

Title: Year 2000 Confidence Assessment - A Bayesian Tutorial

Author: John A. Vitkevich, Jr.

Date: 22 December 1997

For many computer systems, the goal is to survive the Year 2000 (Y2K) rollover by finding all the critical errors that can bring a system down or otherwise impair mission operations. A simple application of Bayes' theorem provides a method for assessing the confidence of surviving the Y2K computer date problem. Introduced into statistical analysis more than 200 years ago, Bayes' theorem provides formulas for adjusting prior knowledge of probabilities to reflect the impact of more than one condition being present. The Bayesian formulas are valid mathematical expressions of conditional probability theory, but many students of introductory probability and statistics often find that the application of Bayes' theorem is hard to grasp. A geometric representation of Bayes' theorem can be helpful.

With Bayes' theorem, it is possible at the end of a given "resolution cycle" to estimate the probability of four types of outcomes, which are also called posterior probabilities, as illustrated in Figure 1. A resolution cycle is defined here to mean one of several iterations attempting to find, correct, and replace those parts of the system that are infected with critical Y2K errors. The posterior probability depends on knowing estimates of three things (also called *a priori* estimates or prior probabilities), namely:

- estimated prior probability that an error is present
- estimated likelihood that an error will be detected in the presence of an actual error
- estimated likelihood that an error will be detected in the absence of an actual error

The outcome is a true positive or "good hit" when an error is detected and the error is actually there. There will be a "false alarm" (also called false positive) when an error is detected, and no errors are actually present. An "undetected error" or false negative will occur when an error is not detected, but an error is actually present. Lastly, there are no errors (i.e., "no news is good news" or true negative) when an error is not detected and no errors are actually present.

The Y2K problem is successfully resolved for a given system when the number of remaining undetected critical errors is reduced to zero before the Y2K rollover occurs. In actual practice, it might take several resolution cycles to reduce the probability of undetected errors to an acceptably low level. Knowledge of the number of errors corrected during a previous resolution cycle can be used to revise the prior probability estimate for applying Bayes' theorem to the follow-on resolution cycle.

Figure 2 shows how the four possible outcomes that can be estimated by Bayes' theorem can be presented geometrically by four distinct rectangular areas, in which the sum of the areas equals one.

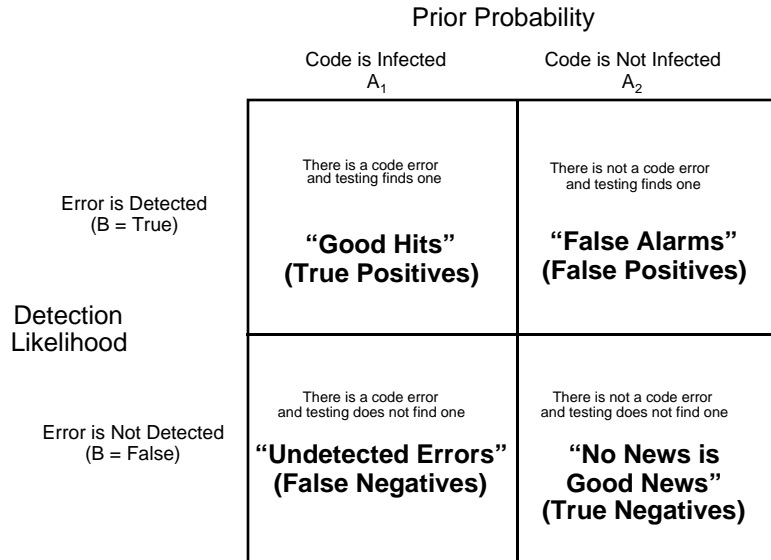


Figure 1. Possible Outcomes Estimated by Bayes' Theorem

The area of each rectangle reflects the absolute probability of the corresponding outcome. The areas of the two upper rectangles in this diagram (above the dark line) are proportional to the probability that an error is detected. The areas of the two lower rectangles (below the dark line) are proportional to the probability that an error is not detected.

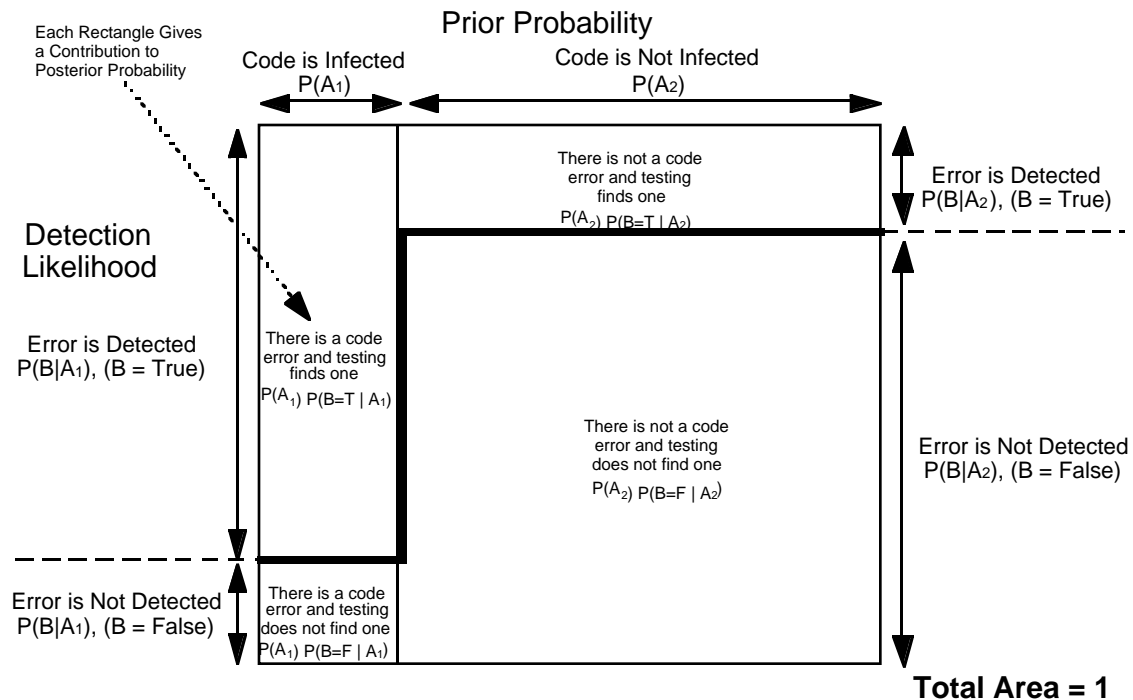


Figure 2. Geometrical Representation of Bayes' Theorem

The areas of the two rectangles on the left side are proportional to the probability that an error is actually present. The areas of the two rectangles on the right side are proportional to

the probability that an error is not actually present. The area of each rectangle can be calculated with the formulas given in Table 1.

<u>Outcome</u>	<u>Description</u>	<u>Bayesian Probability Formula</u>
There is a code error and testing finds one	Good Hits (True Positives)	$P(A_1 B=T) = \frac{P(A_1) P(B=T A_1)}{P(A_1) P(B=T A_1) + P(A_2) P(B=T A_2)}$
There is not a code error and testing finds one	False Alarms (False Positives)	$P(A_2 B=T) = \frac{P(A_2) P(B=T A_2)}{P(A_1) P(B=T A_1) + P(A_2) P(B=T A_2)}$
There is a code error and testing does not find one	Undetected Errors (False Negatives)	$P(A_1 B=F) = \frac{P(A_1) P(B=F A_1)}{P(A_1) P(B=F A_1) + P(A_2) P(B=F A_2)}$
There is not a code error and testing does not find one	No News is Good News (True Negatives)	$P(A_2 B=F) = \frac{P(A_2) P(B=F A_2)}{P(A_1) P(B=F A_1) + P(A_2) P(B=F A_2)}$

Table 1. Bayesian Probability Formulas for Each Outcome

In the example presented in Figure 3, it is initially estimated *a priori* that there is a 10 percent probability that the code is infected with errors. From past knowledge of our testing process, the likelihood of detecting an error when an error is actually present is estimated to be 90 percent. The likelihood of detecting an error when an error is not actually present is estimated to be 10 percent. Using the Bayesian formulas in Table 1, we calculate that if an error is detected, the relative confidence of a “good hit” (50 percent) equals the relative confidence of a “false alarm” (50 percent). This result is surprisingly low, despite our ability to detect errors 90 percent of the time with our testing process.

The resulting low probability to correctly detect an error predicted by Bayes’ theorem, runs counter to our intuition, which is based on our experience with the testing process. However, the explanation for this counter-intuitive result can be easily explained by the geometrical representation in Figure 3. The probability of a good hit is equal to the area of the upper left rectangle, whose area is given by probability of infection times the likelihood of detection, or 0.10 times 0.90, which equals 0.09. On the other hand, the probability of a false positive is equal to the area of the upper right rectangle, whose area is given by the probability that the code is not infected times the likelihood an error is detected when one is not there, or 0.90 times 0.10, which also equals 0.09. Therefore, in this example, whenever an error is detected, we are only 50 percent confident that an error is actually present, even though the likelihood of detection is very high.

The picture also reveals something even more important, namely, the rectangles representing true hits and false positives make up only a very small fraction of the total possible area. The very large rectangle at the bottom right of the picture, representing true negatives, means that when no errors are detected we have a very high probability that no errors are actually present. Finding large numbers of errors might actually be a “red herring” if the testing process still misses a few critical Y2K undetected errors that will cause system failures and/or data corruption. The primary purpose of Y2K resolution is to reduce the number of undetected errors to zero, not to maximize the number of total errors detected. The expense of finding an error (either a true or a false error) during pre-rollover

testing and correcting will be nowhere nearly as costly as the occurrence of a single catastrophic undetected error during Y2K rollover. Consequently, the scarce resources available for fixing and testing Y2K errors should be used to reduce undetected errors. These concepts are summarized for this example in Table 2.

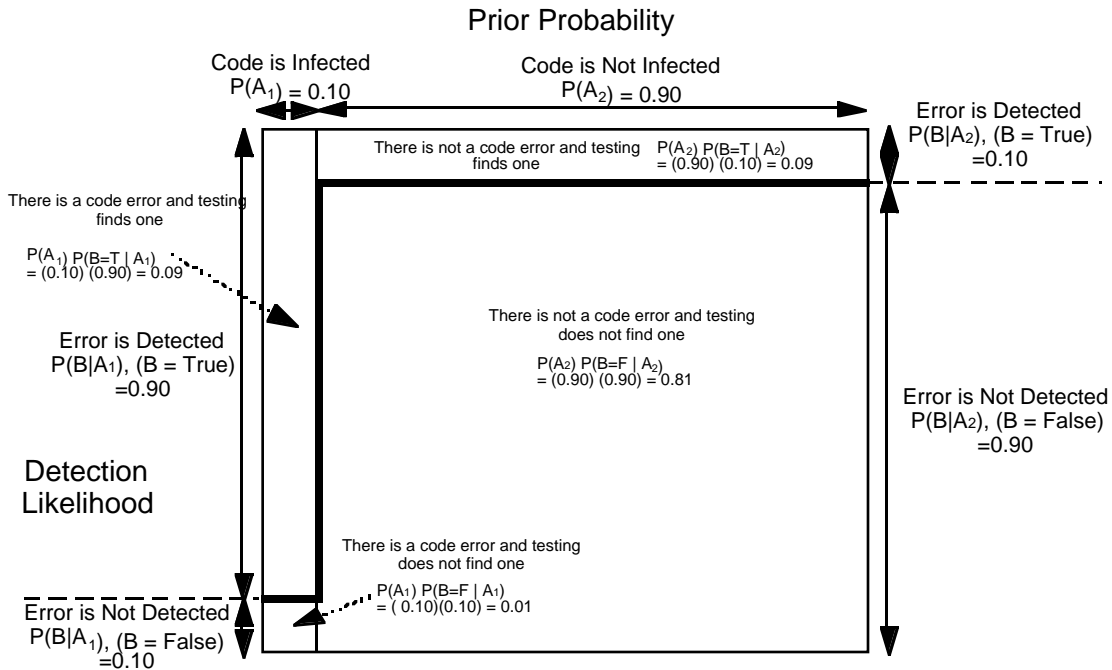


Figure 3. Geometrical Representation of the Example

The 50 percent posterior probability estimated for both true positives and false positives are surprisingly high, in view of the 90 percent detection likelihood. However, the probability of an undetected error is only one percent, which is a better measurement of our ability to avoid catastrophe.

Suppose the current resolution cycle reduces the error infection rate to 5 percent, and the next resolution cycle can achieve an improved error detection likelihood of 95 percent, then the resulting posterior undetected error probability predicted by Bayes' theorem would drop to 0.25 percent, or a factor of four improvement during the next cycle. On the other hand, the posterior probability estimates for true positives and false positives would still each be 50 percent, which illustrates how looking only for positive test results can be misleading.

<u>Outcome Description</u>	<u>Absolute Probability</u>	<u>Relative Confidence by Test Outcome</u>
Good Hits (True Positives)	9 %	50 % (Tests positive)
False Alarms (False Positives)	9 %	50 % (Tests positive)
<hr style="border-top: 1px dashed black;"/>		
Undetected Errors (False Negatives)	1 %	1 % (Tests negative)
No News is Good News (True Negatives)	81 %	99 % (Tests negative)
	100 %	

Table 2. Results of Bayesian Confidence Example

When a multi-cycle resolution process is implemented, prior Y2K confidence estimates can be revised at the end of each resolution cycle to reflect the most recent results from error testing activities, to obtain improved estimates of a successful outcome. Many systems will complete their initial resolution cycle by the end of 1998. This initial cycle will have included the early assessment and correction of most of the Y2K errors. However, the possibility that critical Y2K errors may still exist after the initially planned resolution cycle is completed will make it necessary to perform one (or possibly more) “last ditch” resolution cycle in 1999 to fix remaining errors, as shown in Figure 4. During 1999 this “last ditch” effort must be done in earnest to reduce all critical Y2K errors to zero. The ultimate “test” will occur at Y2K rollover, where hopefully no emergency actions will be necessary to deal with crises.

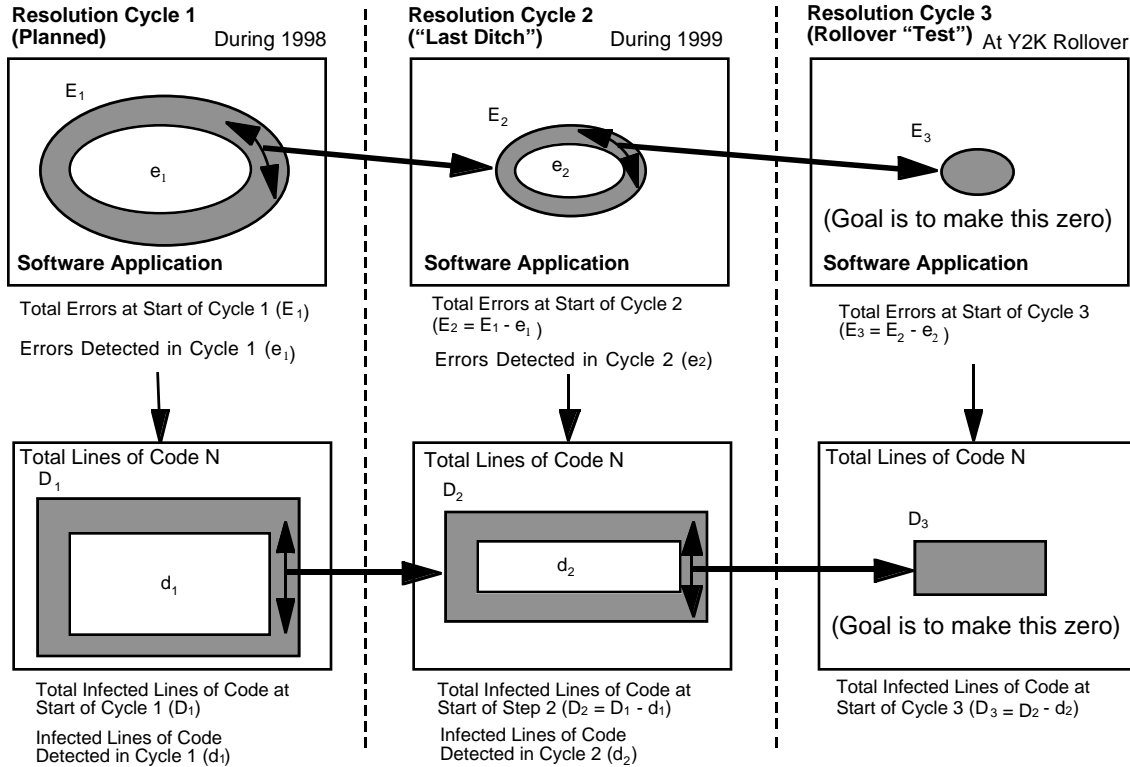


Figure 4. Multi-Cycle Year 2000 Resolution Process

Bayes' theorem can be useful for analyzing the confidence level associated with the process to resolve the Y2K problem. Bayesian confidence assessment can be applied at various points between now and the Y2K rollover, especially at the end of a completed resolution cycle. Each system must provide its own specific input data needed to perform the Bayesian confidence assessment. Initially the quality of the data might be low, due to the need to rely more heavily on judgment for the *a priori* estimates when there is larger uncertainty.

When input data is highly uncertain, especially early in the process, data should be estimated as ranges, perhaps by using the results of past testing experience or by using "gray beard" teams to assure independence and greater objectivity. Gray beard teams are used by industry and government to bring a structured review process to a particular problem or question. The team members represent an appropriate combination of expertise and experience needed to address the key issues. A key service for Y2K that could be performed by gray beard teams would be to determine the relative expected percentage mix of minor, major, and critical Y2K errors in each system. In addition, the results of test measurements can be used to estimate *a priori* probabilities and detection likelihood needed for Bayesian analysis. Subjective assessments can be subsequently revised as experience with working the Y2K problem grows.

Acknowledgments

The author acknowledges the contributions of Dennis Mangsen, who gave much helpful guidance on project issues, and made the suggestion for developing a pictorial representation of Bayes' theorem that could provide a more intuitive way of understanding

its application to confidence assessment than is found in standard textbooks on the subject. His comments about how to interpret the Bayesian pictures have been particularly insightful. Also greatly appreciated were the useful discussions with George Stark about Bayesian problem formulation and methods for utilizing available test data for confidence analysis.