Tutorial Paper on Quantitative Risk Assessment

Mohammad Reza Sohizadeh Abyaneh
Seyed Mehdi Mohammed Hassanzadeh
Håvard Raddum

Abstract

This paper shows how to carry out a quantitative risk assessment, describing how each step in the process is carried out. We use the grade management system at the University of Bergen as a case study, evaluating the risk of wrong grades ending up in the university grade database.

1 Introduction

Most of the current risk assessment methods are based on *scoring*. Risk is defined as the *likelihood of a scenario times its consequence*. The problem with this definition is that it loses the distinction between high likelihood, low consequence events and low likelihood, high consequence events. Furthermore, the definition is often interpreted as an *expected value* for risk, which contains too little information to be useful in practice. Considering central issues about perception of risks and *uncertainties* is also one of the most important drawbacks of the scoring methods. That is why some researchers in the literature has some doubt in their usefulness, e.g. Cox in [1] has concluded the scoring methods are often "*worse than useless*".

However, the scoring methods are widely used to assess risk in different fields. They also serve as a basis for important decisions regarding terrorism, engineering disasters, and a range of business decisions. Scoring methods are used even though their weaknesses are known.

Quantitative Risk Assessment Methods

To avoid the problems with the conventional scoring methods, *quantitative risk assessment* (QRA) methods have been introduced with a new definition of risk. In these methods, the term risk is referred to as "likelihood of conversion of a source of danger into actual delivery of loss, injury, or some form of damage" [2]. This notion involves both *uncertainty* and some kind of *loss or damage* that might be received. It is worth mentioning that the main difference between QRA methods and conventional scoring methods is the notion of uncertainty in the definition of risk.

The first problem we face in assessing risk by this definition is how to measure uncertainty. Measurement of uncertainty is done by assigning a set of *probabilities* to a set of *possibilities*. In contrast to scoring methods, many of the assumptions

This paper was presented at the NIK-2011 conference; see http://www.nik.no/.

in consequence analysis and likelihood measuring in QRA, are represented in probabilistic format. The QRA methods make the risk assessment more realistic and truly probabilistic.

The most useful analytical form for expressing the concept of risk is a "set of triplets" [5]. When we ask the question "What is the risk?" we are really asking three questions; What can go wrong?, how likely is it to happen? and what are the consequences if it does happen?
The first question is answered by a set of scenarios called the "risk scenarios". The "end states" of the scenarios are the consequences in terms of fatalities, injuries, etc., and the "likelihood" refers to the likelihood of the scenarios. More formally, we can say that the risk, R, is the set of triplets:

$$R = \{\langle S_i, P_i, X_i \rangle\},\ i = 1, 2, ..., N \tag{1}$$

where S_i is a scenario identification or description; P_i, is the probability of that scenario; and X_i, is the consequence or evaluation measure of that scenario, i.e., the measure of damage.

Case Study

For the case study, we have chosen to focus on the system for handling grades given to students at the University of Bergen (UoB). This system is essentially the process from the moment an evaluator of an exam assigns a grade to a student, until the grade ends up in a database at UoB. The risk in this system is the possibility that some grades are changed from the evaluator's intent through the system.

We have chosen this system because wrong grades may have serious consequences for affected individuals. Also, if there is doubt about the accuracy of the grades in the database at UoB it can significantly hurt the reputation of the university. As we can not tolerate many mistakes in the grade management system, we believe it is important to do a risk analysis to find the largest risks and provide an estimate of how many wrong grades there actually are in UoB's grade database.

2 Finding Risk Scenarios

This case study presents a QRA of *honest human errors*(mistakes) as a source of danger for the grading process at UoB. We follow a six-step QRA described in [2]:

1. Define the target system in terms of what constitutes normal operation.
2. Identify and characterize the sources of danger (hazards).
3. Develop "what can go wrong" scenarios to establish levels of damage and consequences while identifying points of vulnerability.
4. Quantify the likelihoods of the different scenarios and their levels of damage based on the totality of evidence available.
5. Assemble the scenarios according to damage levels, and present the results into risk curves.
6. Interpret the results to guide the risk management process.

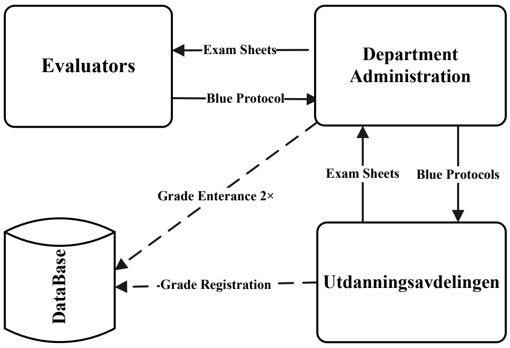


Figure 1: Flow of exam grades generated by one university department

Step 1: System During Normal Conditions

Figure 1 shows a sketch of the flow of grades coming from a university department. As can be seen, the system consists of four components: evaluator (EV), department administration (DA), Division of Student Affairs (Utdanningsavdelingen, UA) and a database. The three first components might be a source of danger in our scope so we will focus on the scenarios originating from them. The database system is managed by an environment called *Felles Studentsystem* (FS).

The flow of information in the grading system during normal operation can be summarized in the steps below.

1. Exam sheets produced by the students are collected by an exam supervisor and delivered to UA.
2. UA sorts exam sheets according to departments, and delivers them to DA.
3. DA sorts exam sheets according to courses, and delivers them to the respective EVs.
4. An EV grades an exam, and enters the grades into the *blue protocol*.
5. EV signs the grade protocol and delivers it back to DA.
6. DA enters the grades in FS
7. DA enters the grades in FS once more. The FS system checks that the grades are consistent with the values enetered the first time.
8. DA sends the grade protocol to UA
9. UA enters the grades in FS a third time. FS checks consistency with the grades already in the database.
10. UA publishes grades on studentweb and archives the grade protocol.

There are small variations in how some of the steps are carried out, and the description does not cover all aspects relevant for grades, such as oral exams and compulsory exercises. However, the steps above should give the reader a good overview of how the system works.

Step 2: Identification and Characterization of Hazards

The definition of risk in the QRA method includes loss or damage caused by a source of danger. For UoB's grade management system there is only one danger we are particularly concerned about: The possibility of a student receiving a different grade from what the evaluator intended. This kind of error may be due to honest mistakes, or due to active attacks. It is difficult to get meaningful estimates on the likelihood of attacks, so for the QRA we have limited ourselves to consider honest mistakes. We have identified three sources of such errors:

1. The evaluator mixes students or exam sheets when doing the evaluation, and writes a wrong grade on the grade protocol sheet.
2. The person at DA entering the grades in FS makes a mistake.
3. The person at UA entering grades in FS makes a mistake.

For every grade given in the system, any combination of these events can occur.

Step 3: Structuring of the Failure Scenarios and Consequences

The system operates normally when none of the three possible mistakes happen. For the other seven combinations of potential errors, we assume three combinations to be impossible. If DA made no mistake, but the person at UA enters a grade incorrectly from the grade protocol, FS will complain and the mistake will be corrected at once. These events are preceded by an evaluator who did or did not make a mistake, so they account for two scenarios. Similarly, if EV made no mistake, DA entered the grade wrongly in FS, but UA enters it correctly, DA's mistake will be corrected, leading to no damage. There are then four scenarios left to consider.

- *Scenario 1:* EV makes no mistake, but DA makes mistake when entering grades, and UA repeats the same mistake.
- *Scenario 2:* EV makes a mistake, DA does not make a mistake, and naturally UA can not make a mistake when entering grades in FS.
- *Scenario 3:* EV makes a mistake, DA makes a mistake, but the latter mistake is detected by the UA.
- *Scenario 4:* EV makes a mistake, DA makes a mistake, and the latter mistake is not detected by the UA.

Structuring Consequences

We now structure the consequences of the aforementioned possible scenarios in terms of four *damage levels*. Higher damage levels indicate more serious consequences. We categorize errors into two classes, mistakes in the grade protocol, and mismatch between grade in FS and protocol. We consider errors in the grade protocol to be a more serious mistake than a mismatch between digital data and the protocol. This is because there is no mechanism that can find an error in the grade protocol, but a mismatch between paper based and digital data can always be detected and corrected if someone sees it. We get the following damage levels.

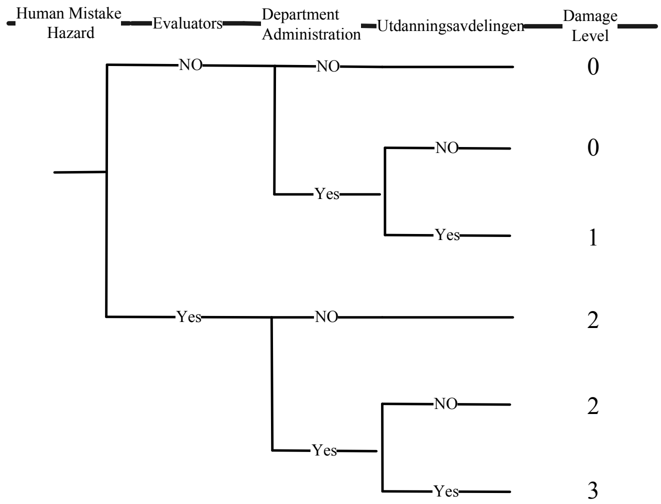


Figure 2: Normal operation and mapping scenarios onto damage levels

- *Damage Level 0*: No mistake
- *Damage Level 1*: Mistake in the digital data only (mismatch from grade protocol, which is correct)
- *Damage Level 2*: Mistake in grade protocol, propagating to digital data in FS
- *Damage Level 3*: Mistake in grade protocol, and mismatch between digital and paper based data

Figure 2 illustrates the possible paths originating from honest mistakes and terminating in a damage level. This figure also depicts a mapping from the scenarios to their consequences (damage levels).

3 Quantifying the Risks

We now proceed to estimate the likelihoods of the scenarios. All estimates involve the uncertainty by expressing them as Probability Density Functions (PDF).

Step 4 : Quantifying Likelihood of Scenarios

In the QRA methodology, the idea of likelihood is expressed by *frequency*. We ask how often a particular scenario occurs. Since we do not know the exact answer, the best we can do is to express our state of knowledge about the scenario in the form of a *probability of frequency* curve. The area on the horizontal axis where we get the highest function values is the area where the true frequency is mostly likely to exist. To quantify the likelihood of the scenarios we take the following four-step approach:

1. Identification of basic actors.
2. Gathering information on the frequencies of actors' mistakes.
3. Combining the information from step 2 to find the PDF of actors' mistakes.
4. Finding the PDF of scenarios based on the PDF's of basic components.

Evaluators	Frequency of Mistakes
EV1	1/100
EV2	0
EV3	1/300
EV4	$[0, 2/100]$
EV5	$[0, 10/100]$

Table 1: Frequency of mistakes made by EV

Department Administration	Frequency of Mistakes
DA1	1/60
DA2	1/75
DA3	1/200
DA4	1/250
DA5	1/300

Table 2: Frequency of mistakes made by DA

Identification of Basic Components

As Figures 1 and 2 illustrate, we have three actors (EV, DA and UA). Each actor has a risk of making a mistake when registering grades.

Gathering Information on the Frequencies of Basic Components

There estimate how often evaluators makes a mistake when writing grades, we interviewed evaluators and employees entering grades in FS, and combined their opinions. We talked to five people with experience as evaluators, and five people responsible for entering grades in FS, and asked them how often they think mistakes occur. The results are presented in Tables 1 and 2, respectively. Since the process of entering grades into the computer system is the same at UA and DA, we did not interview employees at UA (see Section 3).

As can be seen, the information is given in two forms; *A ratio* (the number of mistakes per grade) and *An interval* (with 90% confidence).

Finding the Likelihood (PDF) of Basic Components

In this section, we quantify the likelihood of the three basic components in terms of PDFs. We follow the approach taken in [4]. Let λ denote the frequency of making *mistakes per grade*. We express the information we have before any sampling in a prior PDF ($P_0(\lambda)$).

If we do not have any information about the system prior to sampling, we can chose a "flat" prior PDF, $P_0(\lambda) = 1.0$ for $0 \leq \lambda \leq 1$. In this way, we say that before we have any data, we consider all values to be equally likely.

Now, suppose that we receive some information, giving a piece of evidence E such that:

$$E = \{k \text{ mistakes are made in } m \text{ grades}\} \tag{2}$$

For instance, $k = 1$ and $m = 100$, for the first sample of the evaluators in Table 1. Then, we are able to combine this new evidence with our prior knowledge by Bayes' theorem, as follows:

$$P(\lambda|E) = P_0(\lambda)\frac{P(E|\lambda)}{P_0(E)} \tag{3}$$

Where $P(\lambda|E)$ here is the new PDF, which expresses our state of knowledge after we have become aware of E.

If we have information in the form of a ratio, we can use (4) to calculate $P(E|\lambda)$.

$$P(E|\lambda) = \binom{m}{k}\lambda^k(1-\lambda)^{m-k} \tag{4}$$

In order to use the information given in the form of intervals, we need to convert the intervals to PDFs. This is feasible by using (6) which gives us a *normal* PDF with the *mean* (μ) and *standard deviation* (σ) values computed from the lower and upper bounds of an interval $[a, b]$ with 90% confidence [3].

$$\mu = mean = \frac{(a+b)}{2}, \quad \sigma = \text{standard deviation} = \frac{(b-a)}{3.29} \tag{5}$$

$$P(E|\lambda) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\lambda-\mu)^2}{2\sigma^2}} \tag{6}$$

$P_0(E)$ can be calculated as:

$$P_0(E) = \int_0^1 P_0(\lambda) P(E|\lambda) d\lambda \tag{7}$$

Thus, our new PDF can be attained by:

$$P(\lambda|E) = P_0(\lambda) \frac{P(E|\lambda)}{\int_0^1 P_0(\lambda) P(E|\lambda) d\lambda} \tag{8}$$

In this way, each sample can be used as evidence to update our state of knowledge. The more information or samples we have, the less uncertainty we get about λ.

All the computations in this paper were done using Matlab. The functions were represented on the intervals in question as vectors of function values. Function values were taken at many points, separated with small distances. Integrals were computed numerically, using trapezoidal integration found in Matlab.

Evaluators

After updating our knowledge with all five samples from the evaluators (Table 1), the resulting PDF $P(e) = P(\lambda|E)$ shows how often evaluators make mistakes. This function is plotted in Figure 3(a). From the graph we see that the frequency of grading mistakes made by evaluators is, with very high probability, between 0 and 0.02. In other words it is very unlikely that evaluators make mistakes in more than 1 out of 50 grades.

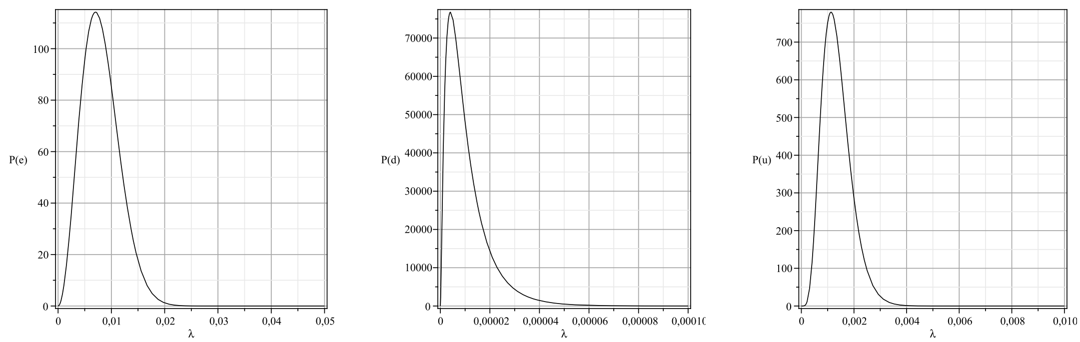
Department Administration

By applying the same steps to the values in Table 2, we obtain a PDF for mistakes made by someone at DA entering grades in FS. The question we asked when gathering this information was how often they find a mistake when entering grades the second time. The values in Table 2 are therefore the frequencies of mistakes when entering only *once*, and we denote this frequency by d' and its PDF by $P_{d'}(d')$.

To obtain the real probability density for mistakes being made in DA, we should get the frequency of entering a wrong grade *twice* in FS, *where both mistakes are identical*. So, we need to calculate the PDF of the random variable $d = \frac{d' \times d'}{5}$. We divide by five because the same mistake must be made the second time, among the five possible choices for a wrong grade.

To find the PDF of d from $P_{d'}(d')$, we can use (9).

$$P(d) = P(\sqrt{5d'}) \tag{9}$$



(a) EV (b) DA (c) UA

Figure 3: PDF(likelihood) of making mistakes

The result is plotted in Figure 3(b). As we should expect, the frequency of mistakes made dy DA and passed on to UA is very low, probably around 1 in $100,000$ and almost surely less than 1 in $25,000$.

Utdanningsavdelingen
For the frequency of mistakes made when grades are entered the third time at UA, we use the values in Table 2. We can consider the frequency of mistakes made at UA as $u = d'$. Therefore, the PDF for this component will be $P(u) = P_{d'}(u)$. This function is plotted in Figure 3(c).

Note that this function only models how often a grade is punched wrongly by someone entering the grades the third time. Since FS checks the consistency of each grade with what is already in the database, (almost) all mistakes are corrected as they happen. Only when an actual mistake has been made at the department administration, and the same mistake is repeated at UA will a grade have been wrongly entered in FS.

Finding the PDF of the Scenarios
Let e, d and u denote the likelihood of mistakes made by EV, DA and UA, respectively. Also, let a letter with bar be the complement of its value, e.g. $\overline{x} = 1-x$.

The likelihood of each scenario should be presented as a function of the likelihood of the basic components. Let us denote the likelihood, or PDF, of the scenarios by $P(S_1), P(S_2), P(S_3)$ and $P(S_4)$. The PDFs as functions of e, d and u are as follows:

$$\begin{aligned} P(S_1) &= P(\overline{e}.d.u) = P(d.u - e.d.u) \\ P(S_2) &= P(e.\overline{d}.\overline{u}) = P(e - e.u - e.d + e.d.u) \\ P(S_3) &= P(e.d.\overline{u}) = P(e.d - e.d.u) \\ P(S_4) &= P(e.d.u) \qquad (10) \end{aligned}$$

In order to use (10) to calculate the PDF of each scenario, we should first know how to calculate the PDF of a *product* and *sum* of two variables.

Let X and Y be continuous random variables with joint PDF $P_{X,Y}(x, y)$. In [6], an algorithm is proposed to calculate $P_V(v)$. This algorithm is based on the Theorem below.

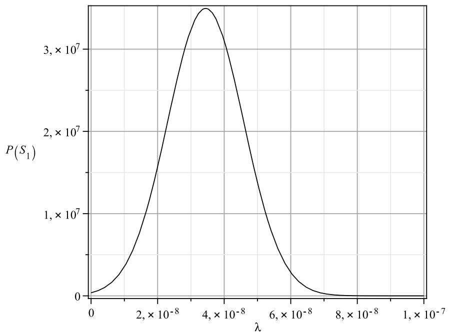


Figure 4: PDF of Scenario 1.

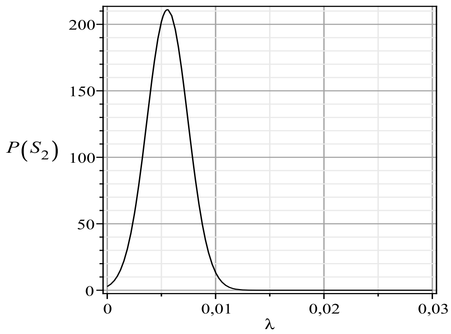


Figure 5: PDF of Scenario 2.

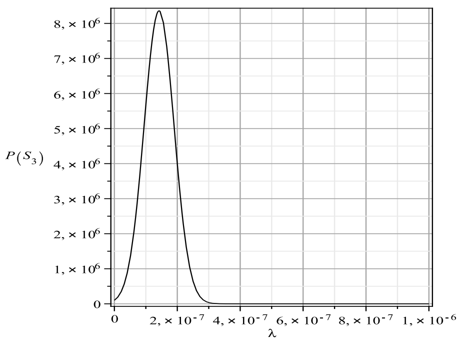


Figure 6: PDF of Scenario 3.

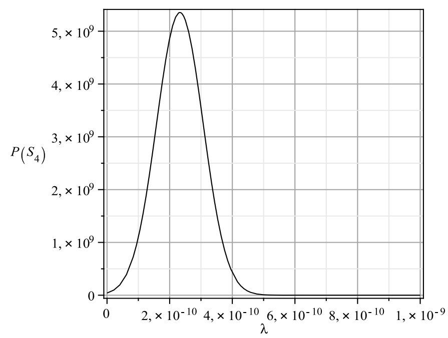


Figure 7: PDF of Scenario 4.

Theorem 1.[6] *Let X be a random variable of the continuous type with PDF $P_X(x)$ which is defined and positive on the interval (a, b), where $0 < a < b < \infty$. Similarly, let Y be a random variable of the continuous type with PDF $P_Y(y)$ which is defined and positive on the interval (c, d), where $0 < c < d < \infty$. The PDF of $V = XY$ is:*

$$P_V(v) = \begin{cases} \int_a^{v/c} P_Y(\frac{v}{x})P_X(x).\frac{1}{x}dx, & ac < v < ad \\ \int_{v/c}^{v/d} P_Y(\frac{v}{x})P_X(x).\frac{1}{x}dx, & ad < v < bc \\ \int_b^{v/d} P_Y(\frac{v}{x})P_X(x).\frac{1}{x}dx, & bc < v < bd \end{cases} \tag{11}$$

Finding the PDF of the summation of two random variables is more straightforward. Let X and Y be two continuous random variables with density functions $P_X(x)$ and $P_Y(y)$, respectively. Then the sum $W = X + Y$ is a random variable with density function $P_W(W)$,

$$P_W(W) = \int_{-\infty}^{\infty} P_Y(W - x).P_X(x)dx \tag{12}$$

By using (11) and (12), we can calculate the PDF curve for each scenario. Figures 4, 5, 6 and 7 depict the results.

Step 5 : Assembly of Scenarios into Measures of Risk

In Step 5, the scenarios are assembled into measures of risk. This means combining all scenarios that terminate in a specific damage level. In our case study scenarios 2 and 3, which both terminate to damage level 2, should be combined. The scenarios should then be arranged in order of increasing damage levels (Table 3).

Scenarios	**Likelihood**	**Damage Level**
S_1	$P(S_1)$	1
$S_2 + S_3$	$P(S_2 + S_3)$	2
S_4	$P(S_4)$	3

Table 3: Scenario List

Scenarios	**Likelihood**	**Damage Level**	**Cumulative Probability**
S_1	$P(S_1)$	1	$\Phi_1 = \Phi_2 + S_1$
S_2	$P(S_2)$	2	$\Phi_2 = \Phi_3 + S_2$
S_3	$P(S_3)$	2	$\Phi_3 = \Phi_4 + S_3$
S_4	$P(S_4)$	3	$\Phi_4 = S_4$

Table 4: Scenario List with Cumulative Probability

The most common form for representing risk is the classical risk curve. If we assume that in (1), the likelihood P is a function of frequency ϕ then we have:

$$R = \{\langle S_i, P_i(\phi_i), X_i \rangle\},\ i = 1, 2, ..., N \tag{13}$$

We illustrate how to determine a risk curve in Table 5. First, we consider the fourth column, where we write the cumulative probability, adding from the bottom, where Φ_i is damage i or greater.

Then, we calculate the PDF of each Φ_i denoted as $\prod_i(\Phi_i)$ by:

$$\prod_i(\Phi_i) = \int_0^{\Phi_i} \prod_{i+1}(\Phi_{i+1} P_i(\Phi_i - \Phi_{i+1}) d\Phi_{i+1} \tag{14}$$

Now we can plot the risk curves in terms of frequency vs. damage level as follows. For a particular damage level X,

1. Choose a certain probability p.
2. Find the frequency (Φ) with the probability p on $\prod_X(\phi_X)$.
3. Let the pair (X, Φ) be a point on the curve.

Plotting the points for different damage levels on logarithmic scales generates a risk curve and changing p gives us a set of curves as shown in Figure 8. The p's most often selected in practice are 0.05 and 0.95 (i.e. the 5th and 95th percentiles). A popular third choice is the *mean*.

We explain how to read Figure 8: Suppose p has the value of 0.95, and suppose we want to know the risk of suffering consequence X_1 or greater at the 95% confidence level. According to the figure, we are 95% confident that the frequency of a consequence X_1 or greater is Φ_1 which represents our cumulative risk.

The curves in Figure 8 show the total frequency for exceeding a particular damage level and the corresponding uncertainties in that frequency. For instance, the curves show that the total mean frequency of exceeding damage level 1 for our case study is 0.00845 mistakes per grade, or one in 118 grades. The uncertainty analysis indicates

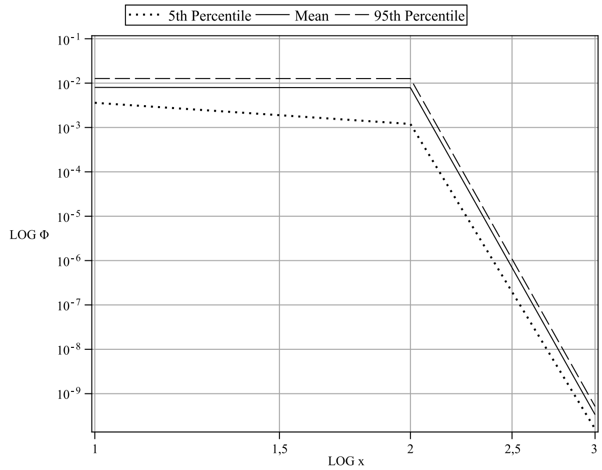


Figure 8: Cumulative risk curves. Horizontal axis shows damage levels, vertical axis shows likelihoods.

that there is 90% confidence that the true frequency lies within 0.0036 and 0.0127. In other words, the frequency of mistakes may be as high as 1 in 79 grades, or as low as 1 in 278 grades, and we are 90% sure the true value is somewhere in between these two values.

Figure 8 shows us that there is hardly any decrease in likelihood between damage level 1 and 2. This means damage level 2 is dominating the risk, and that damage level 1 happens very rarely. Between damage level 2 and 3 the mistake frequency decreases sharply. The mean frequency of suffering damage level 3 is approximately 3.367E-10 mistakes per grade, or once in 3 billion grades.

Step 6: Interpretation of the Results

Table 5 summarizes selected parameters of the uncertainty distributions for all damage levels. It is evident from this table that the likelihood of storing wrong grades in the database is much greater if there is a mistake in the paper data (damage level 2). Precisely speaking, we are 90% confident that the frequency of this consequence is between 3.54E-3 and 1.26E-2, or between once in 79 and 282 grades, respectively. This consequence is mainly influenced by mistakes made by evaluators when they enter grades in the grade protocol (Scenario 2). There is no mechanism for correcting mistakes made by evaluators, as there is for entering grades in FS. Evaluators thus become a single point of failure in the system.

Consequences which include mistakes in digital data (damage level 1 and 3) are quite unlikely. Namely, the frequencies of damage level 1 and 3 are on average 3.43E-8 and 3.37E-10 respectively. That is, approximately once in 29 million and 2.9 billion grades. This is so low that it has probably not happened in practice. The QRA shows that the multiple entering of grades in FS is sufficient to have complete correctness when grades are copied from paper to digital form. There are approximately 60,000 grades given at UoB per year, and the double check done at the DAs when entering grades in FS is so good that if the third check at UA is removed, it should not introduce more than at most one extra mistake per year.

Human Mistake Rate (Mistakes per Grade)			
Damage Level	**5th Percentile**	**95th Percentile**	**Mean**
Any Damage	3.53E-3	1.27E-2	8.03E-3
1	1.61E-8	5.33E-8	3.43E-8
2	3.54E-3	1.26E-2	8.04E-3
3	1.60E-10	5.16E-10	3.37E-10

Table 5: Selected parameters of uncertainty distribution for each level of damage

4 Conclusions and Future works

In this paper, we have shown how to carry out a QRA where the uncertainties of the underlying data is taken into account. For the grade management system at UoB, we found that the by far most likely scenario for getting wrong grades entered in the database is from evaluators to mix students and exam papers. There are approximately 60,000 grades given at UoB every year. Among these we expect that between 212 and 759 grades are different from the evaluator's intent.

When grades are entered into the database there are safeguards in place that should catch typing errors. The grades have to be entered three times into the database and have been consistent each time. We conclude that the current protection against typing errors is sufficient.

Due to lack of data on other parts of the grading system, like, the protection of grading servers, we were unable to assess the risk of those parts and we had to limit the scope of our work to only consider human errors. Therefore, the main possible extension of this work can be inclusion of other parts of the system to make the risk assessment more precise. Gathering more information on how often evaluators confuse students, exam papers and grades could be used to narrow the interval of where the actual frequency of mistakes is found.

References

[1] Douglas W. Hubbard, *The Failure of Risk Management: Why It's Broken and How to Fix It*, Published by John Wiley and Sons, Inc., Hoboken, New Jersey, ISBN 978-0-470-38795-5, 2009.

[2] B. John Garrick, *Quantifying and Controlling Catastrophic Risks*, Academic Press is an imprint of Elsevier, ISBN: 978-0-12-374601-6, 2008.

[3] Douglas W. Hubbard, *How to Measure Anything*, Published by John Wiley and Sons, Inc., Hoboken, New Jersey, ISBN 978-0-470-53939-2, 2010.

[4] Stan Kaplan, *ON THE USE OF DATA AND JUDGMENT IN PROBABILISTIC RISK AND SAFETY ANALYSIS*, Nuclear Engineering and Design, Vol. 93, pp. 123 – 134, North-Holland, 1986.

[5] Stanley Kaplan' and B. John Garrick, *On The Quantitative Definition of Risk*, Risk Analysis, Vol. 1, No. 1, pp. 11 – 27, 1981.

[6] Andrew G. Glen, Lawrence M. Leemis, John H. Drew, *Computing the distribution of the product of two continuous random variables*, Computational Statistics and Data Analysis, Elsevier, 2002.