AEROSPACE
Assuring Space Mission Success

Risks of Software Risk Management – Focusing on the Probability and Statistics Foundations

Dr. Peter Hantos
The Aerospace Corporation

Computers and Software Division/Software Engineering Subdivision
11 April 2013

Acknowledgements

- This work would not have been possible without the support of the following people of The Aerospace Corporation
  - *Asya Campbell*
  - *B. Zane Faught*
  - *Leslie J. Holloway*
  - *Dan X. Houston*
- Funding Source
  - The Aerospace Corporation’s Aerospace Technical Investment Program (ATIP,) Software Acquisition Long Term Capability Development (LTCD) Project

The Challenge

“Prediction is very difficult,
especially if it is about the future.”
**~~~ Niels Bohr**

“Surveys show that 117 percent of Americans don’t
understand statistics and 92 percent don’t care.”
**~~~ Jeffrey Goldberg, The Atlantic, JULY/AUGUST 2012**

Introduction - Risk and Statistics

- **Risk**
  - Risk is the potential for a negative future reality that may or may not happen*
  - A risk associated with an event has two components
    - **Probability** (likelihood,) expressing the fact that the event may or may not happen
    - **Consequence** (impact,) expressing the unwanted result of the event
- **Statistics**
  - **Descriptive Statistics**, which discipline is about the effective organization, summarization and communication of data
    - The term statistics in everyday use refers to descriptive statistics
  - **Inferential Statistics**, a discipline which purpose is arriving to conclusions that extend beyond the data
    - Inferential statistics deal with uncertainty and the relevant, essential concepts are based on the theory of probability

* Source: [DAU 2013]

A Feud of Statisticians You Were Probably Not Aware Of…

- **Frequentists see probability as the long-run expected frequency of occurrence**

$$P(A) = n/N$$

  - Frequentists base inferences for the unknown value from the distribution of statistics derived from repeated trials.

> **"In God we trust; all others must bring data."**
> **~~~ The Ultimate Frequentist, W. E. Deming***

- **Bayesians view of probability is related to degree of belief**
  - Bayesians base inferences about the unknown value under some a-priori model for the data; they believe that probability P(A) can have meaning for singular, unrepeated events

Legend:

P(A)	Probability of event **A**
N	Number of trials
n	number of times event **A** occurs during **N** trials

* W. E. Deming (1900 – 1993) was a famous American statistician, professor, author, lecturer and consultant

A Hardware Example

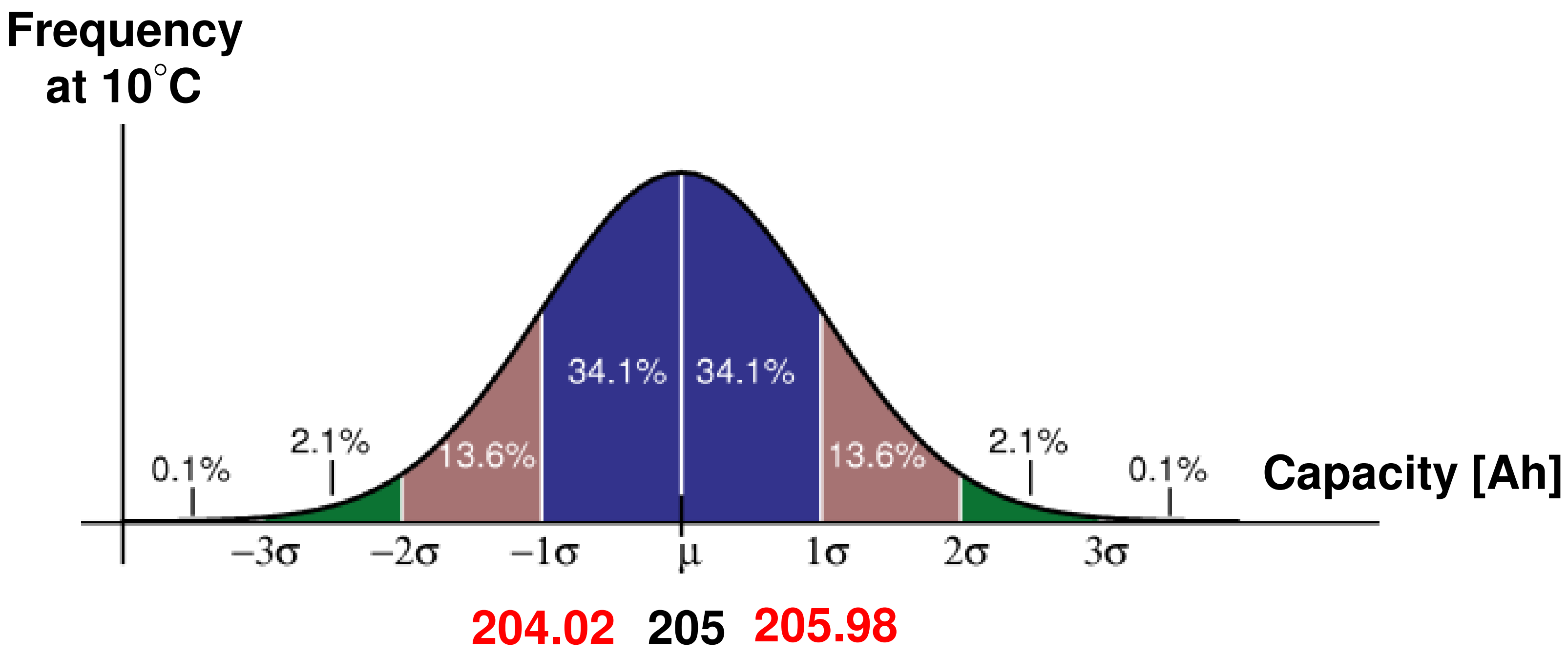
- **Details for Nickel-Hydrogen (NiH_2) battery development***
  - Objective is to develop a new family of 5.5” battery cells for satellites
  - Life cycle performance expectation is 120,000 charge/discharge cycles
    - *Low Earth Orbit (LEO) profile of a cycle: 54min charge – 36min discharge*
  - *Capacity test results at various temperatures for a cycle*

Temperature [°C]	Capacity – Mean (μ) [Ah]
20	188
10	205
0	220
-10	210

  - *Assumptions about unpublished but relevant information*
    - *A typical testing scenario may involve several batches of 100 cells*
    - *At the time of publication cell life test data was shown for the first 100 LEO cycles and it was stated that the test was continuing*
    - *Standard deviation (usually denoted as **σ**,) has been also computed*

* Based on [Caldwell 1998]

Risk Example – Picking a (Bad) Battery



<u>Question:</u>

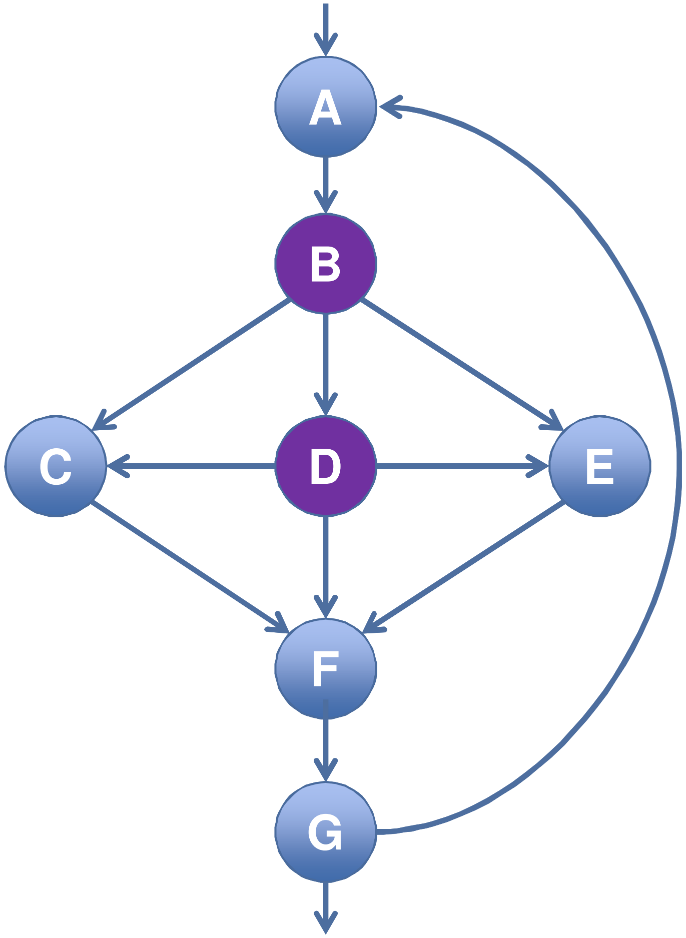
Assuming that the distribution is normal, what is the likelihood that a battery randomly picked from the batch will have a capacity lower than **204.02** Ah?

<u>Answer:</u>

The likelihood is (pretty close to…) 50 - 34.1 = **15.9%**

The frequentist view of probability fits well hardware

A Software Example*

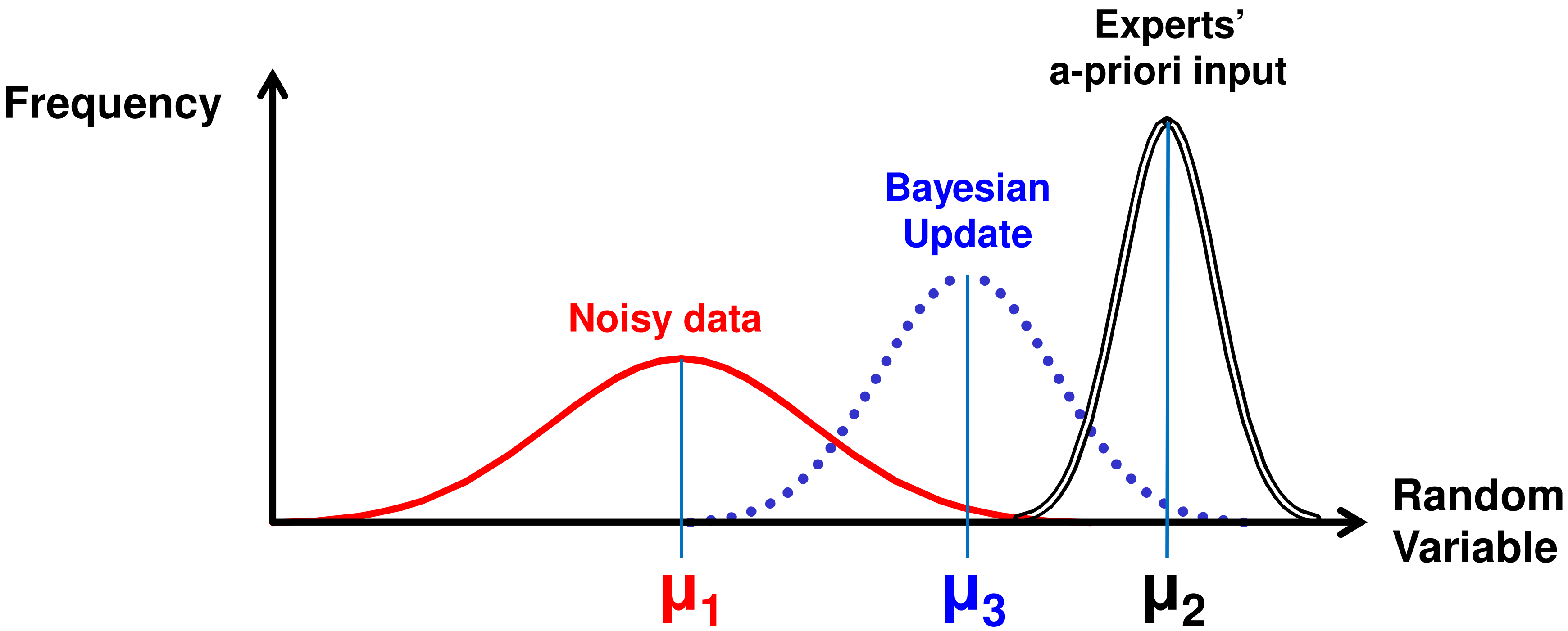


- **Path testing is just one of the numerous tests to discover software defects**
  - It is a structural testing method that involves the source code to attempt to find every possible executable path
- **Program graph on the left shows a simple program's topology**
  - Only 7 program nodes; two of them are conditional branches (B and D)
  - One feedback loop (G → A)
- **Assuming only 18 repetitions for the G → A loop**
  - The number of distinct possible execution paths we would need to test would be **4.77 trillion**

Clearly, blindly applying a frequentist approach is not feasible

* [Jorgensen 2002]

When Frequentists and Bayesians Meet…



Example: Bayesian calibration of noisy data*

- We are unhappy with the precision of μ_1 (The associated σ is too wide)
- A Delphi session of experts is organized that yields a mean (μ_2) that is different, but seems to be more precise (Narrower σ)
- A Bayesian calibration gives a stronger weight to this, a-priori information, causing the new, posterior mean (μ_3) to be closer to the a-priori mean

* Based on [Chulani 1998]

Now a Trick Question

- **A Contractor's 185-page Software Development Plan was reviewed and the comments were categorized as follows:**
  - **0** Critical
  - **92** Substantive
  - **10** Administrative
- **An assessment was made that the probability of program failure if they proceed development is p>50%. Which of the following statements is true?**

  A – This is definitely the correct probability

  B – The probability of failure is much higher, around 90%

  C – The program might not fail at all

Now the (Trick) Answer…

- **A Contractor's 185-page Software Development Plan was reviewed and the comments were categorized as follows:**
  - **0** Critical
  - **92** Substantive
  - **10** Administrative
- **An assessment was made that the probability of program failure if they proceed development is p>50%. Which of the following statements is true?**

  A – This is definitely the correct probability

  B – The probability of failure is much higher, around 90%

  [✓ C] – The program might not fail at all

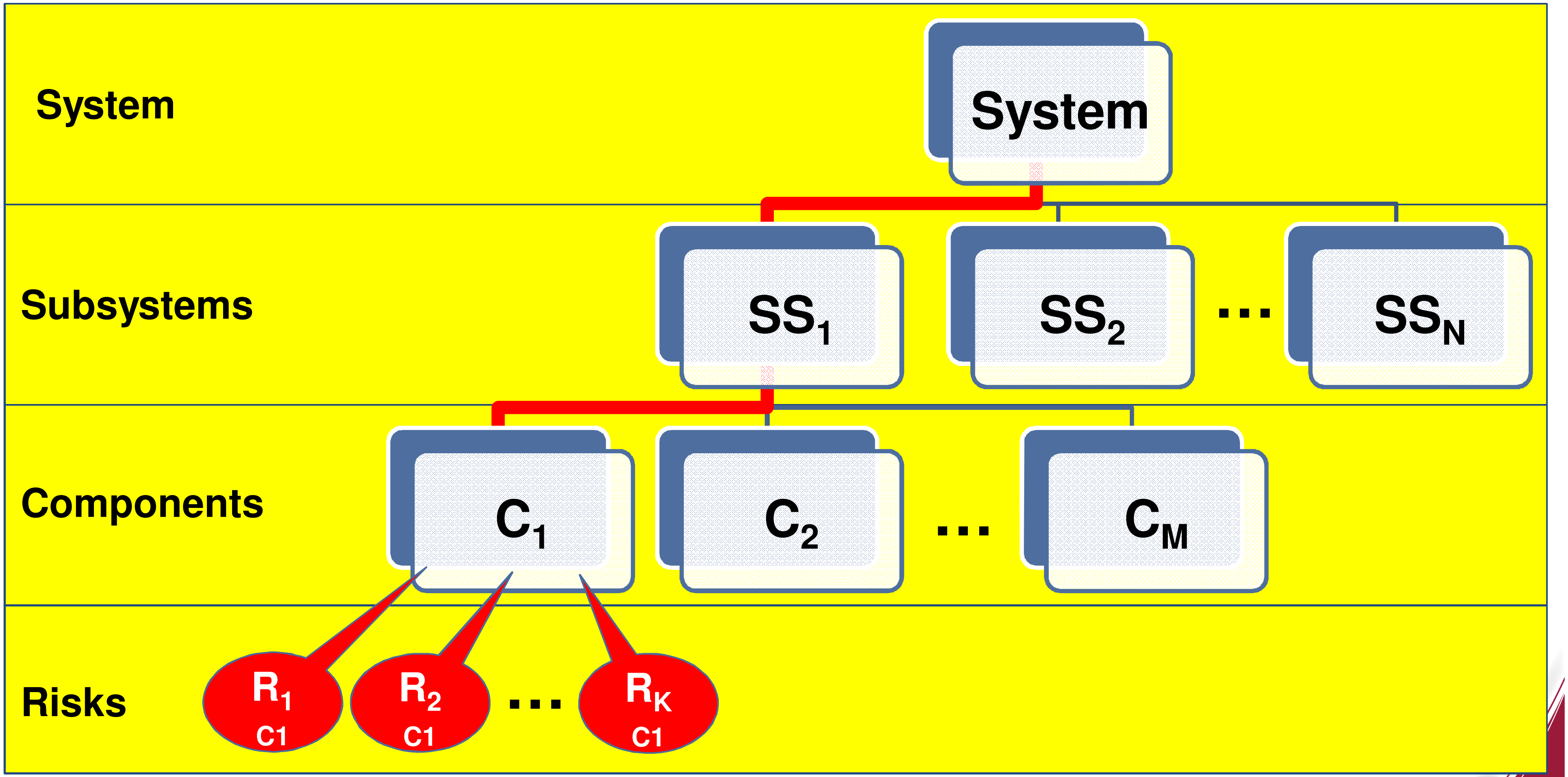
Despite of our high level of discomfort with this document, there is no way to accurately estimate the likelihood of program failure

Risk Roll-up (Risk Aggregation)

- **Decision makers may ask for the aggregated risk in a project**
  - Aggregation of risks is expected on two dimensions
    - Multiple risks associated with a particular system component
    - Along the Work Breakdown Structure (WBS) hierarchy dealing with risks of integrated system components
- **Risks can have two types of relationships to each other**
  - Independent
  - Inter-dependent

In this presentation we only discuss independent risks

Risk Aggregation in the WBS



K risks affecting a parent asset through its children assets

Aggregated Probability and Consequence

- **Aggregated probability formulas:**

$$P_A = 1 - \prod_{i=1}^{K} (1 - P_i)$$

$$S_A = \sum_{i=1}^{K} (P_i \cdot S_i)$$

Legend

$\mathbf{P_A}$	Aggregated probability of risk
K	Number of risks affecting asset
$\mathbf{P_i}$	Probability of i-th risk affecting asset
$\mathbf{S_A}$	Aggregated risk impact
$\mathbf{S_i}$	Impact of i-th risk affecting asset

- **Let's look at them just for two risks:**
  - $\mathbf{P_A} = 1 – (1 - P_1)(1 - P_2) = 1 – (1 – P_1 – P_2 + P_1P_2) = \mathbf{P_1 + P_2 – P_1P_2}$
    - This is essentially the sum of probabilities of either risks materializing separately minus the probability that they materialize at the same time
  - $\mathbf{S_A} = \mathbf{P_1S_1 + P_2S_2}$
    - This is essentially the probability-weighted sum of the individual impacts

How Useful is the Aggregated Risk?

- **Let's use some numbers to make the example easier**
  - – Risk_1 (R1) → Probability 0.2, Impact $100
  - – Risk_2 (R2) → Probability 0.002, Impact $100,000
  - – Aggregated probability 1 – (1 – 0.2) x (1-0.002) = 1 – (0.8 x 0.998) ≈ **0.2**
  - – Aggregated impact 0.2 x 100 + 0.002 x 100,000 = 20 + 200 = **$220**
- **What can we do with the probability-weighted impact figures?**
  - We can prioritize the risks and say that R2 should be addressed before R1
    - **This is a popular approach in creating "Top Ten" Risk lists**
  - However, beyond prioritization and comparison, the probability-weighted impact is not very useful and does not provide any help in answering some basic questions
    - How to prepare for these risks, what to expect if the risks materialize?
    - How much money should be spent on mitigation?
    - How quickly we need to complete the mitigation action(s)?

The Frequentist Background of Aggregated Risk

- **The Frequentist interpretation of risk**
  - If a mission were repeated a large number of times, the aggregated probability (**P_A**) would represent the per-mission rate by which any risk would actually materialize (Or, (**$1-P_A$**) represents the rate of successful missions)
  - The probability-weighted impact represents the average impact experienced over many repetitions of the mission
  - A blind, frequentist interpretation makes the decision-maker prone to the fallacy of the maturity of chances (Also known as the gambler's fallacy – a belief in the existence of a winning streak)
- **The risk of relying on aggregated risk**
  - Even if it is realistic to assume a large number of repetitions of a mission, due to the randomness of the risks, R2 might be realized way before R1, despite of the fact that R2's probability is much lower than R1's
  - When R2 is realized, the impact is $100,000, much larger than $220

Yet Another Trick (Although Simple) Question

- **The contractor reports that with a 0.001 probability the program is facing a risk that would have $1,000,000 impact. Which of the following statements is correct?**

  A – We will not need to reserve more than $1,000 for this risk

  B – The cost reserve must be kept under $1,000

  C – We might eventually spend much less than $1,000

The (Trick) Answer

- **The contractor reports that with a 0.001 probability the program is facing a risk that would have $1,000,000 impact. Which of the following statements is correct?**

A – We will not need to reserve more than $1,000 for this risk

B – The cost reserve must be kept under $1,000

[✓ C] – We might eventually spend much less than $1,000

The provided risk information (probability and impact) has no bearing on what mitigation might actually cost

Sample* Qualitative Rating Scheme to Estimate Software Risk Likelihood for a Single Risk

- **In case of software, a qualitative rating of risks is appropriate**
- **The following, 5-level rating scheme can be used:**

  (1) Remote

  (2) Unlikely

  (3) Likely

  (4) Highly likely

  (5) Near certain

* It is only a sample, because it may depend on the specific context

Sample* Qualitative Risk Consequence Rating

Level	Severity of impact	Impact on an essential, operational or mission capability	Impact on other capabilities or program execution
1	Minimal or no impact	-	-
2	Minor	Results in user or operator inconvenience but does not affect essential capabilities	Results in inconvenience for development or maintenance personnel but does not prevent the accomplishment of their responsibilities
3	Moderate	Adversely affects the accomplishment of an essential capability but a work-around is known	Adversely affects technical, cost, or schedule risks to the project or the life cycle support of the system but a work-around is known
4	Significant	Adversely affects the accomplishment of an essential capability and there is no work-around	Adversely affects technical, cost, or schedule risks to the project or the life cycle support of the system and there is no work-around
5	Severe	Prevents the accomplishment of an essential capability	Jeopardizes safety, security, or other requirement designated as "critical"

The idea is to offer a taxonomy of impact possibilities

* It is only a sample, because it may depend on the specific context

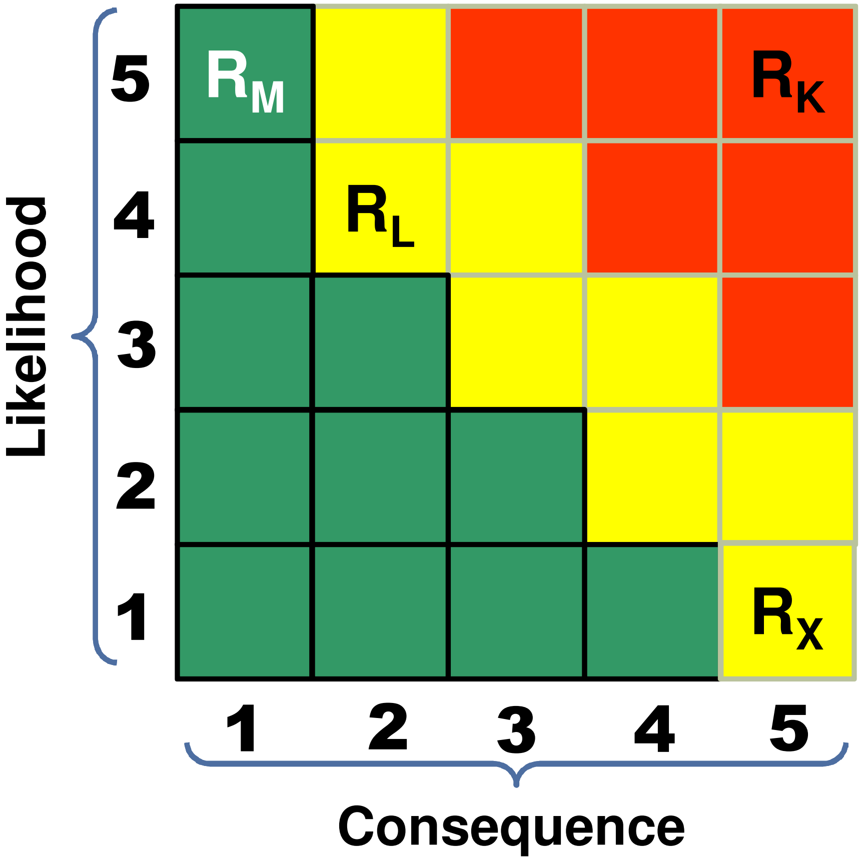
The Risk of Spurious Accuracy

- **First a joke***
  - A man was asked about the age of a certain river. His reply was that it was **3,000,004** years old. When asked how he could give such an accurate information, his answer was that four years ago, when he first did his research, the river’s age was given as three million years.
- **Do you think this is only a joke?**
  - A recent book on Agile Programming** is providing the following formulas to estimate Return on Investment (ROI) of Pair Programming on the basis of 29 data points:
    - Basis for the underlying size estimate: 10,000 Lines of Code (LOC)
      - (*Be wary of any LOC data if the programming language is not declared*)
    - Benefits: (10000/0.8507 + 33,3333*10*100)*100-265436 = **$4,243,397**
    - Costs: (10000/33.4044 + 2.355*10*100)*100 = **$265,436**
    - ROI = (4,243,397-265,436)/265,436*100 = **1,499%**

Don’t ask any questions, the book did not present more details either ☺

* [Campbell 2002], ** [Rico 2009]

Reporting Qualitative Assessment Results with a Risk Assessment Matrix*



Likelihood					
5	🟩 R_M	🟨	🟥	🟥	🟥 R_K
4	🟩	🟨 R_L	🟨	🟥	🟥
3	🟩	🟩	🟨	🟨	🟥
2	🟩	🟩	🟩	🟨	🟨
1	🟩	🟩	🟩	🟩	🟨 R_X
	1	2	3	4	5
	Consequence				

Qualitative Rating for the Severity of Risk Consequence

🟥 High (red)

🟨 Medium (yellow)

🟩 Low (green)

Examples:

R_K (5, 5) → high severity

R_L (4, 2) → medium severity

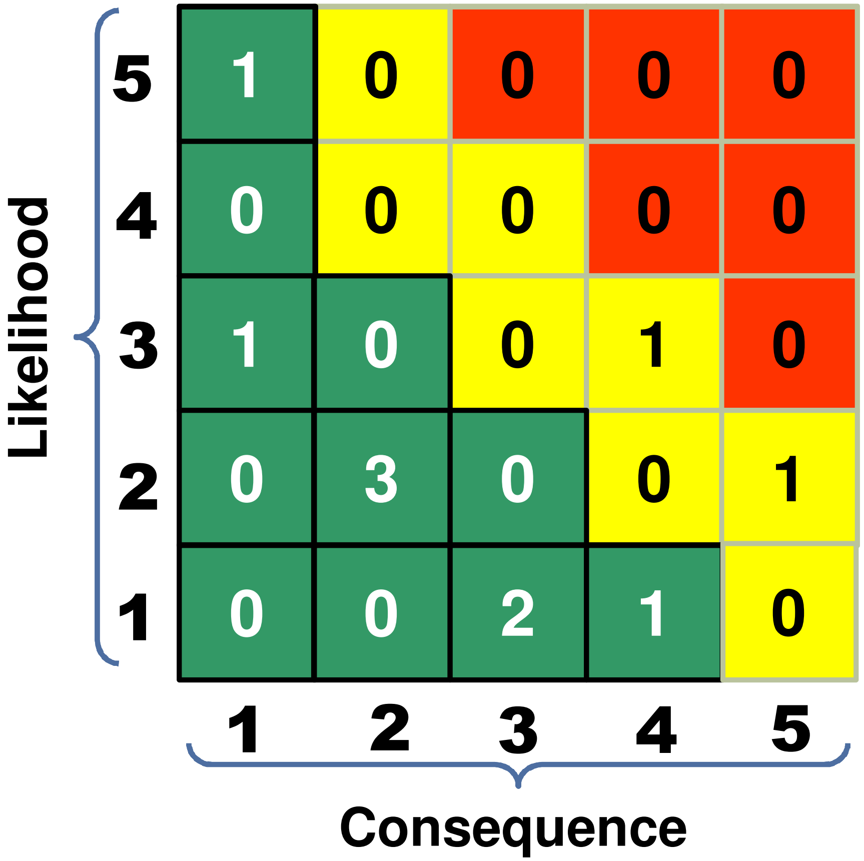
R_M (5,1) → low severity

R_X (1,5) → medium (not low) severity

Qualitative risk assessment results are rated on an ordinal scale

* Source: [DAU 2013]

Multiple Risks of a Component or Subsystem



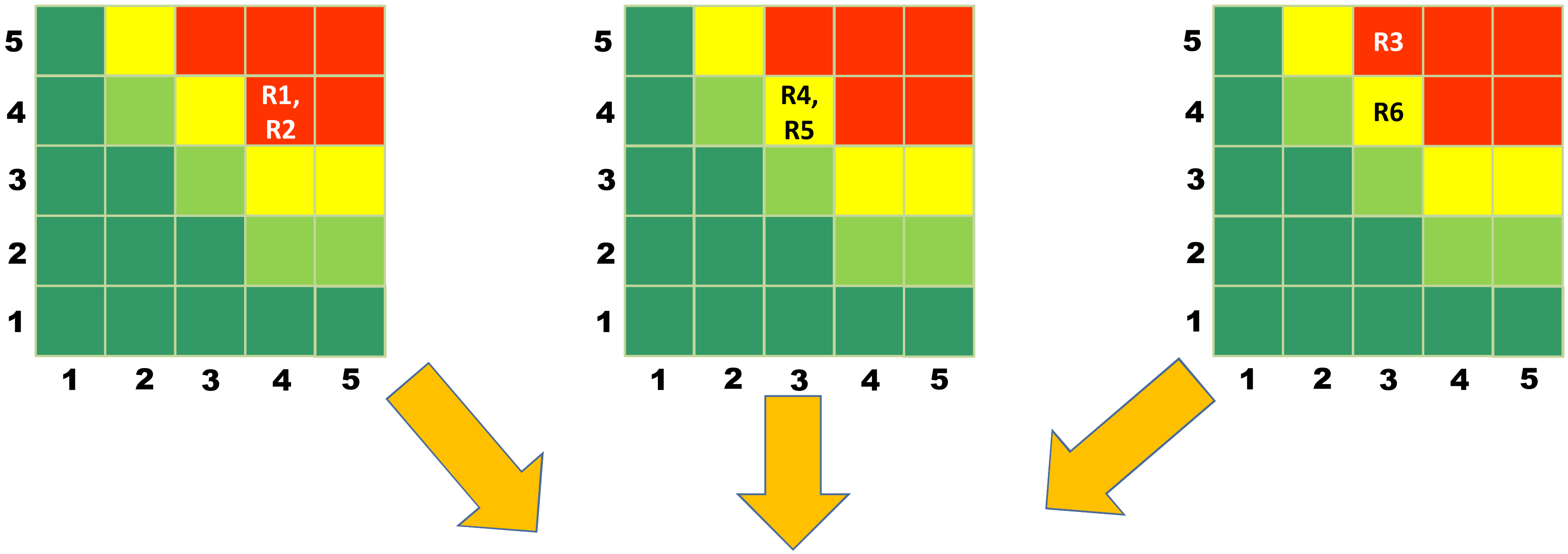
Likelihood	5	🟩 1	🟨 0	🟥 0	🟥 0	🟥 0
	4	🟩 0	🟨 0	🟨 0	🟥 0	🟥 0
	3	🟩 1	🟩 0	🟨 0	🟨 1	🟥 0
	2	🟩 0	🟩 3	🟩 0	🟨 0	🟨 1
	1	🟩 0	🟩 0	🟩 2	🟩 1	🟨 0
		1	2	3	4	5
		Consequence				

🟥 High	0
🟨 Medium	2
🟩 Low	8
Total # of Risks	10

Legend:

- Likelihood and consequence are rated on 1-5 ordinal scales
- Numbers in a cell representing the number of risks with the given likelihood and consequence

Risk Aggregation Example*



Risks in Component-1

Risks in Component-2

Risks in Component-3

High	3
Medium	3
Low-Medium	0
Low	0
Total # of Risks	6

*Also shows how the organization may introduce unique prioritization

Caveats of Ordinal Scales

- **It is a common mistake to attempt mathematical manipulation of ordinal (subjective) scales**
  - Multiplying ordinal likelihood with ordinal consequence values
  - Trying to roll-up multiple risks into a single measure
- **First let's see when ordinal scale is normalized to the upper value of actual consequence**

Risks	Expected Consequence [$]	Normalized to Upper Value (c)
$Risk_4$	100K	1.00
$Risk_3$	10K	0.10
$Risk_2$	5K	0.05
$Risk_1$	1K	0.01

Caveats of Ordinal Scales

- **It is a common mistake to attempt mathematical manipulation of ordinal (subjective) scales**
  - Multiplying ordinal likelihood with ordinal consequence values
  - Trying to roll-up multiple risks into a single measure
- **Now the ordinal scale is normalized to the upper value of the raw subjective scale**

Risks	Expected Consequence [$]		Raw Subjective Scale	Normalized to Upper Value (R)	
$Risk_4$	100K		4	1.00	
$Risk_3$	10K		3	0.75	
$Risk_2$	5K		2	0.50	
$Risk_1$	1K		1	0.25	

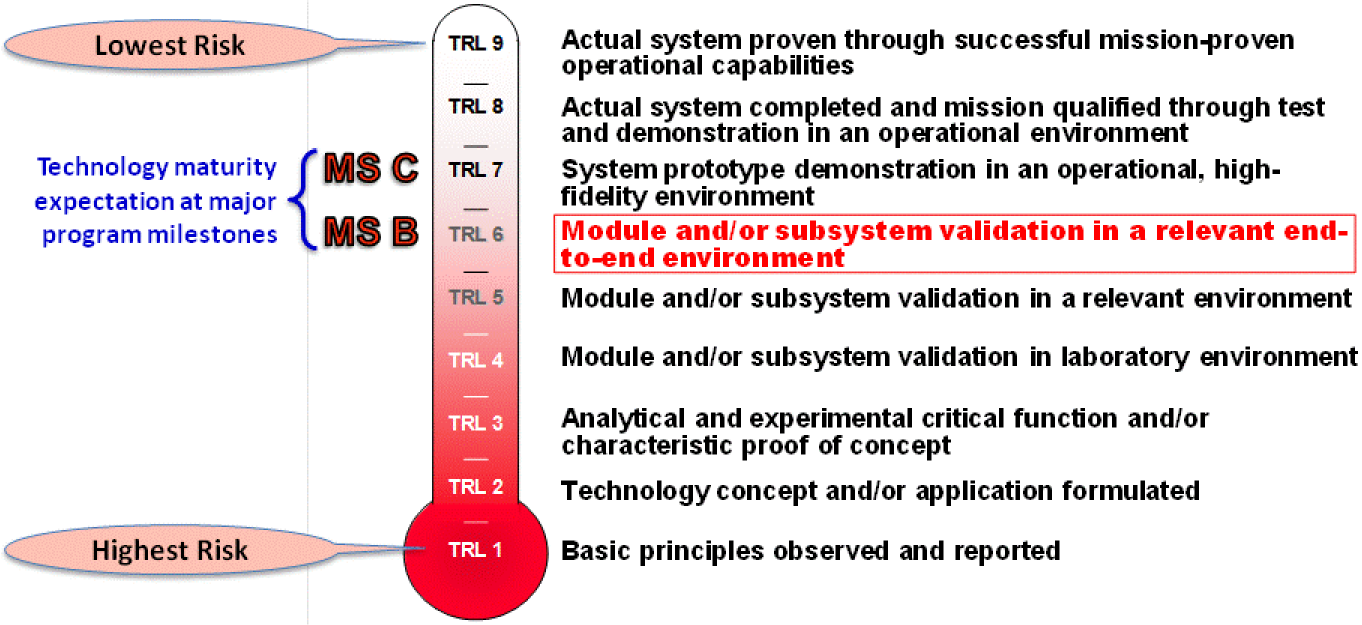
Caveats of Ordinal Scales

- **It is a common mistake to attempt mathematical manipulation of ordinal (subjective) scales**
  - Multiplying ordinal likelihood with ordinal consequence values
  - Trying to roll-up multiple risks into a single measure
- **Finally, the table shows the error when raw subjective scale values are used in such calculations**

Risks	Expected Consequence [$]	Normalized to Upper Value (C)	Raw Subjective Scale	Normalized to Upper Value (R)	% Error ((R – C)/C)*100
Risk_4	100K	1.00	4	1.00	0%
Risk_3	10K	0.10	3	0.75	650%
Risk_2	5K	0.05	2	0.50	900%
Risk_1	1K	0.01	1	0.25	2400%

Example: Algebraic Manipulation of TRLs

- **Technology Readiness Assessment (TRA)**
  - The TRA's objective is to identify technology risks and prevent programs from relying on immature technologies
  - The program's Critical Technology Elements (CTEs) are rated on a 1-9 Technology Readiness Level (TRL) scale



Lowest Risk

Highest Risk

Technology maturity expectation at major program milestones

MS C

MS B

TRL 9	Actual system proven through successful mission-proven operational capabilities
TRL 8	Actual system completed and mission qualified through test and demonstration in an operational environment
TRL 7	System prototype demonstration in an operational, high-fidelity environment
TRL 6	Module and/or subsystem validation in a relevant end-to-end environment
TRL 5	Module and/or subsystem validation in a relevant environment
TRL 4	Module and/or subsystem validation in laboratory environment
TRL 3	Analytical and experimental critical function and/or characteristic proof of concept
TRL 2	Technology concept and/or application formulated
TRL 1	Basic principles observed and reported

Warning: No separate impact and likelihood designation; the TRL is a single number on an ordinal scale

Example: Algebraic Manipulation of TRLs - 2

- **TRLs do not reflect the impact of not attaining the next level of maturity or the probability of attaining the next level of maturity**
  - TRL is a representative example for a discrete, subjective risk scale
  - It well demonstrates the reason why subjective scales are used, i.e., the data that would be needed for measuring the risk on a continuous scale is not attainable
- **In case of TRLs the temptation is particularly big for roll-up**
  - However, as it was pointed out, such manipulation yields bogus results
  - Example: Consider a program with two CTEs, where **TRL_1=4** and **TRL_2=8**
    - An attempt to average them yields a single number, **TRL=6**, which happens to be the acceptable value for passing **Milestone B** and starting the program
      - However, this single number would cover up the fact for the Milestone Decision Authority (MDA) that CTE_1 is immature and still has a long way to go to reach expected maturity. If CTE_1 is indeed “critical” then its immaturity can sink the whole program

Before We Wrap Up, a Difficult, Trick Question

- **Background**
  - Following the unsuccessful launch of Intelsat 27 on February 1, 2013 by Sea Launch, Boeing Co., one of the key, US partner in the service sued its Russian and Ukrainian partners for more than **$350M**
    - Boeing claims in the lawsuit that the partners agreed that if the endeavor fails, they would pay their fair share of reimbursement on the basis of their ownership percentage
  - Sea Launch track record:
    - 31 successful launches out of 35 (~ **89%** success rate)
      - However, one of the earlier, failed launches pushed Sea Launch into bankruptcy in 2009 and at that point the financing banks forced Boeing to pay out **$449M** (of which **$99M** was Boeing's fair share)
  - The (estimated) cost of the Boeing-built, lost satellite is ~ **$200M**
- **What seems to be an appropriate, up-front risk-mitigation cost reserve for Boeing?**

Not Really Answering Just Pondering this Question

- **Background**
  - – Following the unsuccessful launch of Intelsat 27 on February 1, 2013 by Sea Launch Boeing Co., one of the key, US partner in the service sued its Russian and Ukrainian partners for more than **$350M**
    - Boeing claims in the lawsuit that the partners agreed that if the endeavor fails, they would pay their fair share of reimbursement on the basis of their ownership percentage
  - – Sea Launch track record:
    - 31 successful launches out of 35 (~ **89%** success rate)
      - – However, one of the earlier, failed launches pushed Sea Launch into bankruptcy in 2009 and at that point the financing banks forced Boeing to pay out **$449M** (of which **$99M** was Boeing's fair share)
  - – The (assumed) cost of the Boeing-built, lost satellite is ~ **$200M**
- **What seems to be an appropriate, up-front risk-mitigation cost reserve for Boeing?**

The considerations would not strongly depend on the success rate

Instead of a Summary - A Final Joke

What is the Return on Risk Management?

After listening to an excruciating presentation on risk management, an executive asked the presenter what she could expect as a return on her risk management efforts. Without a pause the presenter replied, “Twelve-and-a-half to one.”

“How did you come up with that number?” the executive asked.

“Well, I simply took an average of two well-known cases,” the presenter explained. “A stitch in time saves nine,” and “An ounce of prevention is worth a pound of cure.”

Acronyms

Ah	Ampere hours
C	Component
CTE	Critical Technology Element
LEO	Low Earth Orbit
LOC	Lines of Code
MDA	Milestone Decision Authority
MS	Milestone
NiH_2	Nickel-Hydrogen
R	Risk
ROI	Return on Investment
SS	Subsystem
TRA	Technology Readiness Assessment
TRL	Technology Readiness Level

References

[Caldwell 1998] Caldwell, D., et al., *Development of the Large Diameter (5.5”) Nickel-Hydrogen Battery Cell*, Proceedings of the 5th European Space Power Conference, Tarragona, Spain, September 21-25, 1998

[Campbell 2002] Campbell, S. K., Flaws and Fallacies in Statistical Thinking, Dover Publications, 2002

[Chulani 1998] Chulani, S., et al., *Calibration Approach and Results of the COCOMO II Post Architecture Model,* 20th Annual Conference of the International Society of Parametric Analysts (ISPA,) June 1998

[DAU 2013] Defense Acquisition University website -> ACC Practice Center -> Risk Management, last accessed on February 19, 2013: https://acc.dau.mil/CommunityBrowser.aspx?id=38397

[Jorgensen 2002] Jorgensen, P., Software Testing: A Craftsman’s Approach, 2nd edition, CRC Press, Inc., 2002

[Rico 2009] Rico, D. F., et al., The Business Value of Agile Software Methods, J. Ross Publishing, 2009

Use of Trademarks, Service Marks and Trade Names

Use of any trademarks in this material is not intended in any way to infringe on the rights of the trademark holder. All trademarks, service marks, and trade names are the property of their respective owners.

AEROSPACE
Assuring Space Mission Success

Thank you.