MARY KAY O'CONNOR PROCESS SAFETY CENTER
TEXAS A&M ENGINEERING EXPERIMENT STATION

23rd Annual Process Safety International Symposium
October 20-21, 2020 | College Station, Texas

Does your facility have the flu? Use Bayes rule to treat the problem instead of the symptom

Keith Brumbaugh
aeSolutions
Millennium Tower, 10375 Richmond Ave, #800, Houston, TX 77042
Keith.brumbaugh@aesolns.com; keithbru@gmail.com

Introduction

Is our industry addressing the problems facing it today? We idealize infinitesimally small event rates for highly catastrophic hazards, yet are we any safer? Have we solved the world's problems? Layers of protection analysis (LOPA) drives hazardous event rates to 10^{-4} per year or less, yet industry is still experiencing several disastrous events per year.
If one estimates 3,000 operating units worldwide and industry experiences approximately 3 major incidents per year, the true industry accident rate is a staggering 3 / 3,000 per year (i.e. 10^{-3}). All the while our LOPA calculations are assuring us we have achieved an event rate of 10^{-6}. Something is not adding up! Rather than fussing over an unobtainable numbers game; wouldn't it be wiser to address protection layers which are operating below requirements? We are (hopefully) performing audits and assessments on our protection layers and generating findings. Why are we not focusing our efforts on the results of these findings? Instead we demand more bandages (protect layers) for amputated limbs (LOPA scenarios) instead of upgrading those bandages to tourniquets. Perhaps the dilemma is we cannot effectively prioritize our corrective actions based on findings. Likely we have too much information and the real problems are lost in the chaos. What if there was a way to decipher the information overload and visualize the impact of our short comings? Enter Bayes rule to provide a means to visualize findings through a protection layer health meter approach; to prioritize action items and staunch the bleeding.

Keywords:

Bayes, Bayes rule, Bayes theory, LOPA, IPL, SIS, SIF, SIL Calculations, systematic failure, human factors, human reliability, operations, maintenance, IEC 61511, ANSI/ISA 61511, hardware reliability, proven in use, confidence interval, credible range, safety lifecycle, functional safety assessment, FSA stage 4, health meter.

The State of Our Industry

The objectives of this paper are to look at some issues with the contemporary safety lifecycle industry and provide solutions. A major trend across industry has been an overall decrease of the tolerable catastrophic event likelihood (i.e. multiple fatalities) down to one such event every 100,000, or even 1,000,000 years[1]. This lowering of the target has the good intention of making a facility safer. After all, superb targets will make superb facilities, should it not? The downside of extravagant targets is they are harder to achieve. Realistically these targets are impossible to achieve once one considers the real uncertainties of physical systems (systematic error). Note that it is hard-enough meeting the targets already set, how will making the targets even smaller help matters?

Smaller tolerable risk targets will have the result of producing more Independent Protection Layers (IPL) with greater integrity requirements. This leads to a "Forest-for-the-tree syndrome" where a plant is trying to manage more IPLs than it can handle, missing the bigger picture of plant health and safety (e.g. key performance indicators). If every IPL has its own multifaceted maintenance and management requirements, how can a facility ever manage all responsibilities effectively?

Figure 1- Can't see the Forest for the Trees

The solution in this paper is simple, a facility should focus on managing what it is capable of managing. Minuscule targets have good intentions of making everyone safer by providing more protection but throw enough IPLs at a problem and one will soon reach the tipping point, the

[1] One event every 100,000 or 1,000,000 years is a target of 10^{-5} and 10^{-6} respectively.

straw that broke the camel's back. When everything is a problem due to much information nothing will be managed effectively. Furthermore, when there are fewer IPLs to distract a plant management team, the team is free to focus on the real problems. If one can identify problems, then effective management will occur. One of the best ways to identify problems is with a periodic health check of the IPLs. This paper will present one such health check called the "Bayes truth meter." The concept with the Bayes truth meter is to strip away all guess work and generic data, instead showing true IPL health reflecting a facility's own systematic biases.

To back up the claim that targets of 10^{-5} and 10^{-6} are unobtainable, consider figure 2 which is a list of current investigations from the Chemical Safety Board (CSB) over the course of the last year (circa August 2020).

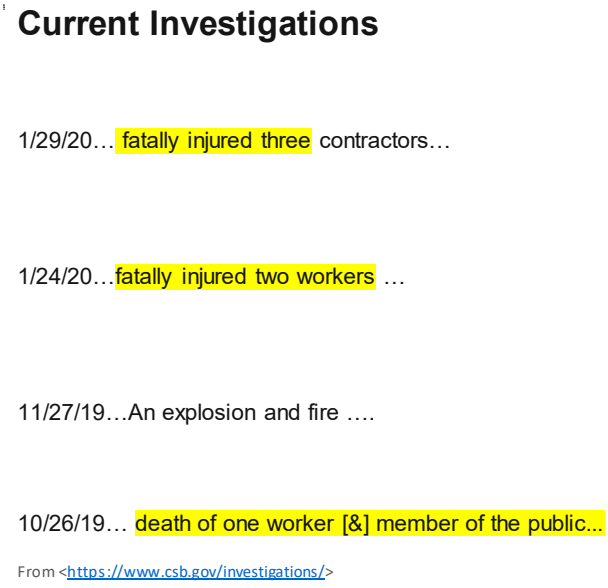

Figure 2- Chemical Safety Board list of open incident investigations for catastrophic events

There were three major multiple fatality accidents (and one near miss) over the course of one year. This sort of catastrophe would likely earn the maximum hazard mitigation target from any client following OSHA. If targets of 10^{-5} or 10^{-6} were achievable, at least 300,000 to 3,000,000 operating facilities would be needed to average out the three major accidents from year. Unfortunately, the current estimated number of petrochemical facilities in the United States is around 2,300 (keep in mind the CSB only covers US operations).

If the number of facilities is generously rounded up to 3,000 and divided by the three catastrophic accidents, the average industry catastrophic event rate is 10^{-3}, far short of the 10^{-5} or 10^{-6} targets! To put this into a visual form consider figure 3, pretend the 100 boxes represent the industry's 10^{-5} mitigation target.

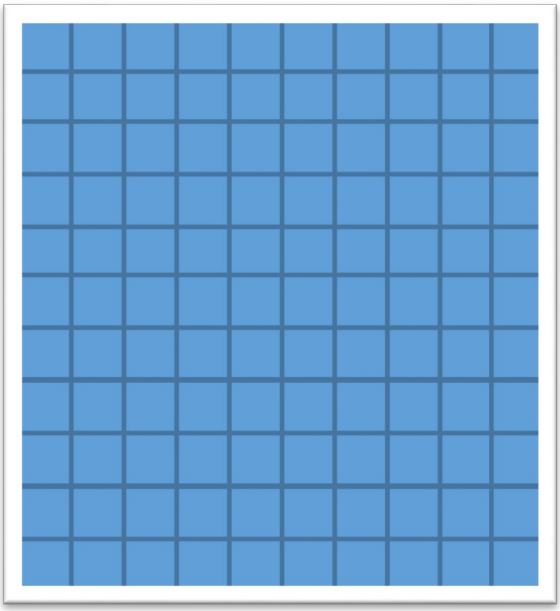

Figure 3- Industry target, 100 boxes represent a hypothetical 10^{-5} (10^{-6} would be 1,000 boxes!)

Figure 4 represents how far off the industry is from the mitigation target (10^{-3}):

Figure 4 - Reality, 1 box represents where the industry is, 10^{-3}! Compare to 10^{-5} above

As the graphics show the industry accident rate is off target by at least 2 orders of magnitude (3 orders if 10^{-6} is considered). This should be a wakeup call that the industry is missing something major. Something interesting to note is a 10^{-3} catastrophic event rate is the same as the lower bound of Human Error Probability (i.e. HEP). HEP is systematic error and is not typically considered in any calculations outside of Human Reliability Analysis (HRA)[2]. Maybe this is the factor the industry is are missing?

The Issues with our Current Approach

Returning to the ideas touched on previously, of increasing risk reduction targets to make everything "safer." There are good intentions behind ultra-low targets, however meeting these obscene targets would require throwing every feasible protection layer that can be mustered at the problem, hoping that something "sticks" (i.e. is effective). This leads to safety barrier overload. The idea is when there are too many things to manage, and everything is important, how can the signal be separated from the noise. If everything is important, what is *drastically*

[2] For more information on HRA and Human Factors consider the following two papers:
Conducting a Human Reliability Assessment to support PHA and LOPA, Dave Grattan - https://www.aesolns.com/post/conducting-a-human-reliability-assessment-to-support-pha-and-lopa

Can we achieve Safety Integrity Level 3 (SIL 3) without analyzing Human Factors?, Keith Brumbaugh and Dave Grattan - https://www.aesolns.com/post/can-we-achieve-safety-integrity-level-3-sil-3-without-analyzing-human-factors

important? What is behaving fine and doesn't need attention? What is behaving fine right now but might be a problem a few years down the line? What has no chance of working when demanded, putting the facility at risk right now, and is going to drive the system over the edge of a cliff?
When there are too many protection layers to manage due to astronomical LOPA targets, no one will know which warning signs are important and which are just a nuisance. Protection layer overload leads to a "forest-for-the-trees-syndrome." As an example, the most sophisticated protection layers designed by top dollar consulting companies will be all for naught once they fall out of maintenance. How good is the gold plated protection layer when the valve has polymerized stuck due to never being tested? What might lead to maintenance oversights? Perhaps there were too many protection layers with not enough manpower to manage them.

Another problem with astronomical protection layer targets is Safety Instrumented Functions (SIFs) will need to be applied with high safety integrity calculations to meet a target. These calculations will "prove" a protection layer is good enough to close a 1 in 10,000 year gap, yet is that number real? Theoretically a Safety Integrity Level (SIL) calculation is correct if we consider hardware failures alone and the system operates in a vacuum, yet as soon as a human touches the system good luck maintaining that integrity level without highly sophisticated management practices. And the problem only gets worse as more high integrity protection layers are added to the facility.

All of this naysaying may seem to be blasphemous coming from safety system engineer, implying lofty risk reduction targets cannot be met, but has reality been considered? As previously mentioned, the industry catastrophe rate is sitting around the lower bound of Human Error probability (10^{-3}). This is key to understanding what has not been addressed in traditional LOPA math and SIL calculations. It is the authors' opinion that some very major degradation factors are being missed when modelling protection layer integrity. The elephant in the room is systematic error.

This is all not to say that the industry is in a bad state. There are a lot of good people out there doing important work, trying to make everyone safer. Their contributions should not be discounted as they are all based on lessons learned in blood. Yet it feels like the industry has gotten as far as it can with its current practices, floundering between moving forward or backwards depending who is asked. The next step forward in process safety needs to be in the best direction possible.

Do it Better with Bayes

The Bayesian approach allows matching optimistic rare event assumptions and IPLs with real-world observations, turning fantasy into reality. This approach allows one to base plant health metrics on observed evidence. Otherwise the industry is stuck with using generic data which is not specific to the facility's own systematic biases. If typical SIL calculation modelling data is based on industry averages, figure 4 shows how good industry average is.

A Bayes approach will likely show a facility isn't as good as it hoped it was. The problem is inductive reasoning has been used to predict catastrophic rare events. This is like the black swan

theory, historically the world used to think all swans were white since a black swan had never been observed in nature, but then low and behold, they were discovered eventually. Bayes rule would have allowed factoring in the possibility of a black swan occurring. A black swan would always be in the realm of possibilities, as more evidence was gathered such as feathers, third party sighting reports, occurrences in other similar species, etc; the model could have been updated to better predict where reality laid. The traditional model would have said "There has been millions to billions of sightings of white swans, no black swan has ever been seen in nature, therefore there is no black swan."

Back to the process safety industry, one can make a similar comparison between the current industry and a Bayesian approach. The current industry approach is based on frequentist-based statistics. This approach requires enormous amounts of data in order to derive a conclusion, such as the millions to billions of white swans and no black swans. If the analogy is given little thought, all that is known with certainty is there is a good chance of operating in a safe state, but there is no idea of the dangerous state; how bad is it, how it would unfold, and how likely it is. The only way to know the answer to the dangerous state questions is to collect data (which is likely a trick that can only be performed once). A frequentist approach requires enormous amounts of data to definitively state a comparative frequency. It should be obvious that a facility will not have, nor will it ever want enormous amounts of data for a rare catastrophic event. Contrast the frequentist approach with the Bayesian approach. The Bayesian approach allows for the input of subjective data in a logical manner. This method allows *all* relevant evidence to be factored into the model. To again use the black swan analogy; things like feathers, third person accounts, and occurrences in similar species can be directly related to near misses, audit results, and similar process accidents. Bayes allows one to factor in systematic biases and errors. Frequentist methods cannot do this.

Bayes rule can answer the question, "is a catastrophic event rate of 10^{-6} obtainable." Most likely if Bayes rule was embraced, the results would show that the industry is aiming for something that is unachievable, "biting off more than one can chew."

When Bayes shows that 10^{-6} can't be met, a facility will need to step back and ask, "what are we really trying to do here." The answer that makes the most sense is the facility is trying to make the most money while not "going boom." Since the facility's resources (time and money) are limited, then the "not going boom" part needs to be focused on the systems that need the most help while not spending all the surplus resources. Bayesian methods can provide an outlook on how each individual protection layer is behaving. Advanced warnings can be given based on evidence, staunching the bleeding of a bad acting barrier.

How to Apply Bayes Rule to Process Safety?

Implementing Bayes in process safety can be as simple or as difficult as one cares to make it. The author's previous paper[3] went over a simple approach to implement Bayesian Methods into the management of Safety Instrumented Functions. It is suggested to review the referenced paper

[3] *What is Truth – Do our SIL calculations reflect reality*, Keith Brumbaugh 2019, https://www.aesolns.com/post/what-is-truth

for further details as well as a rough “how to” example. The approach does not need to be limited to SIFs; all protection layers are ripe for a Bayesian management approach.

Begin a Bayes model with any protection layer, with any theoretical achieved Probability of Failure on Demand (PFD). Convert the achieved PFD point value, such as 0.01, into a probability distribution (Poisson or example). If the distribution is known this would be preferable, but often times the distribution is not known.

The probability distribution should represent all of the possible PFD the protection layer could hold. The boundary of the distribution should contain all realistic values which the protection layer could ever be. For example, it might be expected a SIF can operate somewhere between a PFD 0.1 to 0.0001. The probability distribution also assigns the likelihood that the protection layer *is* any of the particular PFD values.

The initial probability distribution is known as the prior. The Bayesian prior distribution is then updated over time with new evidence to form a posterior. Below is a conceptualized representation of a SIF which has undergone a Bayesian conversion and updating process.

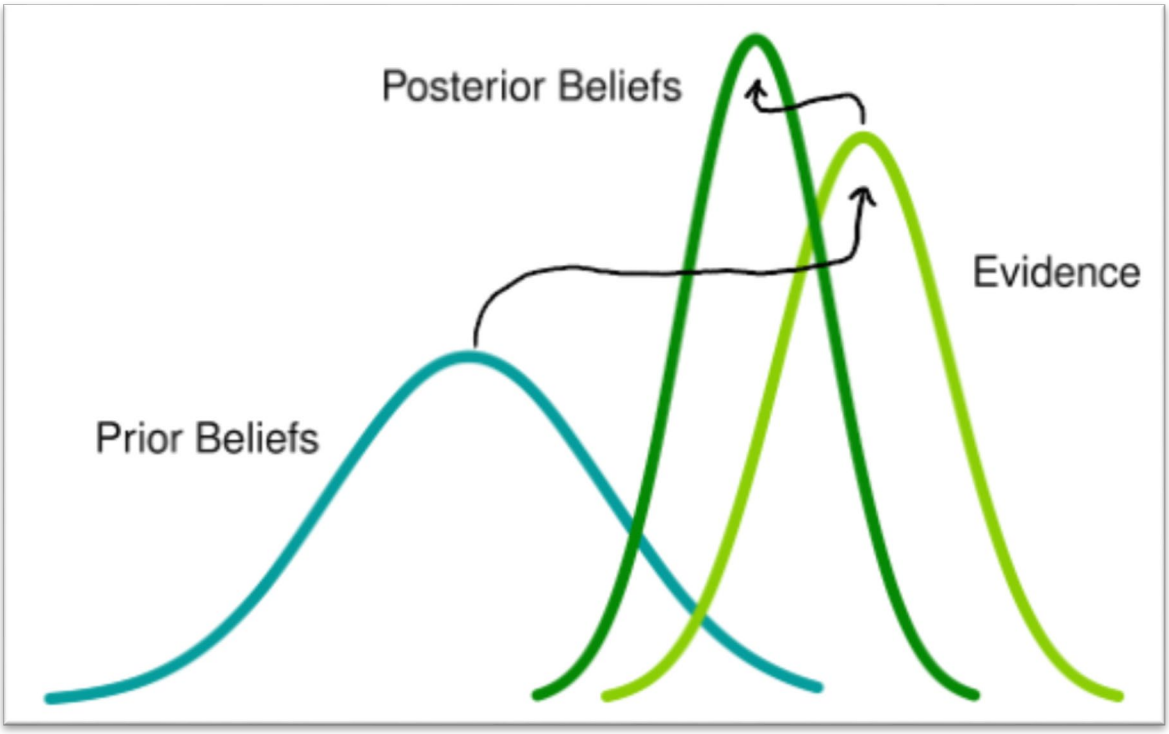

Figure 5- Example of probability distribution with prior, evidence, and posterior

There are two types of evidential data that can be used, the first is quantitative data. Quantitative data can be absolutely proven as true or false (a Bernoulli trial). The protection layer is subject to a test and the result is recorded. Items that fall under this category are proof test results and actual demands on the system (planned or unplanned).

The other type of evidential data is qualitative. Qualitative data is based upon expert opinion. The application of qualitative data may seem subjective, but the precedent of using qualitative data has already been set for process safety. Today it is accepted industry practice to use subjective judgements in LOPA, don't forget that LOPA drives everything (most of the time)! If qualitative judgements are documented and justified, the industry has no problem with them. A qualitative Bayesian update can fall under the same scrutiny. So long as the application of a qualitative Bayesian update is made to be repeatable and predicable there should be no issues. This can be achieved with a repeatable checklist from a common assessment task. Data which fits the qualitative bill are audits, Functional Safety Assessments (FSA), and Human Reliability Analysis. All of this data is aimed at discovering systematic errors by a repeatable and well-established practices.

The probability distribution from figure 5 can also be represented as a cumulative distribution. The data source of a cumulative distribution is the exact same as a probability distribution, the difference is the likelihoods are summed from 0% to 100%.

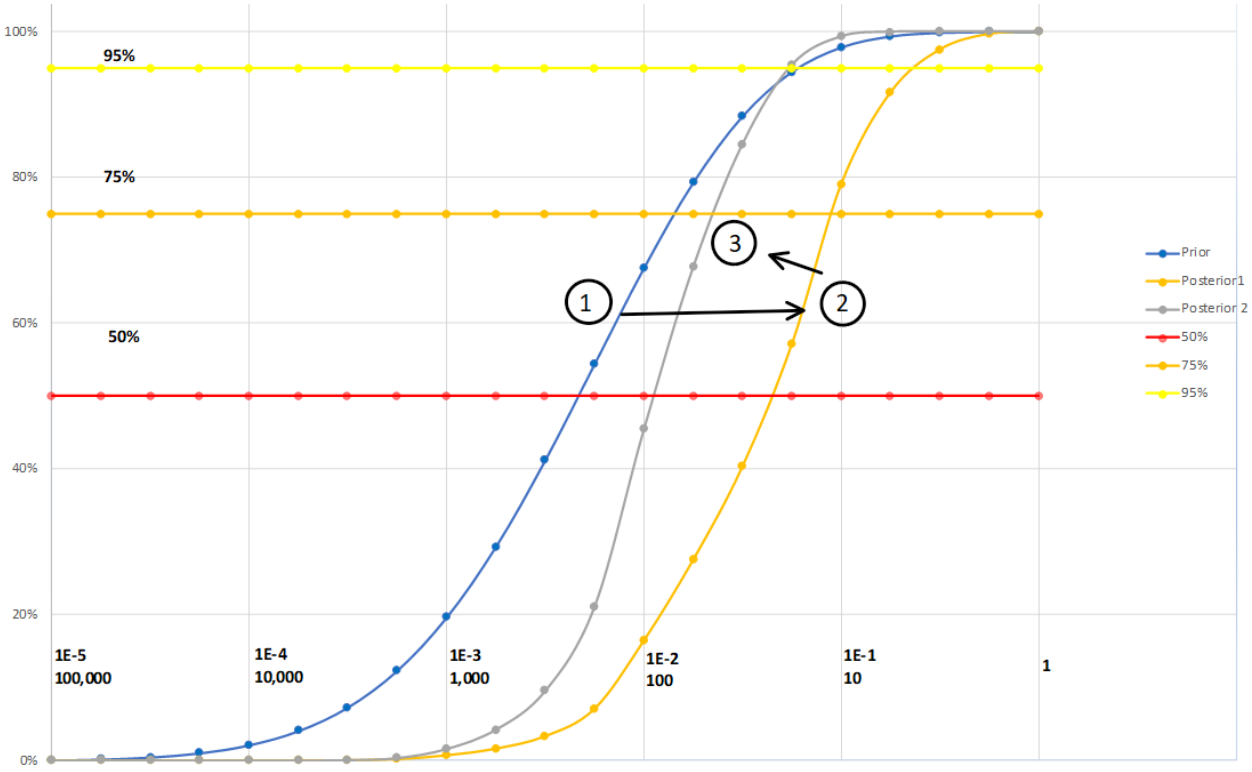

Figure 6- A cumulative distribution of a Prior, and 2 subseqeunt posterior updates

Figure 6 is a cumulative distribution of the probability distribution from figure 5. In figure 6, the X-axis represents an IPLs probability of failure on demand. The X-axis is unitless, representing either PFD on the top, or Risk Reduction Factor (RRF) on the bottom (inverse of PFD). Each vertical line represents one order of magnitude with 4 gradients within (logarithmic PFD). For example, looking at the far-right side of the graph, the four different gradients from right to left

represent 2.5, 5, 7.5, up to 10 RRF. The important take away is the further to the left the point of interest, the better (i.e. smaller PFD).

The Y-axis represents the cumulative likelihood from 0% to 100%. The 3 colored lines at 95%, 75%, and 50% are credibility levels. The point where these credibility level lines intersect the cumulative distribution curve represents the upper credible interval, which says an IPL is a particular PFD value or better. Note that these credibility levels are arbitrary, however they align with examples from the IEC 61511 standard.[4] As an example, the curve labelled "1" intersects the 95% credibility limit around 25 RRF. With this intersection, a statement can be made that it is 95% credible that the IPL represented by distribution "1" is 25 RRF or better.

It might be apparent that as better RRF numbers are targeted, the credibility decreases. For example, on the same "1" curve there is only a 50% credibility the IPL is 125 RRF or better. This is the key concept when trying to determine how "good" an IPL is in a Bayesian system, enabling one to state the credibility an IPL is meeting a certain performance target.

With the basic concepts of credibility and probability addressed, this example can move onto the concept of a Bayesian update. In figure 7, the curves "1," "2," and "3" represent a theoretical Safety Instrumented Function. The system starts with a SIL calculation result converted to a distribution as seen in curve "1." The SIL calculation using traditional methods returned a PFD of $1.3x10^{-2}$ (77 RRF), this achieved RRF value seeded a Poisson distribution to intersect 77 RRF at the 75% credibility level.

Next, pretend there is a failure during the first proof test. This is a simple Bayesian update with quantitative data represented in curve "2", 1 test, 1 failure. Curve "2" has shifted to the right, a worse result. Now the 75% upper credibility limit is around 12 RRF. Once this warning sign is discovered from the Bayesian update, pretend the management team initiates a root cause analysis, identifies the problem, fixes the problem, then performs a Management of Change Functional Safety Assessment with a positive assessment result. This positive assessment is a qualitative update and is applied as curve "3." Observe curve "3" has shifted back to the left, a better result. As the example ends the system is better than after the failed test of "2," but not as good as it originally started in curve "1."

Making Bayes more Intuitive

The graphs in the previous section were full of details but unfortunately, they are not intuitively obvious. In order to make the concept of a Bayesian update more intuitive it would be beneficial to simplify the information into what is important. After all, management of IPLs, and addressing bad actors is the most valuable application of Bayes. This paper introduces a concept called the "Bayes Truth Meter." Imagine that a corporate criterion will accept an IPL upper credibility limit of at least with 50% credibility, preferably 75%, and over achievement at 95% (see IEC 61511-2018, Part 2, Figure A.7 as basis for the levels). Stripping away all of the distribution "mumbo jumbo," the IPL example of the Prior from Figure 6 (i.e. curve "1") is shown in Figure 7 converted to the Bayes Truth Meter.

[4] IEC 61511-2018 - Part 2, Figure A.7 – Typical probabilistic distribution target results […]

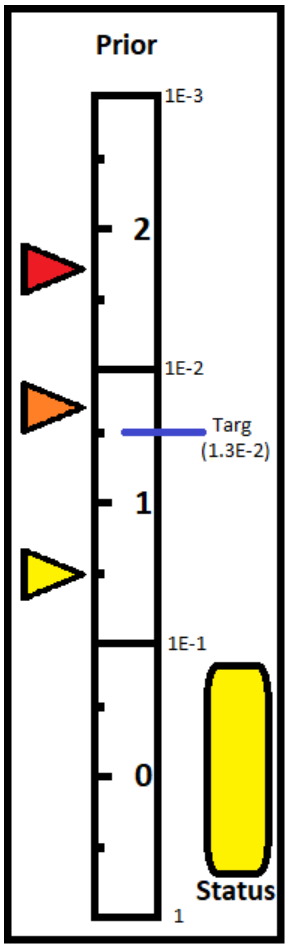

Figure 7- Bayes Truth Meter. Prior "1" from Figure 6.

To quickly describe the meter, the red, orange, and yellow pointers show where the 50%, 75%, and 95% upper credibility limits cross the cumulative distribution curve.

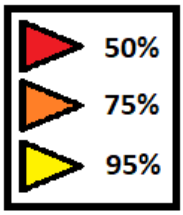

Figure 8 - Credibility markers (key)

The blue bar shows the PFD target (recall the previous example was a SIF with a target of $1.3x10^{-2}$, i.e. 77 RRF).

There is a "Status" button in the lower right corner to quickly tell how an IPL is operating in relation to its target. Red is bad, Orange is Ok, Yellow is good, and Green is Great.

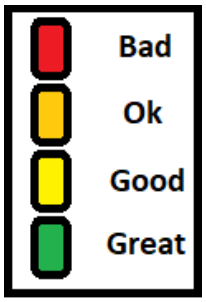

Figure 9 - Status lights (key)

The status light changes based on where an IPLs target (the blue line) lies in relation to its upper credibility levels. For the meter in Figure 7, the 75% credibility marker is better than the target, so Status light is yellow (good).

Following the previous example, the system encounters its first Bayesian update, a failed test.

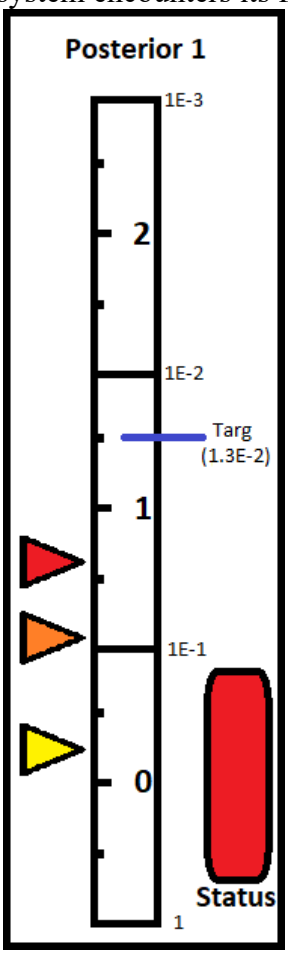

Figure 10 - Bayes Truth Meter. Posterior 1 (i.e. curve "2" from Figure 6).

Figure 10 depicts a failure during the first proof test. The meter shows the target is worse than even the 50% credibility marker. It is not depicted in the meter, but the target is only 20% credible! This poor result has put the status light in the “Bad” zone. Management would know

that it is time to focus attention on this SIF before a problem develops, and at least try to recover to the “OK” status.

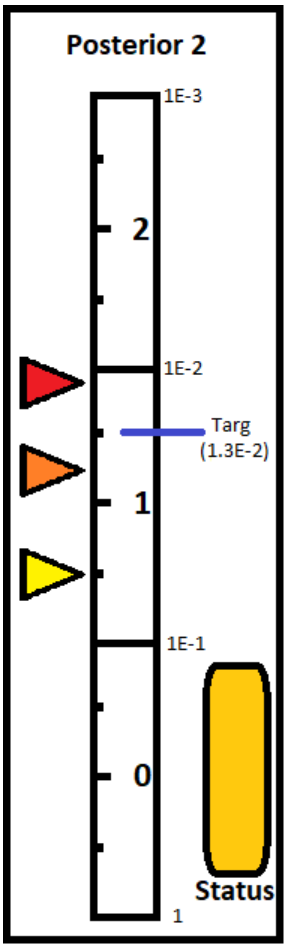

Figure 11 - Bayes Truth Meter. Posterior 2 (i.e. curve "3" from Figure 6).

The final update shown in figure 11 represents the root cause analysis was performed to determine the source of the failure, the cause was fixed, and then an MOC functional safety assessment was run with favorable results. This subjective judgement updates the meter from Figure 10 and shows the system has recovered to the “OK” Status. It is not back to “Good,” but there are likely more important issues demanding attention at this point.

Does Bayes Prove we are Aiming too High?

Returning to a previous point made in this paper, once a Bayes engine has been implemented, the difficulty in achieving the lofty targets set by current industry practice will become apparent (10^{-5} and 10^{-6}). To prove the point, consider the same trial run previously in figures 7 through 11, but with a SIL 2 SIF instead of a SIL 1.

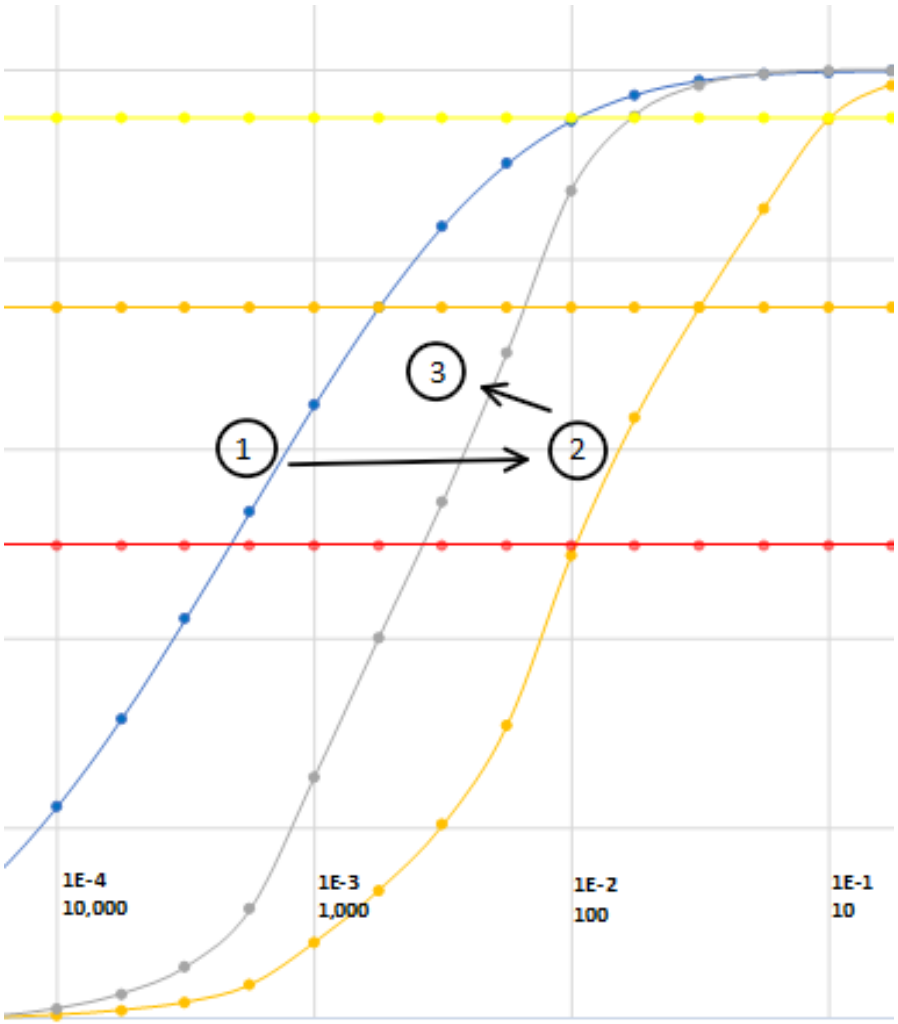

Figure 12 - SIL 2 cumulative distribution with 1 prior and 2 posterior updates

The example in figure 11 is a SIL 2 SIF with near identical parameters as the previous SIL 1 SIF an order of magnitude greater. This example has seeded the prior target of $1.3x10^{-3}$ at the 75% upper credibility limit. The poor results are shown in this table, but the Bayes truth meter makes it easier to understand.

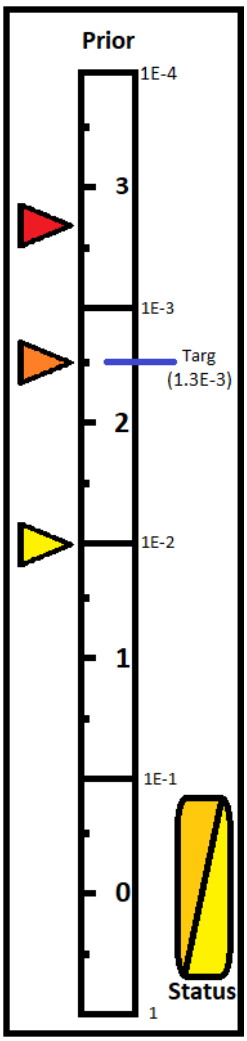

Figure 13 - Bayes Truth Meter. SIL 2 Prior "1" from Figure 12

Figure 13 shows the prior from figure 12 (plot 1) converted to the Bayes truth meter. The status light indicates the system is on the line between “Good” and “Ok” (i.e. the target PFD is at the 75% credibility marker). Next the system is subjected to the same one test, one failure.

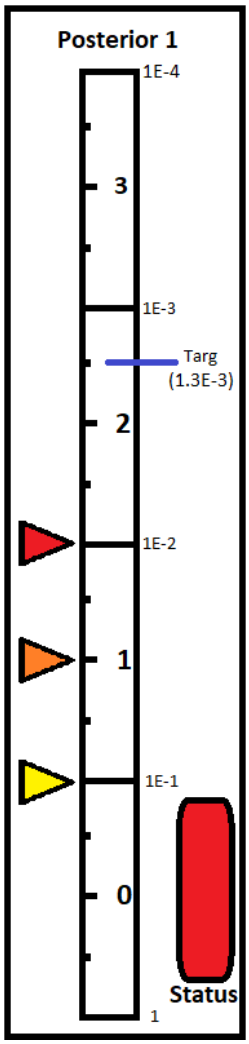

Figure 14 - Bayes Truth Meter. SIL 2 Posterior 1 (i.e. curve "2" from Figure 12).

It can be seen just as in in the previous SIL 1 example (figure 10), the system has dropped significantly. Notice however, this drop is more drastic. Like last time the target has a very low credibility, in fact it is only 13% credible that the SIF is operating at the target. Next a similar root cause analysis with a recovery factor is applied to the SIF.

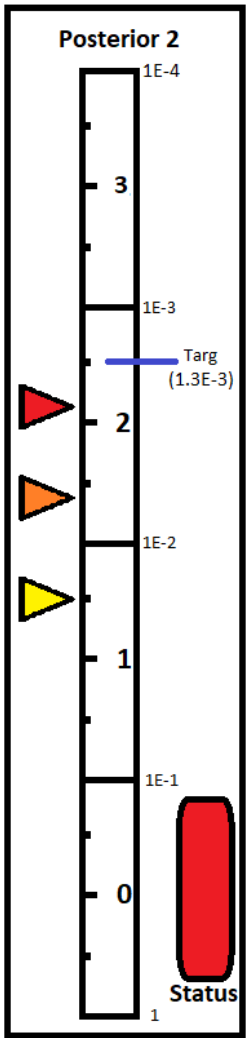

Figure 15- Bayes Truth Meter. SIL 2 Posterior 2 (i.e. curve "3" from Figure 12).

Unfortunately, even after a recovery factor has been applied the SIF is still in the “Bad” zone. It may seem like not enough recovery was applied, however the recovery factor applied was an entire order of magnitude greater than the SIL 1 example (due to the SIL target was an order of magnitude greater). In fact the recovery may have been granted too much weight (recovery factors are also subjective).

This might seem unfair but to put everything into context, consider the target again. The target is around one failure every 1000 years, however in just 5 years there was a failure. How much evidence would it take to convince someone that the function has been fixed and is operating back at the 1 in 1000 level?

Consider hurricane Harvey. That hurricane was a 1 in 1000 level event, yet Houston, Texas experienced this major hurricane a few years back. The city of Houston has implemented new safety measures to help combat any future flooding events, but would anyone living in Houston today claim that there will never be another hurricane Harvey in their lifetime? The answer is

most likely no, and every hurricane season for the next 20+ years the entire city will be on full alert (until the next generation comes along, thinking they know better than their elders).

Conclusion

It is the authors' belief that if the industry started to implement Bayes into its models it will quickly be demonstrated that lofty 10^{-5} and 10^{-6} event frequency targets can never be met. As witnessed in the SIL 2 SIF example, just one failure at any time during a facility's operating history will quickly shatter the illusion that a SIF can reach SIL 2 (not to mention SIL 3). Imagine that a target of 10^{-6} would require at least three IPLs of this same magnitude to mitigate the target. Good luck!

When a facility pretends it can meet 10^{-6} it is ignoring the elephant in the room, systematic errors and common cause failures. These failures are real, but their impacts are largely unknown. A Bayesian model can prove that they are worse than the industry gives credit.

If the industry were to acknowledge that 10^{-6} isn't possible, then what is possible? Back on figure 4 it was shown the industry is operating around a 10^{-3} catastrophic event rate on average, but that doesn't seem like a good target. To compare the process safety industry to the airline industry, the approximate probability of dying in an airplane crash is also 10^{-5} (see footnote [5]). Keep in mind the airline industry has much simpler systems designed for one purpose only, yet still has multiple fatality accidents. It might be best to split the difference between where the process safety industry is, and where the airline industry is. This admits the difficulties due to the complexity of process safety systems, realizing that process safety can never be as simple as an airplane's safety system.

If the industry were able to accept a lower target it would be much easier to close a LOPA gap. Lower targets would equate to few IPLs to manage. Keep in mind also that these IPLs would show a realistic number based on the Bayesian model, updated with real evidence. Less IPLs would lead to more effective management of IPLs and greater ease on maintenance. The Bayes Truth Meter approach allows a plant management to focus on bad actors. Finally, with less "trees" (IPLs) to manage, a facility is free to focus on the "forest" as a whole (overall plant health).

[5] In the year 2019 there were 10 major airline crashes with multiple fatalities. There are approximately 18 million flights per year on average. The odds of dying in a plane crash is around 10^{-5} to 10^{-6}. https://www.1001crash.com/. Airline systems are very sophisticated, and have one goal in mind, technologies are mostly the same, and failure modes are well understood. Compare to the process safety industry.

Figure 16 - Less trees = easier to manage the forest