this difference can be calculated, we obtain the population value $\mathcal{D}_n$.

---

**Box Title Here**

Another family aims at achieving the *uninformative principle*: The published table should provide the attacker with little additional information beyond the background knowledge. There should not be a large difference between the prior and posterior beliefs; otherwise, there is a privacy threat [**?**, **?**]. Many privacy models in this family are designed for statistical database and do not distinguish attributes in $T$ into $QID$, but some of them could also thwart record, attribute, and table linkages. Section 12.1 studies this family of privacy models.

Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.

---

**Theorem 5** *Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.*

**Proof 5** *Most of the proof of this theorem is routine. It is clear that $0 \in Z_m$ and $1 \in Z_m$ are the zero element and identity element. If $a \in Z_m$ and $a \neq 0$, then $m - a$ is the additive inverse of $a$. If $a \in Z_m$ and $a \neq 0$, then the greatest common divisor of $a$ and $m$ is 1, and hence there exist integers $s$ and $t$ such that $sa + tm = 1$. Thus $sa = 1 - tm$ is congruent to 1 modulo $m$. Let $s^*$ be the integer in $Z_m$ congruent to $s$ modulo $m$. Then we also have $s^*a \equiv 1 \mod m$. Hence $s^*$ is the multiplicative inverse of $a$ modulo $m$. Verification of the rest of the field properties is now routine.*

## 5.2   Record Linkage Model

In the privacy attack of *record linkage*, some value *qid* on $QID$ identifies a small number of records in the released table $T$, called a *group*. If the victim's

**FIGURE 5.1**
Figure caption goes here.

*QID* matches the value *qid*, the victim is vulnerable to being linked to the small number of records in the group [**?**]. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group.

## 5.2.1   A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

### 5.2.1.1   H3 A component part

A component part for an electronic item is manufactured at one of three [**?**] different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A
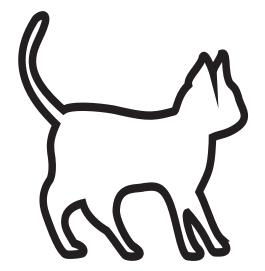
**FIGURE 5.2**
Figure caption goes here. Figure caption goes here. Figure caption goes here.
Figure caption goes here. Figure caption goes here. Figure caption goes here.

component is picked at random from the assembly line. What is the probability that it is faulty?

A fundamental notion [**?**] is that of a subspace of $F^n$. Let $V$ be a nonempty subset of $F^n$. Then $V$ is a *subspace* of $F^n$ provided $V$ is closed under vector addition and scalar multiplication, that is,

    (a)   For all $u$ and $v$ in $V$, $u + v$ is also in $V$.

    (b)   For all $u$ in $V$ and $c$ in $F$, $cu$ is in $V$.

Let $u$ be in the subspace $V$. Because $0u = 0$, it follows that the zero vector is in $V$. Similarly, $-u$ is in $V$ for all $u$ in $V$. A simple example of a subspace of $F^n$ is the set of all vectors $(0, a_2, \ldots, a_n)$ with first coordinate equal to 0. The zero vector itself is a subspace.

**Definition 5** Let $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ be vectors in $F^n$, and let $c_1, c_2, \ldots, c_m$ be scalars. Then the vector

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)}$$

is called a *linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. If $V$ is a subspace of $F^n$, then $V$ is closed under vector addition and scalar multiplication, and it follows easily by induction that a linear combination of vectors in $V$ is also a vector in $V$. Thus *subspaces are closed under linear combinations*; in fact, this can be taken as the defining property of subspaces. The vectors

**FIGURE 5.3**
The bar charts depict the different risk contributions (top: 99% quantile, bottom: 99.9% quantile) of the business areas of a bank. The black bars are based on a Var/Covar approach, the white ones correspond to shortfall risk.

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ *span* $V$ (equivalently, form a *spanning set* of $V$) provided every vector in $V$ is a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. The zero vector can be written as a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ with all scalars equal to 0; this is a *trivial linear combination*. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are *linearly dependent* provided there are scalars $c_1, c_2, \ldots, c_m$, not all of which are zero, such that

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)} = 0,$$

that is, the zero vector can be written as a *nontrivial linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. For example, the vectors $(1,4), (3,-1)$, and $(3,5)$ in $\Re^2$ are linearly dependent since

$$3(1,4) + 1(3,-2) - 2(3,5) = (0,0).$$

Vectors are *linearly independent* provided they are not linearly dependent. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are a *basis* of $V$ provided they are linearly independent and span $V$. By an *ordered basis* we mean a basis in which the vectors of the basis are listed in a specified order; to indicate that we have an ordered basis we write $(u^{(1)}, u^{(2)}, \ldots, u^{(m)})$. A spanning set $S$ of $V$ is a *minimal spanning set of $V$* provided that each set of vectors obtained from $S$ by removing a vector is not a spanning set for $V$. A linearly independent set $S$ of vectors of $V$ is a *maximal linearly independent set of vectors of $V$* provided that for each vector $w$ of $V$ that is not in $S$, $S \cup \{w\}$ is linearly dependent (when this happens, $w$ must be a linear combination of the vectors in $S$). $\square$

In addition to matrix addition, subtraction, and multiplication, there is one additional operation that we define now. It's perhaps the simplest of them all. Let $A = [a_{ij}]$ be an $m$ by $n$ matrix and let $c$ be a number [**?**]. Then the matrix $c \cdot A$, or simply $cA$, is the $m$ by $n$ matrix obtained by multiplying each entry of $A$ by $c$:

$$cA = [ca_{ij}].$$

The matrix $cA$ is called a *scalar multiple* of $A$.

---

## *Think About It...*

Commonly thought of as the first modern computer, ENTAC was built in 1944. It took up more space than an 18-wheeler's tractor trailer and weighed more than 17 Chevrolet Camaros. It consumed 140,000 watts of electricity while executing up to 5,000 basic arithmetic operations per second. One of today's popular microprocessors, the 486, is built on a tiny piece of silicon about the size of a dime.

With the continual expansion of capabilities, computing power

will eventually exceed the capacity for human comprehension or human control.

The Information Revolution
*Business Week*

## 5.3   Glossary

**360 Degree Review:** Performance review that includes feedback from superiors, peers, subordinates, and clients.

**Abnormal Variation:** Changes in process performance that cannot be accounted for by typical day-to-day variation. Also referred to as non-random variation.

**Acceptable Quality Level (AQL):** The minimum number of parts that must comply with quality standards, usually stated as a percentage.

**Activity:** The tasks performed to change inputs into outputs.

**Adaptable:** An adaptable process is designed to maintain effectiveness and efficiency as requirements change. The process is deemed adaptable when there is agreement among suppliers, owners, and customers that the process will meet requirements throughout the strategic period.

# Part II

# Graph-based methods

# 6

## Cognitive Modeling and Graphs

**CONTENTS**

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

## 6.1   Introduction

The term reliability usually refers to the probability that a component or system will operate satisfactorily either at any particular instant at which it is required or for a certain length of time. Fundamental to quantifying reliability s a knowledge of how to define, assess and combine probabilities [**?**]. This may hinge on identifying the form of the variability which is nherent n most processes. If all components had a fixed known lifetime there would be no need to model reliability.

1. A component part for an electronic item is manufactured at one of three different factories.

2. A component part $x$ for an electronic item is manufactured at one of three different factories.

(a) A component part for an electronic item is manufactured at one of three different factories.

(b) A component part $x$ for an electronic item is manufactured at one of three different factories.

    i. A component part for an electronic item is manufactured at one of three different factories.

    ii. A component part $x$ for an electronic item is manufactured at one of three different factories.

    iii. A component part for an electronic item is manufactured at one of three different factories.

    iv. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

    v. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

(c) A component part for an electronic item is manufactured at one of three different factories.

(d) A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

(e) A component part for enumerate list of an electronic item is manufactured at one of three different factories.

3. A component part for an electronic item is manufactured at one of three different factories.

4. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

5. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

### 6.1.1  A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty [**?**]? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main

**TABLE 6.1**

Now we are engaged $\left(a_g^a\right)$ $\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi[a] | 46, 281 | 6, 898 | 5, 400 |
| Keybrd[b] | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

"A Process is a structured, measured set of activities designed to produce a specific output for a particular customer or market— A process is thus a specific ordering of work activities across time and space, with a beginning, an end. and clearly defined inputs and outputs: a structure for action."

Thomas Davenport
Senior Adjutant to the Junior Marketing VP

**MultiRelational $k$-Anonymity.** Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables. This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k-1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables.

This has proposed a privacy model called *MultiR k-anonymity* to ensure *k*-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the *k*-anonymization is applied at the *record owner* level, not at the *record* level in traditional *k*-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

In most literature on PPDP, they [?] consider a more relaxed, yet more practical, notion of privacy protection by assuming limited attacker's background knowledge. Below, the term "victim" refers to the record owner being linked. We can broadly classify linking models to two families.

> A component part for an electronic item is [?] manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively.

One family considers a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call them *record linkage*, *attribute linkage*, and *table linkage*, respectively. In all types of linkages, we assume that the attacker knows the $QID$ of the victim. In record and attribute linkages, we further assume that the attacker knows the presence of the victim's record in the released table, and seeks to identify the victim's record and/or sensitive information from the table [?]. In table linkage, the attack seeks to determine the present or absent of the victim's record in the released table. A data table is considered to privacy preserved

**TABLE 6.2**

Now we are engaged $\left(a_g^a\right)\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|-------|-----------|-----------|-----------|
| Table Head | | | |
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi | 46, 281 | 6, 898 | 5, 400 |
| Keybrd | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

if the table can effectively prevent the attacker from successfully performing these types of linkages on the table [**?**]. Sections 12.1-12.2 study this family of privacy models.

$$\mathrm{var}\widehat{\Delta} = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \mathrm{var}\,(\hat{\alpha}_j - \hat{\alpha}_k) = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \sigma^2(1/n_j + 1/n_k). \qquad (6.1)$$

An obvious measure of imbalance is just the difference in the number of times the two treatments are allocated

$$D_n = \mathcal{M}|n_A - n_B|. \qquad (6.2)$$

For rules such as deterministic allocation, for which the expected value of this difference can be calculated, we obtain the population value $\mathcal{D}_n$.

---

**Box Title Here**

Another family aims at achieving the *uninformative principle*: The published table should provide the attacker with little additional information beyond the background knowledge. There should not be a large difference between the prior and posterior beliefs; otherwise, there is a privacy threat [**?**, **?**]. Many privacy models in this family are designed for statistical database and do not distinguish attributes in $T$ into $QID$, but some of them could also thwart record, attribute, and table linkages. Section 12.1 studies this family of privacy models.

Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.

---

**Theorem 6** *Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.*

**Proof 6** *Most of the proof of this theorem is routine. It is clear that $0 \in Z_m$ and $1 \in Z_m$ are the zero element and identity element. If $a \in Z_m$ and $a \neq 0$, then $m - a$ is the additive inverse of $a$. If $a \in Z_m$ and $a \neq 0$, then the greatest common divisor of $a$ and $m$ is 1, and hence there exist integers $s$ and $t$ such that $sa + tm = 1$. Thus $sa = 1 - tm$ is congruent to 1 modulo $m$. Let $s^*$ be the integer in $Z_m$ congruent to $s$ modulo $m$. Then we also have $s^*a \equiv 1 \mod m$. Hence $s^*$ is the multiplicative inverse of $a$ modulo $m$. Verification of the rest of the field properties is now routine.*

## 6.2   Record Linkage Model

In the privacy attack of *record linkage*, some value *qid* on $QID$ identifies a small number of records in the released table $T$, called a *group*. If the victim's
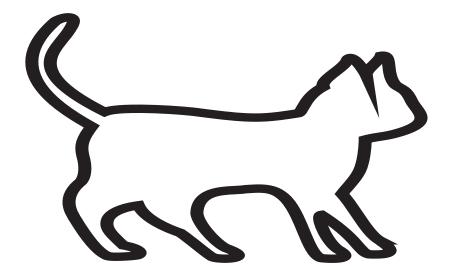
**FIGURE 6.1**
Figure caption goes here.

$QID$ matches the value *qid*, the victim is vulnerable to being linked to the small number of records in the group [**?**]. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group.

## 6.2.1    A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

### 6.2.1.1    H3 A component part

A component part for an electronic item is manufactured at one of three [**?**] different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A

**FIGURE 6.2**
Figure caption goes here. Figure caption goes here. Figure caption goes here.
Figure caption goes here. Figure caption goes here. Figure caption goes here.

component is picked at random from the assembly line. What is the probability that it is faulty?

A fundamental notion [**?**] is that of a subspace of $F^n$. Let $V$ be a nonempty subset of $F^n$. Then $V$ is a *subspace* of $F^n$ provided $V$ is closed under vector addition and scalar multiplication, that is,

(a)  For all $u$ and $v$ in $V$, $u + v$ is also in $V$.

(b)  For all $u$ in $V$ and $c$ in $F$, $cu$ is in $V$.

Let $u$ be in the subspace $V$. Because $0u = 0$, it follows that the zero vector is in $V$. Similarly, $-u$ is in $V$ for all $u$ in $V$. A simple example of a subspace of $F^n$ is the set of all vectors $(0, a_2, \ldots, a_n)$ with first coordinate equal to 0. The zero vector itself is a subspace.

**Definition 6** Let $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ be vectors in $F^n$, and let $c_1, c_2, \ldots, c_m$ be scalars. Then the vector

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)}$$

is called a *linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. If $V$ is a subspace of $F^n$, then $V$ is closed under vector addition and scalar multiplication, and it follows easily by induction that a linear combination of vectors in $V$ is also a vector in $V$. Thus *subspaces are closed under linear combinations*; in fact, this can be taken as the defining property of subspaces. The vectors

(a)

(b)
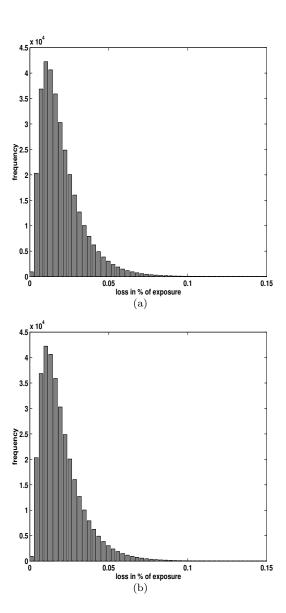
**FIGURE 6.3**
The bar charts depict the different risk contributions (top: 99% quantile, bottom: 99.9% quantile) of the business areas of a bank. The black bars are based on a Var/Covar approach, the white ones correspond to shortfall risk.

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ *span* $V$ (equivalently, form a *spanning set* of $V$) provided every vector in $V$ is a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. The zero vector can be written as a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ with all scalars equal to 0; this is a *trivial linear combination*. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are *linearly dependent* provided there are scalars $c_1, c_2, \ldots, c_m$, not all of which are zero, such that

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)} = 0,$$

that is, the zero vector can be written as a *nontrivial linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. For example, the vectors $(1, 4), (3, -1)$, and $(3, 5)$ in $\Re^2$ are linearly dependent since

$$3(1, 4) + 1(3, -2) - 2(3, 5) = (0, 0).$$

Vectors are *linearly independent* provided they are not linearly dependent. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are a *basis* of $V$ provided they are linearly independent and span $V$. By an *ordered basis* we mean a basis in which the vectors of the basis are listed in a specified order; to indicate that we have an ordered basis we write $(u^{(1)}, u^{(2)}, \ldots, u^{(m)})$. A spanning set $S$ of $V$ is a *minimal spanning set of $V$* provided that each set of vectors obtained from $S$ by removing a vector is not a spanning set for $V$. A linearly independent set $S$ of vectors of $V$ is a *maximal linearly independent set of vectors of $V$* provided that for each vector $w$ of $V$ that is not in $S$, $S \cup \{w\}$ is linearly dependent (when this happens, $w$ must be a linear combination of the vectors in $S$). □

In addition to matrix addition, subtraction, and multiplication, there is one additional operation that we define now. It's perhaps the simplest of them all. Let $A = [a_{ij}]$ be an $m$ by $n$ matrix and let $c$ be a number [**?**]. Then the matrix $c \cdot A$, or simply $cA$, is the $m$ by $n$ matrix obtained by multiplying each entry of $A$ by $c$:

$$cA = [ca_{ij}].$$

The matrix $cA$ is called a *scalar multiple* of $A$.

---

## *Think About It...*

Commonly thought of as the first modern computer, ENTAC was built in 1944. It took up more space than an 18-wheeler's tractor trailer and weighed more than 17 Chevrolet Camaros. It consumed 140,000 watts of electricity while executing up to 5,000 basic arithmetic operations per second. One of today's popular microprocessors, the 486, is built on a tiny piece of silicon about the size of a dime.

With the continual expansion of capabilities, computing power

will eventually exceed the capacity for human comprehension or human control.

The Information Revolution
*Business Week*

## 6.3   Glossary

**360 Degree Review:** Performance review that includes feedback from superiors, peers, subordinates, and clients.

**Abnormal Variation:** Changes in process performance that cannot be accounted for by typical day-to-day variation. Also referred to as non-random variation.

**Acceptable Quality Level (AQL):** The minimum number of parts that must comply with quality standards, usually stated as a percentage.

**Activity:** The tasks performed to change inputs into outputs.

**Adaptable:** An adaptable process is designed to maintain effectiveness and efficiency as requirements change. The process is deemed adaptable when there is agreement among suppliers, owners, and customers that the process will meet requirements throughout the strategic period.

# 7

## *Keyword Extraction and Text Summarization*

**CONTENTS**

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

## 7.1 Introduction

The term reliability usually refers to the probability that a component or system will operate satisfactorily either at any particular instant at which it is required or for a certain length of time. Fundamental to quantifying reliability s a knowledge of how to define, assess and combine probabilities [?]. This may hinge on identifying the form of the variability which is nherent n most processes. If all components had a fixed known lifetime there would be no need to model reliability.

1. A component part for an electronic item is manufactured at one of three different factories.

2. A component part $x$ for an electronic item is manufactured at one of three different factories.

(a) A component part for an electronic item is manufactured at one of three different factories.

(b) A component part $x$ for an electronic item is manufactured at one of three different factories.

    i. A component part for an electronic item is manufactured at one of three different factories.

    ii. A component part $x$ for an electronic item is manufactured at one of three different factories.

    iii. A component part for an electronic item is manufactured at one of three different factories.

    iv. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

    v. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

(c) A component part for an electronic item is manufactured at one of three different factories.

(d) A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

(e) A component part for enumerate list of an electronic item is manufactured at one of three different factories.

3. A component part for an electronic item is manufactured at one of three different factories.

4. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

5. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

### 7.1.1   A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty [**?**]? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main

**TABLE 7.1**

Now we are engaged $\left(a_g^a\right)$ $\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi[a] | 46, 281 | 6, 898 | 5, 400 |
| Keybrd[b] | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

> "A Process is a structured, measured set of activities designed to produce a specific output for a particular customer or market— A process is thus a specific ordering of work activities across time and space, with a beginning, an end. and clearly defined inputs and outputs: a structure for action."
>
> Thomas Davenport
> Senior Adjutant to the Junior Marketing VP

**MultiRelational $k$-Anonymity.** Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables. This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables.

This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

In most literature on PPDP, they [?] consider a more relaxed, yet more practical, notion of privacy protection by assuming limited attacker's background knowledge. Below, the term "victim" refers to the record owner being linked. We can broadly classify linking models to two families.

> A component part for an electronic item is [?] manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively.

One family considers a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call them *record linkage*, *attribute linkage*, and *table linkage*, respectively. In all types of linkages, we assume that the attacker knows the $QID$ of the victim. In record and attribute linkages, we further assume that the attacker knows the presence of the victim's record in the released table, and seeks to identify the victim's record and/or sensitive information from the table [?]. In table linkage, the attack seeks to determine the present or absent of the victim's record in the released table. A data table is considered to privacy preserved

**TABLE 7.2**

Now we are engaged $\left(a_g^a\right) \left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|-------|-----------|-----------|-----------|
| Table Head | | | |
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi | 46, 281 | 6, 898 | 5, 400 |
| Keybrd | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

if the table can effectively prevent the attacker from successfully performing these types of linkages on the table [**?**]. Sections 12.1-12.2 study this family of privacy models.

$$\mathrm{var}\widehat{\Delta} = \sum_{j=1}^{t}\sum_{k=j+1}^{t} \mathrm{var}\,(\hat{\alpha}_j - \hat{\alpha}_k) = \sum_{j=1}^{t}\sum_{k=j+1}^{t}\sigma^2(1/n_j + 1/n_k). \qquad (7.1)$$

An obvious measure of imbalance is just the difference in the number of times the two treatments are allocated

$$D_n = \mathcal{M}|n_A - n_B|. \qquad (7.2)$$

For rules such as deterministic allocation, for which the expected value of this difference can be calculated, we obtain the population value $\mathcal{D}_n$.

---

**Box Title Here**

Another family aims at achieving the *uninformative principle*: The published table should provide the attacker with little additional information beyond the background knowledge. There should not be a large difference between the prior and posterior beliefs; otherwise, there is a privacy threat [**?, ?**]. Many privacy models in this family are designed for statistical database and do not distinguish attributes in $T$ into $QID$, but some of them could also thwart record, attribute, and table linkages. Section 12.1 studies this family of privacy models.

Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.

---

**Theorem 7** *Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.*

**Proof 7** *Most of the proof of this theorem is routine. It is clear that $0 \in Z_m$ and $1 \in Z_m$ are the zero element and identity element. If $a \in Z_m$ and $a \neq 0$, then $m - a$ is the additive inverse of $a$. If $a \in Z_m$ and $a \neq 0$, then the greatest common divisor of $a$ and $m$ is 1, and hence there exist integers $s$ and $t$ such that $sa + tm = 1$. Thus $sa = 1 - tm$ is congruent to 1 modulo $m$. Let $s^*$ be the integer in $Z_m$ congruent to $s$ modulo $m$. Then we also have $s^*a \equiv 1 \mod m$. Hence $s^*$ is the multiplicative inverse of $a$ modulo $m$. Verification of the rest of the field properties is now routine.*

## 7.2 Record Linkage Model

In the privacy attack of *record linkage*, some value *qid* on $QID$ identifies a small number of records in the released table $T$, called a *group*. If the victim's

**FIGURE 7.1**
Figure caption goes here.

$QID$ matches the value $qid$, the victim is vulnerable to being linked to the small number of records in the group [**?**]. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group.

## 7.2.1　A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

### 7.2.1.1　H3 A component part

A component part for an electronic item is manufactured at one of three [**?**] different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A

**FIGURE 7.2**
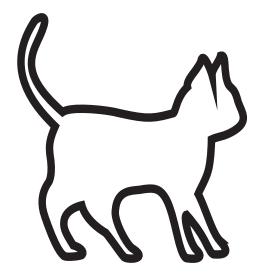Figure caption goes here. Figure caption goes here. Figure caption goes here.
Figure caption goes here. Figure caption goes here. Figure caption goes here.

component is picked at random from the assembly line. What is the probability
that it is faulty?

A fundamental notion [**?**] is that of a subspace of $F^n$. Let $V$ be a nonempty
subset of $F^n$. Then $V$ is a *subspace* of $F^n$ provided $V$ is closed under vector
addition and scalar multiplication, that is,

(a)   For all $u$ and $v$ in $V$, $u + v$ is also in $V$.

(b)   For all $u$ in $V$ and $c$ in $F$, $cu$ is in $V$.

Let $u$ be in the subspace $V$. Because $0u = 0$, it follows that the zero vector
is in $V$. Similarly, $-u$ is in $V$ for all $u$ in $V$. A simple example of a subspace
of $F^n$ is the set of all vectors $(0, a_2, \ldots, a_n)$ with first coordinate equal to 0.
The zero vector itself is a subspace.

**Definition 7** Let $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ be vectors in $F^n$, and let $c_1, c_2, \ldots, c_m$
be scalars. Then the vector

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)}$$

is called a *linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. If $V$ is a subspace of
$F^n$, then $V$ is closed under vector addition and scalar multiplication, and
it follows easily by induction that a linear combination of vectors in $V$ is
also a vector in $V$. Thus *subspaces are closed under linear combinations*; in
fact, this can be taken as the defining property of subspaces. The vectors
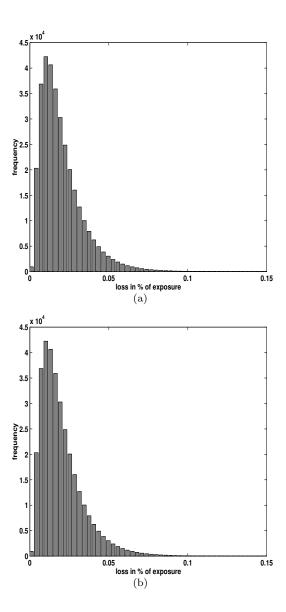
(a)



(b)

**FIGURE 7.3**
The bar charts depict the different risk contributions (top: 99% quantile, bottom: 99.9% quantile) of the business areas of a bank. The black bars are based on a Var/Covar approach, the white ones correspond to shortfall risk.

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ *span* $V$ (equivalently, form a *spanning set* of $V$) provided every vector in $V$ is a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. The zero vector can be written as a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ with all scalars equal to 0; this is a *trivial linear combination*. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are *linearly dependent* provided there are scalars $c_1, c_2, \ldots, c_m$, not all of which are zero, such that

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)} = 0,$$

that is, the zero vector can be written as a *nontrivial linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. For example, the vectors $(1, 4), (3, -1)$, and $(3, 5)$ in $\Re^2$ are linearly dependent since

$$3(1, 4) + 1(3, -2) - 2(3, 5) = (0, 0).$$

Vectors are *linearly independent* provided they are not linearly dependent. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are a *basis* of $V$ provided they are linearly independent and span $V$. By an *ordered basis* we mean a basis in which the vectors of the basis are listed in a specified order; to indicate that we have an ordered basis we write $(u^{(1)}, u^{(2)}, \ldots, u^{(m)})$. A spanning set $S$ of $V$ is a *minimal spanning set of $V$* provided that each set of vectors obtained from $S$ by removing a vector is not a spanning set for $V$. A linearly independent set $S$ of vectors of $V$ is a *maximal linearly independent set of vectors of $V$* provided that for each vector $w$ of $V$ that is not in $S$, $S \cup \{w\}$ is linearly dependent (when this happens, $w$ must be a linear combination of the vectors in $S$). $\square$

In addition to matrix addition, subtraction, and multiplication, there is one additional operation that we define now. It's perhaps the simplest of them all. Let $A = [a_{ij}]$ be an $m$ by $n$ matrix and let $c$ be a number [**?**]. Then the matrix $c \cdot A$, or simply $cA$, is the $m$ by $n$ matrix obtained by multiplying each entry of $A$ by $c$:

$$cA = [ca_{ij}].$$

The matrix $cA$ is called a *scalar multiple* of $A$.

---

## *Think About It...*

Commonly thought of as the first modern computer, ENTAC was built in 1944. It took up more space than an 18-wheeler's tractor trailer and weighed more than 17 Chevrolet Camaros. It consumed 140,000 watts of electricity while executing up to 5,000 basic arithmetic operations per second. One of today's popular microprocessors, the 486, is built on a tiny piece of silicon about the size of a dime.

With the continual expansion of capabilities, computing power

will eventually exceed the capacity for human comprehension or human control.

> The Information Revolution
> *Business Week*

## 7.3   Glossary

**360 Degree Review:** Performance review that includes feedback from superiors, peers, subordinates, and clients.

**Abnormal Variation:** Changes in process performance that cannot be accounted for by typical day-to-day variation. Also referred to as non-random variation.

**Acceptable Quality Level (AQL):** The minimum number of parts that must comply with quality standards, usually stated as a percentage.

**Activity:** The tasks performed to change inputs into outputs.

**Adaptable:** An adaptable process is designed to maintain effectiveness and efficiency as requirements change. The process is deemed adaptable when there is agreement among suppliers, owners, and customers that the process will meet requirements throughout the strategic period.

# Part III

# Embeddings

# 8

## Word Embeddings

**CONTENTS**

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

## 8.1   Introduction

The term reliability usually refers to the probability that a component or system will operate satisfactorily either at any particular instant at which it is required or for a certain length of time. Fundamental to quantifying reliability s a knowledge of how to define, assess and combine probabilities [**?**]. This may hinge on identifying the form of the variability which is nherent n most processes. If all components had a fixed known lifetime there would be no need to model reliability.

1. A component part for an electronic item is manufactured at one of three different factories.

2. A component part $x$ for an electronic item is manufactured at one of three different factories.

(a) A component part for an electronic item is manufactured at one of three different factories.

(b) A component part $x$ for an electronic item is manufactured at one of three different factories.

     i. A component part for an electronic item is manufactured at one of three different factories.

     ii. A component part $x$ for an electronic item is manufactured at one of three different factories.

     iii. A component part for an electronic item is manufactured at one of three different factories.

     iv. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

     v. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

(c) A component part for an electronic item is manufactured at one of three different factories.

(d) A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

(e) A component part for enumerate list of an electronic item is manufactured at one of three different factories.

3. A component part for an electronic item is manufactured at one of three different factories.

4. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

5. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

### 8.1.1    A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty [**?**]? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main

**TABLE 8.1**

Now we are engaged $\left(a_g^a\right)$ $\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi[a] | 46, 281 | 6, 898 | 5, 400 |
| Keybrd[b] | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

"A Process is a structured, measured set of activities designed to produce a specific output for a particular customer or market— A process is thus a specific ordering of work activities across time and space, with a beginning, an end. and clearly defined inputs and outputs: a structure for action."

Thomas Davenport
Senior Adjutant to the Junior Marketing VP

**MultiRelational $k$-Anonymity.** Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables. This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables.

This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

In most literature on PPDP, they [?] consider a more relaxed, yet more practical, notion of privacy protection by assuming limited attacker's background knowledge. Below, the term "victim" refers to the record owner being linked. We can broadly classify linking models to two families.

A component part for an electronic item is [?] manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively.

One family considers a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call them *record linkage*, *attribute linkage*, and *table linkage*, respectively. In all types of linkages, we assume that the attacker knows the $QID$ of the victim. In record and attribute linkages, we further assume that the attacker knows the presence of the victim's record in the released table, and seeks to identify the victim's record and/or sensitive information from the table [?]. In table linkage, the attack seeks to determine the present or absent of the victim's record in the released table. A data table is considered to privacy preserved

**TABLE 8.2**

Now we are engaged $\left(a_g^a\right) \left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|-------|-----------|-----------|-----------|
| Table Head | | | |
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi | 46, 281 | 6, 898 | 5, 400 |
| Keybrd | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

if the table can effectively prevent the attacker from successfully performing these types of linkages on the table [**?**]. Sections 12.1-12.2 study this family of privacy models.

$$\text{var}\widehat{\Delta} = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \text{var}\,(\hat{\alpha}_j - \hat{\alpha}_k) = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \sigma^2(1/n_j + 1/n_k). \qquad (8.1)$$

An obvious measure of imbalance is just the difference in the number of times the two treatments are allocated

$$D_n = \mathcal{M}|n_A - n_B|. \qquad (8.2)$$

For rules such as deterministic allocation, for which the expected value of this difference can be calculated, we obtain the population value $\mathcal{D}_n$.

---

**Box Title Here**

Another family aims at achieving the *uninformative principle*: The published table should provide the attacker with little additional information beyond the background knowledge. There should not be a large difference between the prior and posterior beliefs; otherwise, there is a privacy threat [**?, ?**]. Many privacy models in this family are designed for statistical database and do not distinguish attributes in $T$ into $QID$, but some of them could also thwart record, attribute, and table linkages. Section 12.1 studies this family of privacy models.

Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.

---

**Theorem 8** *Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.*

**Proof 8** *Most of the proof of this theorem is routine. It is clear that $0 \in Z_m$ and $1 \in Z_m$ are the zero element and identity element. If $a \in Z_m$ and $a \neq 0$, then $m - a$ is the additive inverse of $a$. If $a \in Z_m$ and $a \neq 0$, then the greatest common divisor of $a$ and $m$ is 1, and hence there exist integers $s$ and $t$ such that $sa + tm = 1$. Thus $sa = 1 - tm$ is congruent to 1 modulo $m$. Let $s^*$ be the integer in $Z_m$ congruent to $s$ modulo $m$. Then we also have $s^*a \equiv 1 \mod m$. Hence $s^*$ is the multiplicative inverse of $a$ modulo $m$. Verification of the rest of the field properties is now routine.*

## 8.2 Record Linkage Model

In the privacy attack of *record linkage*, some value $qid$ on $QID$ identifies a small number of records in the released table $T$, called a *group*. If the victim's
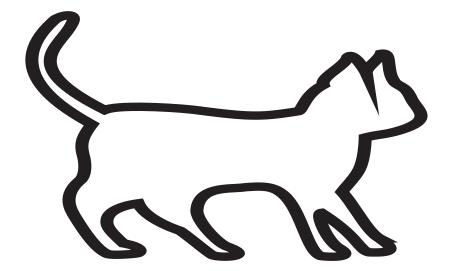
**FIGURE 8.1**
Figure caption goes here.

$QID$ matches the value $qid$, the victim is vulnerable to being linked to the small number of records in the group [**?**]. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group.

### 8.2.1  A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

#### 8.2.1.1  H3 A component part

A component part for an electronic item is manufactured at one of three [**?**] different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A
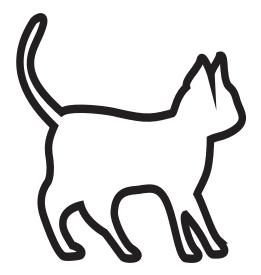
**FIGURE 8.2**
Figure caption goes here. Figure caption goes here. Figure caption goes here.
Figure caption goes here. Figure caption goes here. Figure caption goes here.

component is picked at random from the assembly line. What is the probability
that it is faulty?

A fundamental notion [**?**] is that of a subspace of $F^n$. Let $V$ be a nonempty
subset of $F^n$. Then $V$ is a *subspace* of $F^n$ provided $V$ is closed under vector
addition and scalar multiplication, that is,

(a)   For all $u$ and $v$ in $V$, $u + v$ is also in $V$.

(b)   For all $u$ in $V$ and $c$ in $F$, $cu$ is in $V$.

Let $u$ be in the subspace $V$. Because $0u = 0$, it follows that the zero vector
is in $V$. Similarly, $-u$ is in $V$ for all $u$ in $V$. A simple example of a subspace
of $F^n$ is the set of all vectors $(0, a_2, \ldots, a_n)$ with first coordinate equal to 0.
The zero vector itself is a subspace.

**Definition 8** Let $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ be vectors in $F^n$, and let $c_1, c_2, \ldots, c_m$
be scalars. Then the vector

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)}$$

is called a *linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. If $V$ is a subspace of
$F^n$, then $V$ is closed under vector addition and scalar multiplication, and
it follows easily by induction that a linear combination of vectors in $V$ is
also a vector in $V$. Thus *subspaces are closed under linear combinations*; in
fact, this can be taken as the defining property of subspaces. The vectors
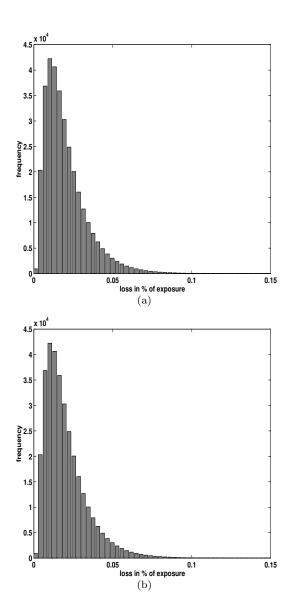
**FIGURE 8.3**
The bar charts depict the different risk contributions (top: 99% quantile, bottom: 99.9% quantile) of the business areas of a bank. The black bars are based on a Var/Covar approach, the white ones correspond to shortfall risk.

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ *span* $V$ (equivalently, form a *spanning set* of $V$) provided every vector in $V$ is a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. The zero vector can be written as a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ with all scalars equal to 0; this is a *trivial linear combination*. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are *linearly dependent* provided there are scalars $c_1, c_2, \ldots, c_m$, not all of which are zero, such that

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)} = 0,$$

that is, the zero vector can be written as a *nontrivial linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. For example, the vectors $(1,4), (3,-1)$, and $(3,5)$ in $\Re^2$ are linearly dependent since

$$3(1,4) + 1(3,-2) - 2(3,5) = (0,0).$$

Vectors are *linearly independent* provided they are not linearly dependent. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are a *basis* of $V$ provided they are linearly independent and span $V$. By an *ordered basis* we mean a basis in which the vectors of the basis are listed in a specified order; to indicate that we have an ordered basis we write $(u^{(1)}, u^{(2)}, \ldots, u^{(m)})$. A spanning set $S$ of $V$ is a *minimal spanning set of $V$* provided that each set of vectors obtained from $S$ by removing a vector is not a spanning set for $V$. A linearly independent set $S$ of vectors of $V$ is a *maximal linearly independent set of vectors of $V$* provided that for each vector $w$ of $V$ that is not in $S$, $S \cup \{w\}$ is linearly dependent (when this happens, $w$ must be a linear combination of the vectors in $S$). $\square$

In addition to matrix addition, subtraction, and multiplication, there is one additional operation that we define now. It's perhaps the simplest of them all. Let $A = [a_{ij}]$ be an $m$ by $n$ matrix and let $c$ be a number [**?**]. Then the matrix $c \cdot A$, or simply $cA$, is the $m$ by $n$ matrix obtained by multiplying each entry of $A$ by $c$:

$$cA = [ca_{ij}].$$

The matrix $cA$ is called a *scalar multiple* of $A$.

---

## Think About It...

Commonly thought of as the first modern computer, ENTAC was built in 1944. It took up more space than an 18-wheeler's tractor trailer and weighed more than 17 Chevrolet Camaros. It consumed 140,000 watts of electricity while executing up to 5,000 basic arithmetic operations per second. One of today's popular microprocessors, the 486, is built on a tiny piece of silicon about the size of a dime.

With the continual expansion of capabilities, computing power

will eventually exceed the capacity for human comprehension or human control.

<div align="right">

The Information Revolution
*Business Week*

</div>

## 8.3   Glossary

**360 Degree Review:** Performance review that includes feedback from superiors, peers, subordinates, and clients.

**Abnormal Variation:** Changes in process performance that cannot be accounted for by typical day-to-day variation. Also referred to as nonrandom variation.

**Acceptable Quality Level (AQL):** The minimum number of parts that must comply with quality standards, usually stated as a percentage.

**Activity:** The tasks performed to change inputs into outputs.

**Adaptable:** An adaptable process is designed to maintain effectiveness and efficiency as requirements change. The process is deemed adaptable when there is agreement among suppliers, owners, and customers that the process will meet requirements throughout the strategic period.

# 9

# *Synchronic and Diachronic Embeddings*

**CONTENTS**

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

## 9.1 Introduction

The term reliability usually refers to the probability that a component or system will operate satisfactorily either at any particular instant at which it is required or for a certain length of time. Fundamental to quantifying reliability s a knowledge of how to define, assess and combine probabilities [**?**]. This may hinge on identifying the form of the variability which is nherent n most processes. If all components had a fixed known lifetime there would be no need to model reliability.

1. A component part for an electronic item is manufactured at one of three different factories.

2. A component part $x$ for an electronic item is manufactured at one of three different factories.

(a) A component part for an electronic item is manufactured at one of three different factories.

(b) A component part $x$ for an electronic item is manufactured at one of three different factories.

    i. A component part for an electronic item is manufactured at one of three different factories.

    ii. A component part $x$ for an electronic item is manufactured at one of three different factories.

    iii. A component part for an electronic item is manufactured at one of three different factories.

    iv. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

    v. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

(c) A component part for an electronic item is manufactured at one of three different factories.

(d) A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

(e) A component part for enumerate list of an electronic item is manufactured at one of three different factories.

3. A component part for an electronic item is manufactured at one of three different factories.

4. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

5. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

### 9.1.1   A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty [**?**]? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main

**TABLE 9.1**

Now we are engaged $\left(a_g^a\right)$ $\left(a_g^a\right)$ in a great civil
war, testing whether that nation, or any
nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|-------|-----------|-----------|-----------|
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi[a] | 46, 281 | 6, 898 | 5, 400 |
| Keybrd[b] | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

assembly line.Of the total number supplied, factory A supplies 50%, factory
B 30%, and factory C 20%. Of the components manufactured at factory A,
1% are faulty and the corresponding proportions for factories B and C are
4% and 2% respectively. A component is picked at random from the assembly
line. What is the probability that it is faulty?

"A Process is a structured, measured set of activities designed to
produce a specific output for a particular customer or market—
A process is thus a specific ordering of work activities across
time and space, with a beginning, an end. and clearly defined
inputs and outputs: a structure for action."

Thomas Davenport
Senior Adjutant to the Junior Marketing VP

**MultiRelational $k$-Anonymity.** Most works on $k$-anonymity focus on
anonymizing a single data table; however, a real-life [**?**] database usually con-
tains multiple relational tables. This has proposed a privacy model called *Mul-
tiR $k$-anonymity* to ensure $k$-anonymity on multiple relational tables. Their
model assumes that a relational database contains a person-specific table $PT$
and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and
some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys,
some attributes in $QID$, and sensitive attributes. The general privacy notion
is to ensure that for each record owner $o$ contained in the join of all tables
$PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share
the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization
is applied at the *record owner* level, not at the *record* level in traditional $k$-
anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and
$Y = \{Pid\}$.

Most works on $k$-anonymity focus on anonymizing a single data table;
however, a real-life [**?**] database usually contains multiple relational tables.

This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k-1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

In most literature on PPDP, they [?] consider a more relaxed, yet more practical, notion of privacy protection by assuming limited attacker's background knowledge. Below, the term "victim" refers to the record owner being linked. We can broadly classify linking models to two families.

> A component part for an electronic item is [?] manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively.

One family considers a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call them *record linkage*, *attribute linkage*, and *table linkage*, respectively. In all types of linkages, we assume that the attacker knows the $QID$ of the victim. In record and attribute linkages, we further assume that the attacker knows the presence of the victim's record in the released table, and seeks to identify the victim's record and/or sensitive information from the table [?]. In table linkage, the attack seeks to determine the present or absent of the victim's record in the released table. A data table is considered to privacy preserved

**TABLE 9.2**

Now we are engaged $\left(a_g^a\right)\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Table Head | | | |
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi | 46, 281 | 6, 898 | 5, 400 |
| Keybrd | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

if the table can effectively prevent the attacker from successfully performing these types of linkages on the table [**?**]. Sections 12.1-12.2 study this family of privacy models.

$$\text{var}\widehat{\Delta} = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \text{var}\,(\hat{\alpha}_j - \hat{\alpha}_k) = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \sigma^2(1/n_j + 1/n_k). \qquad (9.1)$$

An obvious measure of imbalance is just the difference in the number of times the two treatments are allocated

$$D_n = \mathcal{M}|n_A - n_B|. \qquad (9.2)$$

For rules such as deterministic allocation, for which the expected value of this difference can be calculated, we obtain the population value $\mathcal{D}_n$.

---

**Box Title Here**

Another family aims at achieving the *uninformative principle*: The published table should provide the attacker with little additional information beyond the background knowledge. There should not be a large difference between the prior and posterior beliefs; otherwise, there is a privacy threat [**?, ?**]. Many privacy models in this family are designed for statistical database and do not distinguish attributes in $T$ into $QID$, but some of them could also thwart record, attribute, and table linkages. Section 12.1 studies this family of privacy models.

Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.

---

**Theorem 9** *Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.*

**Proof 9** *Most of the proof of this theorem is routine. It is clear that $0 \in Z_m$ and $1 \in Z_m$ are the zero element and identity element. If $a \in Z_m$ and $a \neq 0$, then $m - a$ is the additive inverse of $a$. If $a \in Z_m$ and $a \neq 0$, then the greatest common divisor of $a$ and $m$ is 1, and hence there exist integers $s$ and $t$ such that $sa + tm = 1$. Thus $sa = 1 - tm$ is congruent to 1 modulo $m$. Let $s^*$ be the integer in $Z_m$ congruent to $s$ modulo $m$. Then we also have $s^*a \equiv 1 \mod m$. Hence $s^*$ is the multiplicative inverse of $a$ modulo $m$. Verification of the rest of the field properties is now routine.*

## 9.2 Record Linkage Model

In the privacy attack of *record linkage*, some value *qid* on $QID$ identifies a small number of records in the released table $T$, called a *group*. If the victim's
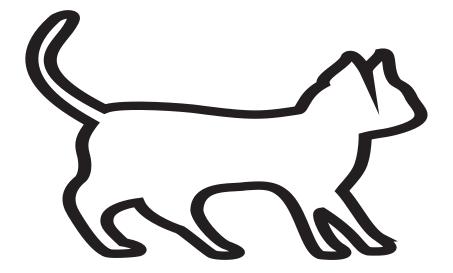
**FIGURE 9.1**
Figure caption goes here.

*QID* matches the value *qid*, the victim is vulnerable to being linked to the small number of records in the group [**?**]. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group.

### 9.2.1 A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

#### 9.2.1.1 H3 A component part

A component part for an electronic item is manufactured at one of three [**?**] different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A
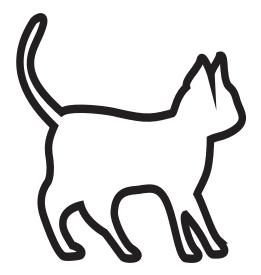
**FIGURE 9.2**
Figure caption goes here. Figure caption goes here. Figure caption goes here.
Figure caption goes here. Figure caption goes here. Figure caption goes here.

component is picked at random from the assembly line. What is the probability
that it is faulty?

A fundamental notion [**?**] is that of a subspace of $F^n$. Let $V$ be a nonempty
subset of $F^n$. Then $V$ is a *subspace* of $F^n$ provided $V$ is closed under vector
addition and scalar multiplication, that is,

(a) For all $u$ and $v$ in $V$, $u + v$ is also in $V$.

(b) For all $u$ in $V$ and $c$ in $F$, $cu$ is in $V$.

Let $u$ be in the subspace $V$. Because $0u = 0$, it follows that the zero vector
is in $V$. Similarly, $-u$ is in $V$ for all $u$ in $V$. A simple example of a subspace
of $F^n$ is the set of all vectors $(0, a_2, \ldots, a_n)$ with first coordinate equal to 0.
The zero vector itself is a subspace.

**Definition 9** Let $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ be vectors in $F^n$, and let $c_1, c_2, \ldots, c_m$
be scalars. Then the vector

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)}$$

is called a *linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. If $V$ is a subspace of
$F^n$, then $V$ is closed under vector addition and scalar multiplication, and
it follows easily by induction that a linear combination of vectors in $V$ is
also a vector in $V$. Thus *subspaces are closed under linear combinations*; in
fact, this can be taken as the defining property of subspaces. The vectors
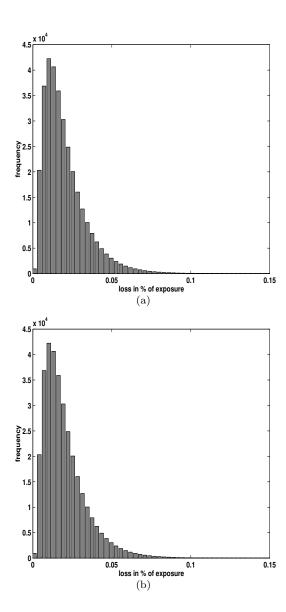
(a)



(b)

**FIGURE 9.3**
The bar charts depict the different risk contributions (top: 99% quantile, bottom: 99.9% quantile) of the business areas of a bank. The black bars are based on a Var/Covar approach, the white ones correspond to shortfall risk.

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ *span* $V$ (equivalently, form a *spanning set* of $V$) provided every vector in $V$ is a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. The zero vector can be written as a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ with all scalars equal to 0; this is a *trivial linear combination*. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are *linearly dependent* provided there are scalars $c_1, c_2, \ldots, c_m$, not all of which are zero, such that

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)} = 0,$$

that is, the zero vector can be written as a *nontrivial linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. For example, the vectors $(1,4), (3,-1)$, and $(3,5)$ in $\Re^2$ are linearly dependent since

$$3(1,4) + 1(3,-2) - 2(3,5) = (0,0).$$

Vectors are *linearly independent* provided they are not linearly dependent. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are a *basis* of $V$ provided they are linearly independent and span $V$. By an *ordered basis* we mean a basis in which the vectors of the basis are listed in a specified order; to indicate that we have an ordered basis we write $(u^{(1)}, u^{(2)}, \ldots, u^{(m)})$. A spanning set $S$ of $V$ is a *minimal spanning set of $V$* provided that each set of vectors obtained from $S$ by removing a vector is not a spanning set for $V$. A linearly independent set $S$ of vectors of $V$ is a *maximal linearly independent set of vectors of $V$* provided that for each vector $w$ of $V$ that is not in $S$, $S \cup \{w\}$ is linearly dependent (when this happens, $w$ must be a linear combination of the vectors in $S$). $\square$

In addition to matrix addition, subtraction, and multiplication, there is one additional operation that we define now. It's perhaps the simplest of them all. Let $A = [a_{ij}]$ be an $m$ by $n$ matrix and let $c$ be a number [**?**]. Then the matrix $c \cdot A$, or simply $cA$, is the $m$ by $n$ matrix obtained by multiplying each entry of $A$ by $c$:

$$cA = [ca_{ij}].$$

The matrix $cA$ is called a *scalar multiple* of $A$.

---

### *Think About It...*

Commonly thought of as the first modern computer, ENTAC was built in 1944. It took up more space than an 18-wheeler's tractor trailer and weighed more than 17 Chevrolet Camaros. It consumed 140,000 watts of electricity while executing up to 5,000 basic arithmetic operations per second. One of today's popular microprocessors, the 486, is built on a tiny piece of silicon about the size of a dime.

With the continual expansion of capabilities, computing power

will eventually exceed the capacity for human comprehension or human control.

The Information Revolution
*Business Week*

## 9.3　Glossary

**360 Degree Review:** Performance review that includes feedback from superiors, peers, subordinates, and clients.

**Abnormal Variation:** Changes in process performance that cannot be accounted for by typical day-to-day variation. Also referred to as non-random variation.

**Acceptable Quality Level (AQL):** The minimum number of parts that must comply with quality standards, usually stated as a percentage.

**Activity:** The tasks performed to change inputs into outputs.

**Adaptable:** An adaptable process is designed to maintain effectiveness and efficiency as requirements change. The process is deemed adaptable when there is agreement among suppliers, owners, and customers that the process will meet requirements throughout the strategic period.

# Part IV

# Advanced topics

# 10

## Clustering and Classification

**CONTENTS**

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

## 10.1 Introduction

The term reliability usually refers to the probability that a component or system will operate satisfactorily either at any particular instant at which it is required or for a certain length of time. Fundamental to quantifying reliability s a knowledge of how to define, assess and combine probabilities [**?**]. This may hinge on identifying the form of the variability which is nherent n most processes. If all components had a fixed known lifetime there would be no need to model reliability.

1. A component part for an electronic item is manufactured at one of three different factories.

2. A component part $x$ for an electronic item is manufactured at one of three different factories.

(a) A component part for an electronic item is manufactured at one of three different factories.

(b) A component part $x$ for an electronic item is manufactured at one of three different factories.

    i. A component part for an electronic item is manufactured at one of three different factories.

    ii. A component part $x$ for an electronic item is manufactured at one of three different factories.

    iii. A component part for an electronic item is manufactured at one of three different factories.

    iv. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

    v. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

(c) A component part for an electronic item is manufactured at one of three different factories.

(d) A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

(e) A component part for enumerate list of an electronic item is manufactured at one of three different factories.

3. A component part for an electronic item is manufactured at one of three different factories.

4. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

5. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

### 10.1.1　A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty [**?**]? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main

**TABLE 10.1**

Now we are engaged $\left(a_g^a\right)$ $\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi[a] | 46, 281 | 6, 898 | 5, 400 |
| Keybrd[b] | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

"A Process is a structured, measured set of activities designed to produce a specific output for a particular customer or market— A process is thus a specific ordering of work activities across time and space, with a beginning, an end. and clearly defined inputs and outputs: a structure for action."

Thomas Davenport
Senior Adjutant to the Junior Marketing VP

**MultiRelational $k$-Anonymity.** Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables. This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables.

This has proposed a privacy model called *MultiR k-anonymity* to ensure *k*-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner *o* contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with *o*. It is important to emphasize that the *k*-anonymization is applied at the *record owner* level, not at the *record* level in traditional *k*-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

In most literature on PPDP, they [**?**] consider a more relaxed, yet more practical, notion of privacy protection by assuming limited attacker's background knowledge. Below, the term "victim" refers to the record owner being linked. We can broadly classify linking models to two families.

> A component part for an electronic item is [**?**] manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively.

One family considers a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call them *record linkage*, *attribute linkage*, and *table linkage*, respectively. In all types of linkages, we assume that the attacker knows the $QID$ of the victim. In record and attribute linkages, we further assume that the attacker knows the presence of the victim's record in the released table, and seeks to identify the victim's record and/or sensitive information from the table [**?**]. In table linkage, the attack seeks to determine the present or absent of the victim's record in the released table. A data table is considered to privacy preserved

**TABLE 10.2**
Now we are engaged $\left(a_g^a\right) \left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Table Head | | | |
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi | 46, 281 | 6, 898 | 5, 400 |
| Keybrd | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

if the table can effectively prevent the attacker from successfully performing these types of linkages on the table [**?**]. Sections 12.1-12.2 study this family of privacy models.

$$\text{var}\widehat{\Delta} = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \text{var}\,(\hat{\alpha}_j - \hat{\alpha}_k) = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \sigma^2(1/n_j + 1/n_k). \qquad (10.1)$$

An obvious measure of imbalance is just the difference in the number of times the two treatments are allocated

$$D_n = \mathcal{M}|n_A - n_B|. \qquad (10.2)$$

For rules such as deterministic allocation, for which the expected value of this difference can be calculated, we obtain the population value $\mathcal{D}_n$.

---

**Box Title Here**

Another family aims at achieving the *uninformative principle*: The published table should provide the attacker with little additional information beyond the background knowledge. There should not be a large difference between the prior and posterior beliefs; otherwise, there is a privacy threat [**?, ?**]. Many privacy models in this family are designed for statistical database and do not distinguish attributes in $T$ into $QID$, but some of them could also thwart record, attribute, and table linkages. Section 12.1 studies this family of privacy models.

Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.

---

**Theorem 10** *Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.*

**Proof 10** *Most of the proof of this theorem is routine. It is clear that $0 \in Z_m$ and $1 \in Z_m$ are the zero element and identity element. If $a \in Z_m$ and $a \neq 0$, then $m - a$ is the additive inverse of $a$. If $a \in Z_m$ and $a \neq 0$, then the greatest common divisor of $a$ and $m$ is 1, and hence there exist integers $s$ and $t$ such that $sa + tm = 1$. Thus $sa = 1 - tm$ is congruent to 1 modulo $m$. Let $s^*$ be the integer in $Z_m$ congruent to $s$ modulo $m$. Then we also have $s^*a \equiv 1 \bmod m$. Hence $s^*$ is the multiplicative inverse of $a$ modulo $m$. Verification of the rest of the field properties is now routine.*

## 10.2   Record Linkage Model

In the privacy attack of *record linkage*, some value $qid$ on $QID$ identifies a small number of records in the released table $T$, called a *group*. If the victim's
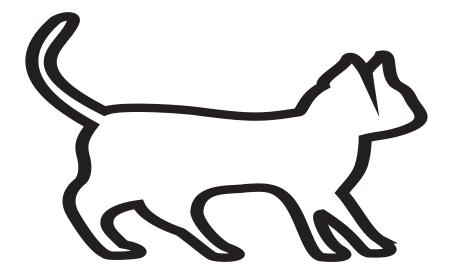
**FIGURE 10.1**
Figure caption goes here.

$QID$ matches the value $qid$, the victim is vulnerable to being linked to the small number of records in the group [**?**]. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group.

### 10.2.1 A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

#### 10.2.1.1 H3 A component part

A component part for an electronic item is manufactured at one of three [**?**] different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A
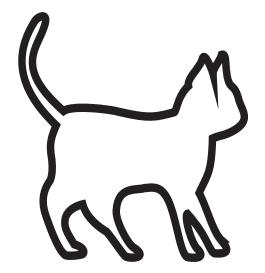
**FIGURE 10.2**
Figure caption goes here. Figure caption goes here. Figure caption goes here.
Figure caption goes here. Figure caption goes here. Figure caption goes here.

component is picked at random from the assembly line. What is the probability
that it is faulty?

A fundamental notion [**?**] is that of a subspace of $F^n$. Let $V$ be a nonempty
subset of $F^n$. Then $V$ is a *subspace* of $F^n$ provided $V$ is closed under vector
addition and scalar multiplication, that is,

   (a)   For all $u$ and $v$ in $V$, $u + v$ is also in $V$.

   (b)   For all $u$ in $V$ and $c$ in $F$, $cu$ is in $V$.

Let $u$ be in the subspace $V$. Because $0u = 0$, it follows that the zero vector
is in $V$. Similarly, $-u$ is in $V$ for all $u$ in $V$. A simple example of a subspace
of $F^n$ is the set of all vectors $(0, a_2, \ldots, a_n)$ with first coordinate equal to 0.
The zero vector itself is a subspace.

**Definition 10** Let $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ be vectors in $F^n$, and let $c_1, c_2, \ldots, c_m$
be scalars. Then the vector

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)}$$

is called a *linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. If $V$ is a subspace of
$F^n$, then $V$ is closed under vector addition and scalar multiplication, and
it follows easily by induction that a linear combination of vectors in $V$ is
also a vector in $V$. Thus *subspaces are closed under linear combinations*; in
fact, this can be taken as the defining property of subspaces. The vectors
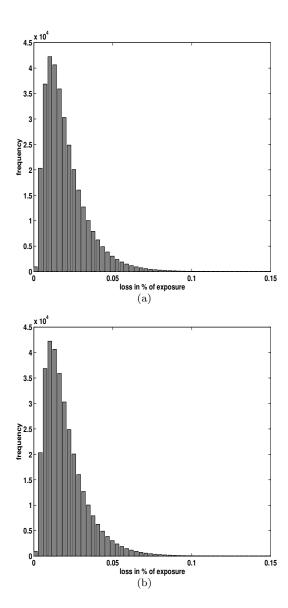
**FIGURE 10.3**
The bar charts depict the different risk contributions (top: 99% quantile, bottom: 99.9% quantile) of the business areas of a bank. The black bars are based on a Var/Covar approach, the white ones correspond to shortfall risk.

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ *span* $V$ (equivalently, form a *spanning set* of $V$) provided every vector in $V$ is a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. The zero vector can be written as a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ with all scalars equal to 0; this is a *trivial linear combination*. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are *linearly dependent* provided there are scalars $c_1, c_2, \ldots, c_m$, not all of which are zero, such that

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)} = 0,$$

that is, the zero vector can be written as a *nontrivial linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. For example, the vectors $(1,4), (3,-1)$, and $(3,5)$ in $\Re^2$ are linearly dependent since

$$3(1,4) + 1(3,-2) - 2(3,5) = (0,0).$$

Vectors are *linearly independent* provided they are not linearly dependent. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are a *basis* of $V$ provided they are linearly independent and span $V$. By an *ordered basis* we mean a basis in which the vectors of the basis are listed in a specified order; to indicate that we have an ordered basis we write $(u^{(1)}, u^{(2)}, \ldots, u^{(m)})$. A spanning set $S$ of $V$ is a *minimal spanning set of $V$* provided that each set of vectors obtained from $S$ by removing a vector is not a spanning set for $V$. A linearly independent set $S$ of vectors of $V$ is a *maximal linearly independent set of vectors of $V$* provided that for each vector $w$ of $V$ that is not in $S$, $S \cup \{w\}$ is linearly dependent (when this happens, $w$ must be a linear combination of the vectors in $S$). $\square$

In addition to matrix addition, subtraction, and multiplication, there is one additional operation that we define now. It's perhaps the simplest of them all. Let $A = [a_{ij}]$ be an $m$ by $n$ matrix and let $c$ be a number [**?**]. Then the matrix $c \cdot A$, or simply $cA$, is the $m$ by $n$ matrix obtained by multiplying each entry of $A$ by $c$:

$$cA = [ca_{ij}].$$

The matrix $cA$ is called a *scalar multiple* of $A$.

---

## Think About It...

Commonly thought of as the first modern computer, ENTAC was built in 1944. It took up more space than an 18-wheeler's tractor trailer and weighed more than 17 Chevrolet Camaros. It consumed 140,000 watts of electricity while executing up to 5,000 basic arithmetic operations per second. One of today's popular microprocessors, the 486, is built on a tiny piece of silicon about the size of a dime.

With the continual expansion of capabilities, computing power

will eventually exceed the capacity for human comprehension or human control.

The Information Revolution
*Business Week*

## 10.3   Glossary

**360 Degree Review:** Performance review that includes feedback from superiors, peers, subordinates, and clients.

**Abnormal Variation:** Changes in process performance that cannot be accounted for by typical day-to-day variation. Also referred to as non-random variation.

**Acceptable Quality Level (AQL):** The minimum number of parts that must comply with quality standards, usually stated as a percentage.

**Activity:** The tasks performed to change inputs into outputs.

**Adaptable:** An adaptable process is designed to maintain effectiveness and efficiency as requirements change. The process is deemed adaptable when there is agreement among suppliers, owners, and customers that the process will meet requirements throughout the strategic period.

# 11

## *Topic Models*

**CONTENTS**

A component part for an electronic item is manufactured at one of three differ-
ent factories, and then delivered to the main assembly line.Of the total number
supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the
components manufactured at factory A, 1% are faulty and the corresponding
proportions for factories B and C are 4% and 2% respectively. A component
is picked at random from the assembly line. What is the probability that it is
faulty?

## 11.1 Introduction

The term reliability usually refers to the probability that a component or
system will operate satisfactorily either at any particular instant at which
it is required or for a certain length of time. Fundamental to quantifying
reliability s a knowledge of how to define, assess and combine probabilities
[?]. This may hinge on identifying the form of the variability which is nherent
n most processes. If all components had a fixed known lifetime there would
be no need to model reliability.

1. A component part for an electronic item is manufactured at one of
   three different factories.

2. A component part $x$ for an electronic item is manufactured at one
   of three different factories.

    (a) A component part for an electronic item is manufactured at one of three different factories.

    (b) A component part $x$ for an electronic item is manufactured at one of three different factories.

        i. A component part for an electronic item is manufactured at one of three different factories.
        ii. A component part $x$ for an electronic item is manufactured at one of three different factories.
        iii. A component part for an electronic item is manufactured at one of three different factories.
        iv. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.
        v. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

    (c) A component part for an electronic item is manufactured at one of three different factories.

    (d) A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

    (e) A component part for enumerate list of an electronic item is manufactured at one of three different factories.

3. A component part for an electronic item is manufactured at one of three different factories.

4. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

5. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

### 11.1.1  A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty [**?**]? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main

**TABLE 11.1**

Now we are engaged $\left(a_g^a\right)$ $\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi[a] | 46, 281 | 6, 898 | 5, 400 |
| Keybrd[b] | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

"A Process is a structured, measured set of activities designed to produce a specific output for a particular customer or market— A process is thus a specific ordering of work activities across time and space, with a beginning, an end. and clearly defined inputs and outputs: a structure for action."

Thomas Davenport
Senior Adjutant to the Junior Marketing VP

**MultiRelational $k$-Anonymity.** Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables. This has proposed a privacy model called *MultiR $k$-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables.

This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \le i \le n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k - 1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

In most literature on PPDP, they [?] consider a more relaxed, yet more practical, notion of privacy protection by assuming limited attacker's background knowledge. Below, the term "victim" refers to the record owner being linked. We can broadly classify linking models to two families.

A component part for an electronic item is [?] manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively.

One family considers a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call them *record linkage*, *attribute linkage*, and *table linkage*, respectively. In all types of linkages, we assume that the attacker knows the $QID$ of the victim. In record and attribute linkages, we further assume that the attacker knows the presence of the victim's record in the released table, and seeks to identify the victim's record and/or sensitive information from the table [?]. In table linkage, the attack seeks to determine the present or absent of the victim's record in the released table. A data table is considered to privacy preserved

**TABLE 11.2**

Now we are engaged $\left(a_g^a\right)\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|-------|-----------|-----------|-----------|
| Table Head | | | |
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi | 46, 281 | 6, 898 | 5, 400 |
| Keybrd | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

if the table can effectively prevent the attacker from successfully performing these types of linkages on the table [**?**]. Sections 12.1-12.2 study this family of privacy models.

$$\text{var}\widehat{\Delta} = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \text{var}\,(\hat{\alpha}_j - \hat{\alpha}_k) = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \sigma^2(1/n_j + 1/n_k). \qquad (11.1)$$

An obvious measure of imbalance is just the difference in the number of times the two treatments are allocated

$$D_n = \mathcal{M}|n_A - n_B|. \qquad (11.2)$$

For rules such as deterministic allocation, for which the expected value of this difference can be calculated, we obtain the population value $\mathcal{D}_n$.

---

**Box Title Here**

Another family aims at achieving the *uninformative principle*: The published table should provide the attacker with little additional information beyond the background knowledge. There should not be a large difference between the prior and posterior beliefs; otherwise, there is a privacy threat [**?**, **?**]. Many privacy models in this family are designed for statistical database and do not distinguish attributes in $T$ into $QID$, but some of them could also thwart record, attribute, and table linkages. Section 12.1 studies this family of privacy models.

Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.

---

**Theorem 11** *Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.*

**Proof 11** *Most of the proof of this theorem is routine. It is clear that $0 \in Z_m$ and $1 \in Z_m$ are the zero element and identity element. If $a \in Z_m$ and $a \neq 0$, then $m - a$ is the additive inverse of $a$. If $a \in Z_m$ and $a \neq 0$, then the greatest common divisor of $a$ and $m$ is 1, and hence there exist integers $s$ and $t$ such that $sa + tm = 1$. Thus $sa = 1 - tm$ is congruent to 1 modulo $m$. Let $s^*$ be the integer in $Z_m$ congruent to $s$ modulo $m$. Then we also have $s^*a \equiv 1 \mod m$. Hence $s^*$ is the multiplicative inverse of $a$ modulo $m$. Verification of the rest of the field properties is now routine.*

## 11.2 Record Linkage Model

In the privacy attack of *record linkage*, some value $qid$ on $QID$ identifies a small number of records in the released table $T$, called a *group*. If the victim's
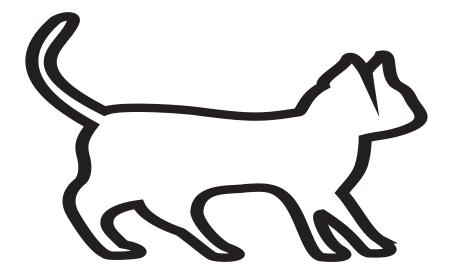
**FIGURE 11.1**
Figure caption goes here.

$QID$ matches the value $qid$, the victim is vulnerable to being linked to the small number of records in the group [**?**]. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group.

### 11.2.1   A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

#### 11.2.1.1   H3 A component part

A component part for an electronic item is manufactured at one of three [**?**] different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A
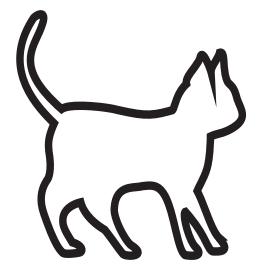
**FIGURE 11.2**
Figure caption goes here. Figure caption goes here. Figure caption goes here.
Figure caption goes here. Figure caption goes here. Figure caption goes here.

component is picked at random from the assembly line. What is the probability
that it is faulty?

A fundamental notion [**?**] is that of a subspace of $F^n$. Let $V$ be a nonempty
subset of $F^n$. Then $V$ is a *subspace* of $F^n$ provided $V$ is closed under vector
addition and scalar multiplication, that is,

(a)  For all $u$ and $v$ in $V$, $u + v$ is also in $V$.

(b)  For all $u$ in $V$ and $c$ in $F$, $cu$ is in $V$.

Let $u$ be in the subspace $V$. Because $0u = 0$, it follows that the zero vector
is in $V$. Similarly, $-u$ is in $V$ for all $u$ in $V$. A simple example of a subspace
of $F^n$ is the set of all vectors $(0, a_2, \ldots, a_n)$ with first coordinate equal to 0.
The zero vector itself is a subspace.

**Definition 11** Let $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ be vectors in $F^n$, and let $c_1, c_2, \ldots, c_m$
be scalars. Then the vector

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)}$$

is called a *linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. If $V$ is a subspace of
$F^n$, then $V$ is closed under vector addition and scalar multiplication, and
it follows easily by induction that a linear combination of vectors in $V$ is
also a vector in $V$. Thus *subspaces are closed under linear combinations*; in
fact, this can be taken as the defining property of subspaces. The vectors
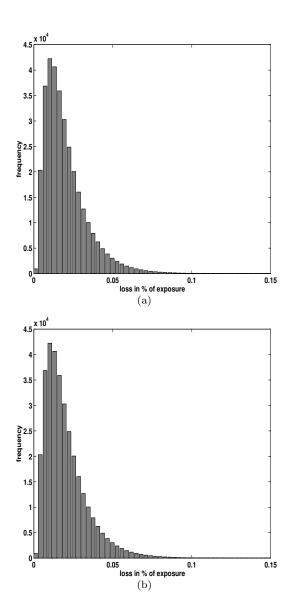
(a)



(b)

**FIGURE 11.3**
The bar charts depict the different risk contributions (top: 99% quantile, bottom: 99.9% quantile) of the business areas of a bank. The black bars are based on a Var/Covar approach, the white ones correspond to shortfall risk.

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ *span* $V$ (equivalently, form a *spanning set* of $V$) provided every vector in $V$ is a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. The zero vector can be written as a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ with all scalars equal to 0; this is a *trivial linear combination*. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are *linearly dependent* provided there are scalars $c_1, c_2, \ldots, c_m$, not all of which are zero, such that

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)} = 0,$$

that is, the zero vector can be written as a *nontrivial linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. For example, the vectors $(1,4), (3,-1)$, and $(3,5)$ in $\Re^2$ are linearly dependent since

$$3(1,4) + 1(3,-2) - 2(3,5) = (0,0).$$

Vectors are *linearly independent* provided they are not linearly dependent. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are a *basis* of $V$ provided they are linearly independent and span $V$. By an *ordered basis* we mean a basis in which the vectors of the basis are listed in a specified order; to indicate that we have an ordered basis we write $(u^{(1)}, u^{(2)}, \ldots, u^{(m)})$. A spanning set $S$ of $V$ is a *minimal spanning set of $V$* provided that each set of vectors obtained from $S$ by removing a vector is not a spanning set for $V$. A linearly independent set $S$ of vectors of $V$ is a *maximal linearly independent set of vectors of $V$* provided that for each vector $w$ of $V$ that is not in $S$, $S \cup \{w\}$ is linearly dependent (when this happens, $w$ must be a linear combination of the vectors in $S$). $\square$

In addition to matrix addition, subtraction, and multiplication, there is one additional operation that we define now. It's perhaps the simplest of them all. Let $A = [a_{ij}]$ be an $m$ by $n$ matrix and let $c$ be a number [**?**]. Then the matrix $c \cdot A$, or simply $cA$, is the $m$ by $n$ matrix obtained by multiplying each entry of $A$ by $c$:

$$cA = [ca_{ij}].$$

The matrix $cA$ is called a *scalar multiple* of $A$.

---

## *Think About It...*

Commonly thought of as the first modern computer, ENTAC was built in 1944. It took up more space than an 18-wheeler's tractor trailer and weighed more than 17 Chevrolet Camaros. It consumed 140,000 watts of electricity while executing up to 5,000 basic arithmetic operations per second. One of today's popular microprocessors, the 486, is built on a tiny piece of silicon about the size of a dime.

With the continual expansion of capabilities, computing power

will eventually exceed the capacity for human comprehension or human control.

The Information Revolution
*Business Week*

## 11.3   Glossary

**360 Degree Review:** Performance review that includes feedback from superiors, peers, subordinates, and clients.

**Abnormal Variation:** Changes in process performance that cannot be accounted for by typical day-to-day variation. Also referred to as non-random variation.

**Acceptable Quality Level (AQL):** The minimum number of parts that must comply with quality standards, usually stated as a percentage.

**Activity:** The tasks performed to change inputs into outputs.

**Adaptable:** An adaptable process is designed to maintain effectiveness and efficiency as requirements change. The process is deemed adaptable when there is agreement among suppliers, owners, and customers that the process will meet requirements throughout the strategic period.

# 12

## *Advanced Topics*

**CONTENTS**

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

## 12.1   Introduction

The term reliability usually refers to the probability that a component or system will operate satisfactorily either at any particular instant at which it is required or for a certain length of time. Fundamental to quantifying reliability s a knowledge of how to define, assess and combine probabilities [**?**]. This may hinge on identifying the form of the variability which is nherent n most processes. If all components had a fixed known lifetime there would be no need to model reliability.

1. A component part for an electronic item is manufactured at one of three different factories.

2. A component part $x$ for an electronic item is manufactured at one of three different factories.

(a) A component part for an electronic item is manufactured at one of three different factories.

(b) A component part $x$ for an electronic item is manufactured at one of three different factories.

    i. A component part for an electronic item is manufactured at one of three different factories.

    ii. A component part $x$ for an electronic item is manufactured at one of three different factories.

    iii. A component part for an electronic item is manufactured at one of three different factories.

    iv. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

    v. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

(c) A component part for an electronic item is manufactured at one of three different factories.

(d) A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

(e) A component part for enumerate list of an electronic item is manufactured at one of three different factories.

3. A component part for an electronic item is manufactured at one of three different factories.

4. A component part 1, 2, 3, 4 for an electronic item is manufactured at one of three different factories.

5. A component part for enumerate list of an electronic item is manufactured at one of three different factories.

### 12.1.1 A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty [**?**]? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line. Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty? A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main

**TABLE 12.1**

Now we are engaged $\left(a_g^a\right)$ $\left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|---|---|---|---|
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi[a] | 46, 281 | 6, 898 | 5, 400 |
| Keybrd[b] | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

---

"A Process is a structured, measured set of activities designed to produce a specific output for a particular customer or market— A process is thus a specific ordering of work activities across time and space, with a beginning, an end. and clearly defined inputs and outputs: a structure for action."

Thomas Davenport
Senior Adjutant to the Junior Marketing VP

---

**MultiRelational $k$-Anonymity.** Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables. This has proposed a privacy model called *MultiR $k$-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k-1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

Most works on $k$-anonymity focus on anonymizing a single data table; however, a real-life [**?**] database usually contains multiple relational tables.

This has proposed a privacy model called *MultiR k-anonymity* to ensure $k$-anonymity on multiple relational tables. Their model assumes that a relational database contains a person-specific table $PT$ and a set of tables $T_1, \cdots, T_n$, where $PT$ contains a person identifier $Pid$ and some sensitive attributes, and $T_i$, for $1 \leq i \leq n$, contains some foreign keys, some attributes in $QID$, and sensitive attributes. The general privacy notion is to ensure that for each record owner $o$ contained in the join of all tables $PT \bowtie T_1 \bowtie \cdots \bowtie T_n$, there exists at least $k-1$ other record owners share the same $QID$ with $o$. It is important to emphasize that the $k$-anonymization is applied at the *record owner* level, not at the *record* level in traditional $k$-anonymity. This idea is similar to $(X, Y)$-anonymity, where $X = QID$ and $Y = \{Pid\}$.

In most literature on PPDP, they [**?**] consider a more relaxed, yet more practical, notion of privacy protection by assuming limited attacker's background knowledge. Below, the term "victim" refers to the record owner being linked. We can broadly classify linking models to two families.

A component part for an electronic item is [**?**] manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively.

One family considers a privacy threat occurs when an attacker is able to link a record owner to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself. We call them *record linkage*, *attribute linkage*, and *table linkage*, respectively. In all types of linkages, we assume that the attacker knows the $QID$ of the victim. In record and attribute linkages, we further assume that the attacker knows the presence of the victim's record in the released table, and seeks to identify the victim's record and/or sensitive information from the table [**?**]. In table linkage, the attack seeks to determine the present or absent of the victim's record in the released table. A data table is considered to privacy preserved

**TABLE 12.2**

Now we are engaged $\left(a_g^a\right) \left(a_g^a\right)$ in a great civil war, testing whether that nation, or any nation so conceived.

| Scene | Reg. fts. | Hor. fts. | Ver. fts. |
|-------|-----------|-----------|-----------|
| Table Head | | | |
| Ball | 19, 221 | 4, 598 | 3, 200 |
| Pepsi | 46, 281 | 6, 898 | 5, 400 |
| Keybrd | 27, 290 | 2, 968 | 3, 405 |
| Pepsi | 14, 796 | 9, 188 | 3, 209 |

if the table can effectively prevent the attacker from successfully performing these types of linkages on the table [**?**]. Sections 12.1-12.2 study this family of privacy models.

$$\text{var}\widehat{\Delta} = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \text{var}\,(\hat{\alpha}_j - \hat{\alpha}_k) = \sum_{j=1}^{t} \sum_{k=j+1}^{t} \sigma^2(1/n_j + 1/n_k). \qquad (12.1)$$

An obvious measure of imbalance is just the difference in the number of times the two treatments are allocated

$$D_n = \mathcal{M}|n_A - n_B|. \qquad (12.2)$$

For rules such as deterministic allocation, for which the expected value of this difference can be calculated, we obtain the population value $\mathcal{D}_n$.

---

**Box Title Here**

Another family aims at achieving the *uninformative principle*: The published table should provide the attacker with little additional information beyond the background knowledge. There should not be a large difference between the prior and posterior beliefs; otherwise, there is a privacy threat [**?, ?**]. Many privacy models in this family are designed for statistical database and do not distinguish attributes in $T$ into $QID$, but some of them could also thwart record, attribute, and table linkages. Section 12.1 studies this family of privacy models.

Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.

---

**Theorem 12** *Let $m$ be a prime number. With the addition and multiplication as defined above, $Z_m$ is a field.*

**Proof 12** *Most of the proof of this theorem is routine. It is clear that $0 \in Z_m$ and $1 \in Z_m$ are the zero element and identity element. If $a \in Z_m$ and $a \neq 0$, then $m - a$ is the additive inverse of $a$. If $a \in Z_m$ and $a \neq 0$, then the greatest common divisor of $a$ and $m$ is 1, and hence there exist integers $s$ and $t$ such that $sa + tm = 1$. Thus $sa = 1 - tm$ is congruent to 1 modulo $m$. Let $s^*$ be the integer in $Z_m$ congruent to $s$ modulo $m$. Then we also have $s^*a \equiv 1 \mod m$. Hence $s^*$ is the multiplicative inverse of $a$ modulo $m$. Verification of the rest of the field properties is now routine.*

## 12.2    Record Linkage Model

In the privacy attack of *record linkage*, some value $qid$ on $QID$ identifies a small number of records in the released table $T$, called a *group*. If the victim's
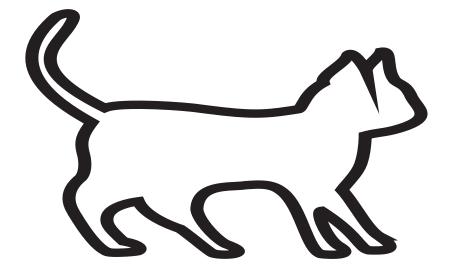
**FIGURE 12.1**
Figure caption goes here.

$QID$ matches the value $qid$, the victim is vulnerable to being linked to the small number of records in the group [**?**]. In this case, the attacker faces only a small number of possibilities for the victim's record, and with the help of additional knowledge, there is a chance that the attacker could uniquely identify the victim's record from the group.

### 12.2.1   A component part

A component part for an electronic item is manufactured at one of three different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A component is picked at random from the assembly line. What is the probability that it is faulty?

#### 12.2.1.1   H3 A component part

A component part for an electronic item is manufactured at one of three [**?**] different factories, and then delivered to the main assembly line.Of the total number supplied, factory A supplies 50%, factory B 30%, and factory C 20%. Of the components manufactured at factory A, 1% are faulty and the corresponding proportions for factories B and C are 4% and 2% respectively. A
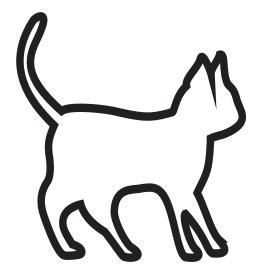
**FIGURE 12.2**
Figure caption goes here. Figure caption goes here. Figure caption goes here. Figure caption goes here. Figure caption goes here. Figure caption goes here.

component is picked at random from the assembly line. What is the probability that it is faulty?

A fundamental notion [**?**] is that of a subspace of $F^n$. Let $V$ be a nonempty subset of $F^n$. Then $V$ is a *subspace* of $F^n$ provided $V$ is closed under vector addition and scalar multiplication, that is,

(a)  For all $u$ and $v$ in $V$, $u + v$ is also in $V$.

(b)  For all $u$ in $V$ and $c$ in $F$, $cu$ is in $V$.

Let $u$ be in the subspace $V$. Because $0u = 0$, it follows that the zero vector is in $V$. Similarly, $-u$ is in $V$ for all $u$ in $V$. A simple example of a subspace of $F^n$ is the set of all vectors $(0, a_2, \ldots, a_n)$ with first coordinate equal to 0. The zero vector itself is a subspace.

**Definition 12** Let $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ be vectors in $F^n$, and let $c_1, c_2, \ldots, c_m$ be scalars. Then the vector

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)}$$

is called a *linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. If $V$ is a subspace of $F^n$, then $V$ is closed under vector addition and scalar multiplication, and it follows easily by induction that a linear combination of vectors in $V$ is also a vector in $V$. Thus *subspaces are closed under linear combinations*; in fact, this can be taken as the defining property of subspaces. The vectors
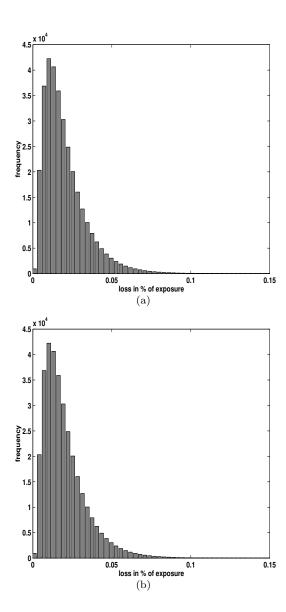
(a)

(b)

**FIGURE 12.3**
The bar charts depict the different risk contributions (top: 99% quantile, bottom: 99.9% quantile) of the business areas of a bank. The black bars are based on a Var/Covar approach, the white ones correspond to shortfall risk.

$u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ *span* $V$ (equivalently, form a *spanning set* of $V$) provided every vector in $V$ is a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. The zero vector can be written as a linear combination of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ with all scalars equal to 0; this is a *trivial linear combination*. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are *linearly dependent* provided there are scalars $c_1, c_2, \ldots, c_m$, not all of which are zero, such that

$$c_1 u^{(1)} + c_2 u^{(2)} + \cdots + c_m u^{(m)} = 0,$$

that is, the zero vector can be written as a *nontrivial linear combination* of $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$. For example, the vectors $(1,4), (3,-1)$, and $(3,5)$ in $\Re^2$ are linearly dependent since

$$3(1,4) + 1(3,-2) - 2(3,5) = (0,0).$$

Vectors are *linearly independent* provided they are not linearly dependent. The vectors $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ are a *basis* of $V$ provided they are linearly independent and span $V$. By an *ordered basis* we mean a basis in which the vectors of the basis are listed in a specified order; to indicate that we have an ordered basis we write $(u^{(1)}, u^{(2)}, \ldots, u^{(m)})$. A spanning set $S$ of $V$ is a *minimal spanning set of $V$* provided that each set of vectors obtained from $S$ by removing a vector is not a spanning set for $V$. A linearly independent set $S$ of vectors of $V$ is a *maximal linearly independent set of vectors of $V$* provided that for each vector $w$ of $V$ that is not in $S$, $S \cup \{w\}$ is linearly dependent (when this happens, $w$ must be a linear combination of the vectors in $S$). $\square$

In addition to matrix addition, subtraction, and multiplication, there is one additional operation that we define now. It's perhaps the simplest of them all. Let $A = [a_{ij}]$ be an $m$ by $n$ matrix and let $c$ be a number [**?**]. Then the matrix $c \cdot A$, or simply $cA$, is the $m$ by $n$ matrix obtained by multiplying each entry of $A$ by $c$:

$$cA = [ca_{ij}].$$

The matrix $cA$ is called a *scalar multiple* of $A$.

---

## *Think About It...*

Commonly thought of as the first modern computer, ENTAC was built in 1944. It took up more space than an 18-wheeler's tractor trailer and weighed more than 17 Chevrolet Camaros. It consumed 140,000 watts of electricity while executing up to 5,000 basic arithmetic operations per second. One of today's popular microprocessors, the 486, is built on a tiny piece of silicon about the size of a dime.

With the continual expansion of capabilities, computing power

will eventually exceed the capacity for human comprehension or human control.

The Information Revolution
*Business Week*

## 12.3   Glossary

**360 Degree Review:** Performance review that includes feedback from superiors, peers, subordinates, and clients.

**Abnormal Variation:** Changes in process performance that cannot be accounted for by typical day-to-day variation. Also referred to as non-random variation.

**Acceptable Quality Level (AQL):** The minimum number of parts that must comply with quality standards, usually stated as a percentage.

**Activity:** The tasks performed to change inputs into outputs.

**Adaptable:** An adaptable process is designed to maintain effectiveness and efficiency as requirements change. The process is deemed adaptable when there is agreement among suppliers, owners, and customers that the process will meet requirements throughout the strategic period.