Metrology Concepts: k=3.9? . . . Why Not??

Speaker/Author:
Howard Zion
Director of Technical Operations
Transcat, Inc.
35 Vantage Point Drive
Rochester, NY, 14624
Phone: (800) 828-1470
Fax: (585) 352-4547
hzion@transcat.com

Abstract

This paper will begin with an elementary review of the differences between TAR (Test Accuracy Ratio) and TUR (Test Uncertainty Ratio), underscoring the reasons for ISO-17025 and the GUM (Guide to the expression of Uncertainty in Measurement). In the discussion on TUR, a demonstration of the application of k=2 will be presented with respect to the UUT's (Unit Under Test) tolerance.

Other papers have been published that discuss an 'Indeterminate' region. These types of calibration results will be addressed and quantified, illustrating the probability that a reading may indeed be OOT (out of tolerance). Conversely, a probability exists that the reading is in tolerance. When attempting to determine this probability using k=2 for the reporting of a TUR, a problem arises if the entire area under the Normal Probability Density Function is not considered: the result is a misrepresentation of the OOT probability.

This will lead to the concept that, although k=2 is a good reporting format for the uncertainty of a measurement, TURs should be standardized using k=3.9. It is the author's hope that this will spark discussion that will take the Metrology industry to the next step in tackling this 'Indeterminate' area.

Introduction

With respect to my Metrology career, I was raised in the school of thought that the accuracy of my standards needed to be better than the accuracy of the instrument that I was calibrating. This comparison of accuracies is known as a Test Accuracy Ratio (TAR). 10:1 was a satisfactory ratio. Good thought, right? Well, it was a start anyway. But this thought is incomplete, as I have learned over the years. **We** weren't looking at the whole picture (this is largely how the Metrology industry approached calibration 25 years ago!).

I have traveled along many paths in my Metrology career and have met quite a few interesting people along the way. Many have come and gone. A few have remained, committed to making heads or tails out of this industry. And, some have resurfaced like an old penny. I'll tell you though; I have met some truly brilliant minds along the way! Did you ever meet someone who seemed to be able to recite more details that you have forgotten? I have had the fortune to work with a few of 'those guys' at the Kennedy Space Center. Paul, Perry, Otto, and Lois come to mind. Oh, and let's not forget Tom! He and Otto were my mentors for the greatest portion of my tenure in the Standards and Calibration Laboratories at KSC. I learned a lot in the decade that I spent working there: regression fitting, effective use of Logarithmic functions, practical database concepts, chemical and gas analysis techniques, and the concept of uncertainties in measurement, among others. Many of these concepts were not taught in the numerous Air Force technical schools I had attended early in my career. So I returned to school to complete an engineering degree while working at KSC. What a powerful combination: absorbing theoretical information at night and applying it in the 'Lab' every day! I highly recommend it!

The concept of uncertainty in measurement is one that considers other sources of error in addition to the accuracy of the standard(s). This was really not a new concept to me, though. PMEL school indeed taught about sources of error but in the respect that, by controlling these errors in a laboratory environment, one could effectively eliminate or diminish these to a point of no concern. I walked away from my Air Force days with the concept that a calibration was a controlled experiment in which one or two variables were isolated and measured against 'known values'. Nothing else really mattered, since all other errors were not appreciable enough to worry about.

I quickly learned that by considering and quantifying (or estimating, if not measurable) all sources of error we come closer to the truth about a measurement, no matter how small or inconsequential these sources or 'components' might seem. And, if we account for this total error as it compounds from the top of the chain of traceability (National Measurement Institute, or NMI) through various calibration labs and on through to the production processes, we can make a more accurate statement about the risk that the producer bears and, more importantly, the risk that the consumer faces. And **that** is what Metrology is all about!

As a participant in this traceability chain, it is my duty to report (to my customers) the cumulative errors that influence my measurements; that is to say, I must consider all of the errors that have come before me, which are included in the uncertainty estimates of the standards I am using. If I have not received an accredited calibration on my standards, which provides these uncertainties in the calibration report, then I have lost a key piece of information and cannot possibly provide a complete estimate to my customers. Just as important, if I am using a supplier for the calibration of my standards whose accreditation process is not traceable through an international oversight, such as ILAC (www.ilac.org), then the uncertainties that are reported may be invalid, incomplete, or incorrect.

In addition to the uncertainties of my standards, I must consider any other factors that influence the direct measurement that I am making. By quantifying and combining all of these potential sources of error, I too can develop an estimate about the uncertainty of my measurement. By reporting the estimate of my uncertainty of measurement to my customer, they can then integrate this into their process of measurement, whether it's another link in the chain of traceability to a subsequent link, or the measurement of the end product. My customer can then follow the same process of calculating the combined uncertainty of their measurement process, which includes the uncertainty that I reported, which included the uncertainty that my supplier reported for my standard(s), and so on – all the way back to the National Measurement Institute that originated the measurement(s).

Having stated this concept of uncertainty in measurement, we can now distinguish between the *accuracy of the standard* being used and the *uncertainty associated with the measurement process*. The accuracy of the standard comes from the manufacturer's specification. The uncertainty of the measurement is derived from an uncertainty budget that can include the *accuracy of the standard* as well as other errors that cannot be ignored. Referring to the instrument we are testing as the Unit Under Test, or UUT, if we want to talk about a good ratio between the UUT and the standard we are using, we can no longer use just the accuracy of the standard in this comparison. We must also include all other errors as the total 'uncertainty' of the measurement. So, instead of dividing the *accuracy of the UUT* by the *accuracy of the standard* (remember, this is TAR), we now must divide the *accuracy of the UUT* by the *uncertainty of the measurement*. This is known as the Test Uncertainty Ratio (TUR), which is a more truthful statement about how good the measurement process is with respect to the instrument being calibrated. Contrasting TAR with TUR: same numerator – more complete denominator.

Allow me to pause for a brief moment for a word on specifications and the *accuracy of the UUT*: Basically, the user of the instrument (or someone in their organization) at one point should have determined the need for the instrument by comparing their product (or process) requirements to the Original Equipment Manufacturer's (OEM) specification for the instrument. These specifications establish tolerances and a time limit (cal interval) over which the instrument can be expected to hold these tolerances. If you are a Quality Engineer or someone else who is responsible for understanding the relationship of your test and measuring equipment to your product or process then you should know that, upon calibration of these instruments, once the instrument exceeds any of these specification 'thresholds', these Out of Tolerance (OOT) conditions become a 'warning flag' to you that indicates you need to go back to your process and determine if this non-compliance condition detrimentally affected your product. Additionally, if anything about these specifications is modified by you or your organization, then there are a number of people who need to know about these modifications (whether a limited range, change of tolerances, etc.). First, in order to get the calibration that is expected, you need to inform the calibration service provider so that the calibration data given mirrors these changes and to reset the trigger points for these 'warning flags'. Second, the calibration label (and calibration report) needs to indicate to all potential users of this instrument that something has been changed about this instrument and that they should not assume that they can simply look up the OEM's specifications and use the instrument based on these tolerances and/or ranges. This could happen with a regular calibration label applied to the instrument (and I've seen this happen often enough to know it is a problem). A Limited Calibration label will take care of this precautionary measure. Even if the instrument is dedicated to a particular system or process, by having the proper information on the label you're ensuring that everyone is aware of a potentially dangerous situation. Without it, you're opening yourself up to risk – big or small – when it could have been eliminated through the use of a simple sticker. Some people are hesitant to use anything that suggests a deviation in fear that it will raise a flag in their quality system. I am here to tell you that, if you do not have some means of flagging those instruments, you are setting yourself up for potential disaster, among which could be: risk of using an uncalibrated function in your process and risk of using the instrument against incorrect tolerances. By **not** indicating that the instrument is 'Limited' or changed from the OEM's specifications you are, in effect, removing those 'warning flags' that the calibration process would otherwise catch, which can lead to an audit finding. Limited Calibration labels have a valid purpose and should not be avoided. So, please don't misunderstand their purpose in reducing risk in your quality system.

OK, enough said about that – now back to the main topic. As much as it pains me to say it, the Metrology industry is headed in the right direction with these concepts being incorporated into laboratory practices more and more, but there is still something missing. There is yet another level to consider in this process. We still haven't fully developed these concepts to the point that we can make the resulting calibration information easy to use for the subsequent step in the chain of traceable measurement. Remember, the purpose of Metrology is to support commerce by ensuring that the best estimate of our measurement errors is known or minimized and to keep the measurement chain in alignment around the globe. Along the steps of the chain, though, we need to be able to make some statement of compliance to the specifications assigned to the UUT so that our customers can rely on those 'warning flags' (i.e., those "not-to-exceed" tolerance limits) they have established in their process. There is more to this than meets the eye, which this paper will cover. It is this author's goal to present a possible solution that will allow for standardization of compliance statements, which will help to simplify the process for addressing Producer and Consumer risk. If the Metrology industry is in agreement with this paper, then I recommend an NCSLI committee be formed to produce a standardized format for industries to follow.

An Illustrative Example

Consider a customer who submits a signal generator for calibration. Using one specific step in the calibration process as an example, we'll consider the calibration of the output signal at 10 MHz. For this particular model, the OEM's specification for frequency accuracy at 10 MHz is ± 0.2 MHz (fig. 1).

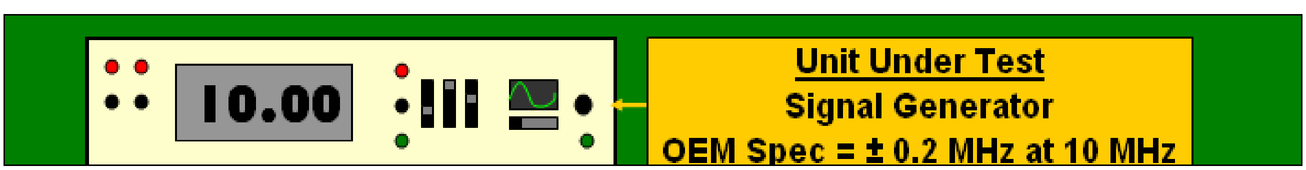


Figure 1: Calibration of signal generator at 10 MHz.

Another way of viewing this is by plotting it on a graph, which is shown in figure 2.

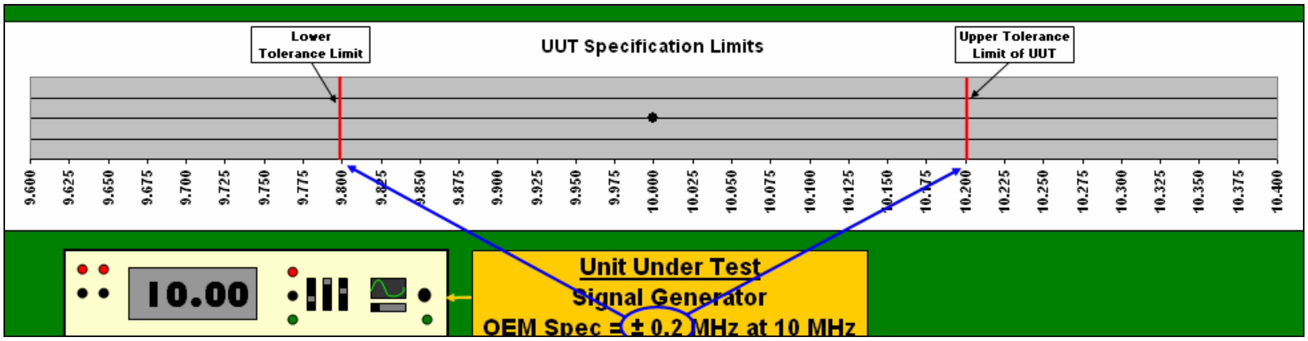


Figure 2: Graph of tolerance at 10 MHz.

Now we need a calibration standard in order to determine the error in this 10 MHz measurement. So, after crunching some numbers in an uncertainty budget, we determine that the standard shown in figure 3 is sufficient for this application.

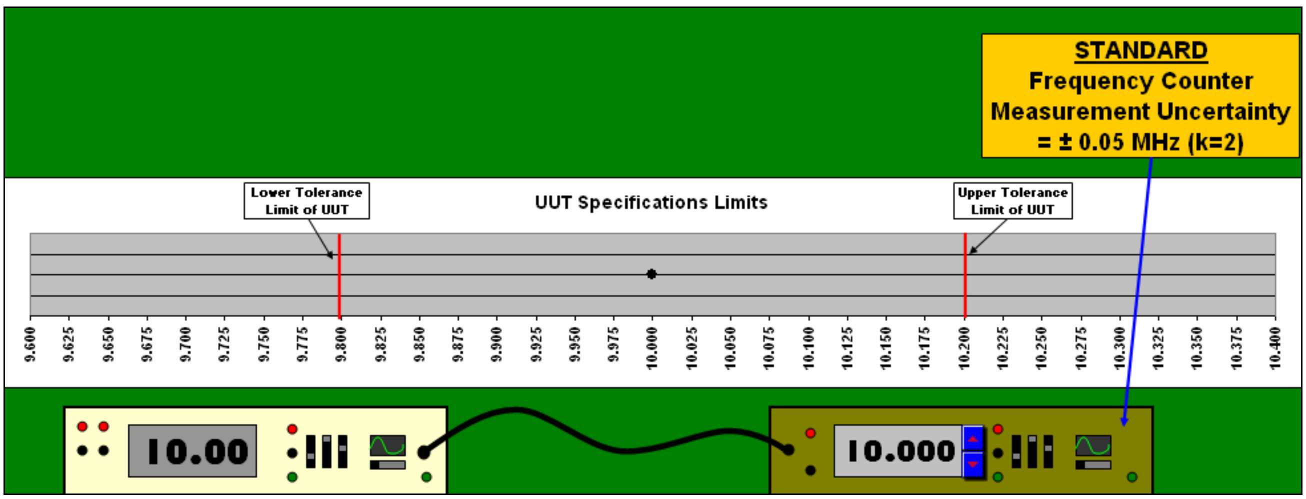


Figure 3: UUT and selected calibration standard.

Note that the tolerance shown for the standard is reported as the *measurement uncertainty* at a 95% confidence interval (k=2 or 2σ). That is to say that this is not merely the *accuracy of the standard* but, rather, the *uncertainty of the measurement process* for this 10 MHz signal which includes (in this example) the accuracy of the standard as one of its components. Figure 4 shows this uncertainty that surrounds the measurement at 10 MHz with the addition of an uncertainty bar.

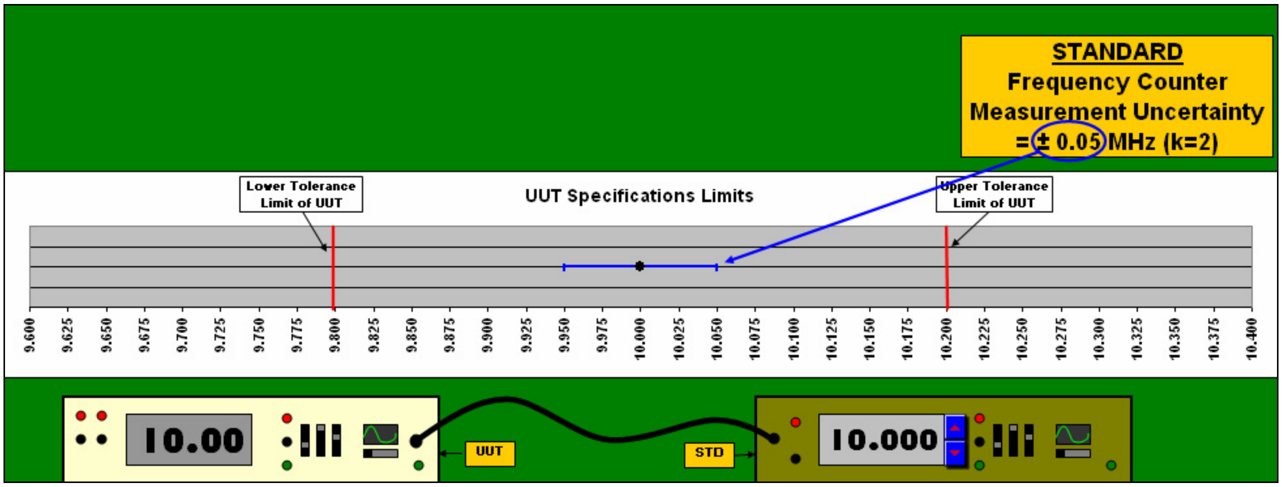


Figure 4: Uncertainty surrounding the measurement.

Given the uncertainty of the measurement, we can calculate the TUR. Referring to figure 5, we see that the measurement process uncertainty is 25% that of the UUT's tolerance – or at least this wording follows the format as it is stated in ANSI/NCSL Z540-1-1994. A simpler way of putting it: the measuring process is 4 times better than the instrument being calibrated.

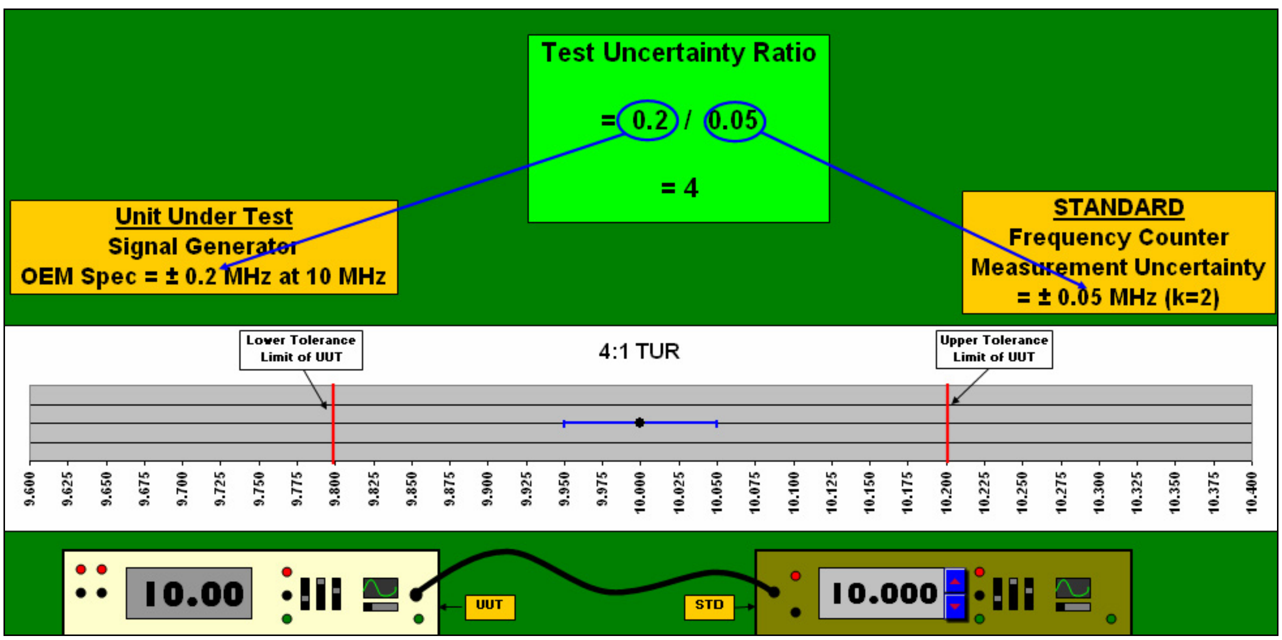


Figure 5: Origins of the 'Test Uncertainty Ratio' calculation.

In figure 5 we see that the standard reads 10.000 MHz, indicating that the UUT is not only in tolerance, but apparently has no error when generating a 10MHz signal. Additionally, this example illustrates the fact that the uncertainty surrounding our measurement has no bearing on the statement we are making about the UUT's compliance to the manufacturer's accuracy specification . . . or does it?

Over time, every instrument drifts. The measurement we made at this moment in time is a static measurement that changes from one moment to the next. We have 'captured' the UUT's error at this particular moment, which also implies that we cannot make a statement about any future value of this 10MHz signal. The reason for bringing up this point is that the "As Found" (or "As Received") readings are critical because this information indicates how the instrument has performed since its last calibration. If there are any OOT conditions, then these are the warning flags that I spoke of earlier. Again, to the QEs (and/or associated positions of responsibility) in the audience, you need be evaluating these OOT test points against the processes in which this instrument was used to determine whether or not this detrimentally affected your product. Although this data cannot tell you what the instrument read one week or one month ago, it can point to a potential problem with your product that must be resolved before you potentially spend thousands or millions of dollars in scrapped product and/or rework. Time is truly of the essence in performing a reverse traceability of the instrument to the product lines!

Although this was an important point to make, the purpose in speaking about instrument drift is that my standard drifts too! The expanded uncertainty (k=2) covers that potential drift over time, or at least most of it as we will see. This uncertainty value is represented by a normal, or Gaussian, distribution (fig. 6).

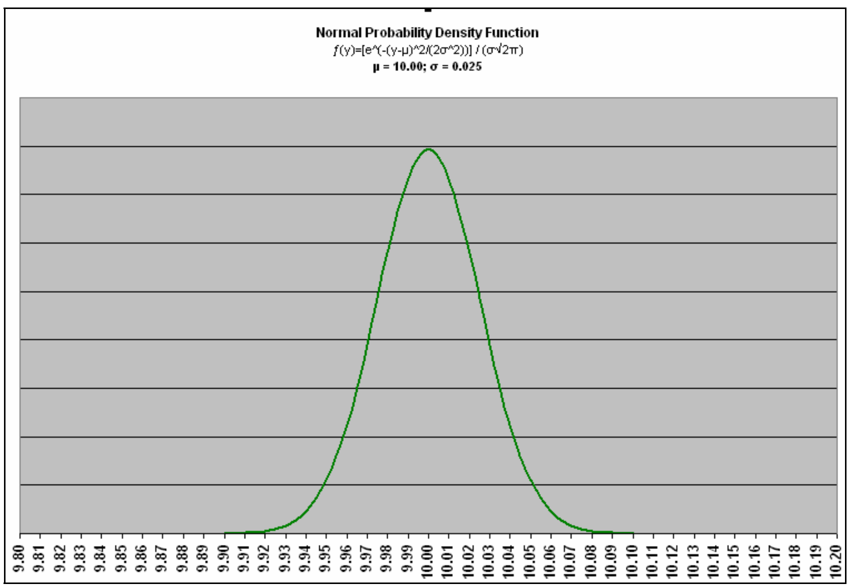


Figure 6: Normal (Gaussian) Probability Distribution.

Hmmmm . . . so what does this mean? Well, the concept of normally distributed probabilities is that, most of the time, the readings are very close to the center of the distribution; the center being the mean (or average) value. Sometimes, but less often, readings will occur further away from the mean value. The further away from the mean, the less likely this event will happen. As an example of this, think about at an hourglass. As the sand falls through the orifice, most of the grains pile up in the middle, while some fall away to the sides. As the sand passes through, the probability that the grains will drop to the center is higher than the probability of falling to the sides. This is *normally* the way that many events occur in nature, whether sands in an hourglass or measurements in a process.

Not all things follow this pattern though. There are many different distributions to represent other events; ones that do not follow a normal distribution of probability: Binomial, Chi-Square, Gamma, Weibull, etc. But, for the measurement process, the Normal distribution and Uniform (or Rectangular) distribution are the ones used in describing the probability of measuring events and in developing uncertainty budgets. Occasionally you'll see a Triangular distribution used.

So, considering the uncertainty of my measurement over time, it is more likely that its absolute value will occur near the center of the distribution (all factors considered; i.e., the combined effect of all components of uncertainty) and less likely that it's true value will occur at one end or the other – although it could happen. This is the premise of the Central Limit Theorem in statistics. Keep this concept in mind as you read on.

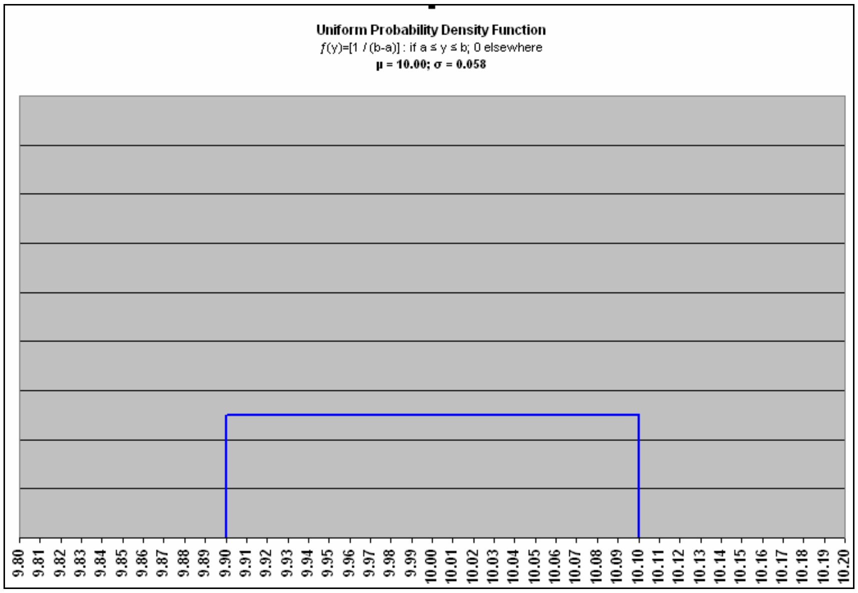


Figure 7: Uniform (Rectangular) Probability Distribution.

The Uniform distribution simply represents the fact that all values, or events under consideration, are equally likely to occur (fig. 7). There is no natural tendency for events to occur at one particular place or another. They all have equal probability.

Now think about a different outcome in the calibration, where we measure the output of the 10MHz signal from the UUT and the standard indicates 10.150 MHz (fig. 8).

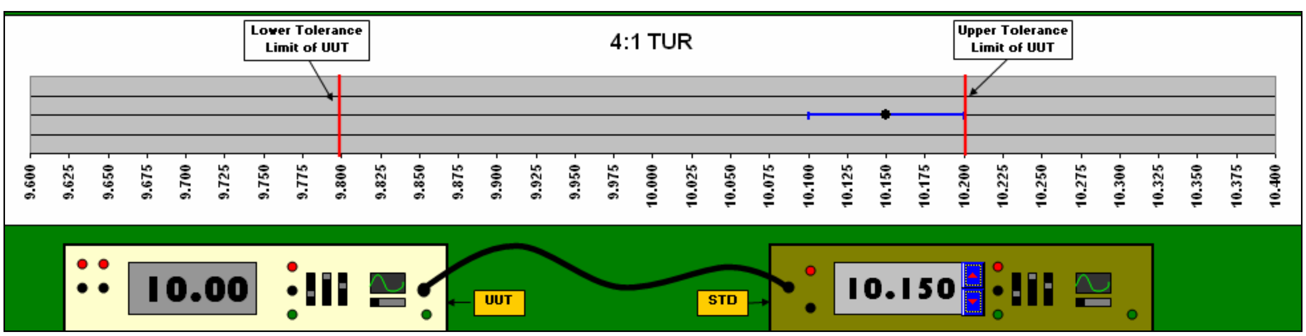


Figure 8: Standard indicates UUT's value is nearing the edge of the tolerance.

Taking into consideration the uncertainty of the measurement, our standard might show a front panel reading of 10.150 MHZ but it's true value could actually be 10.100 MHZ (left end of the uncertainty bar). Or, it could actually be 10.200 MHz (right end of the uncertainty bar), which would indicate that the UUT is about to trip the warning flag to the customer! Problem is, we simply don't know – we are uncertain (within these estimated limits) about the true indication of the standard at this moment in time.

Another possible outcome of the calibration is that the standard indicates 10.200 MHz (fig. 9). The uncertainty around the measurement means that it could really be 10.150 MHZ, in which case there is no problem (left end of uncertainty bar), or it could be 10.250 MHz, in which case it is OOT (right end of uncertainty bar). Because the industry calls this an In-Tolerance condition, this situation does not set off a warning flag to alert the customer that they might want to look at their process. If the ratio between this instrument and their process is large enough, this most likely will not be an issue for them. But sometimes tolerances along the traceability chain are very tight and cannot be avoided due to challenges in the technology available for the measurement of interest, in which case this condition needs to be evaluated by the customer.

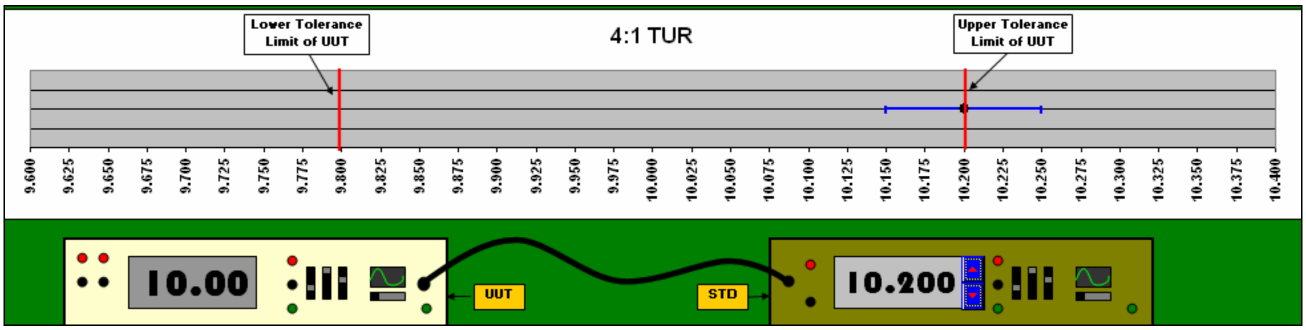


Figure 9: Standard indicates UUT's value is at the edge of the tolerance.

This range of inability to state clearly whether the instrument complies with the specs or not is often referred to as an Indeterminate area. This is also known as the Guardband range (figs. 10 through 13), within which it is highly recommended that the instrument be adjusted to return it as closely to its nominal value as possible.

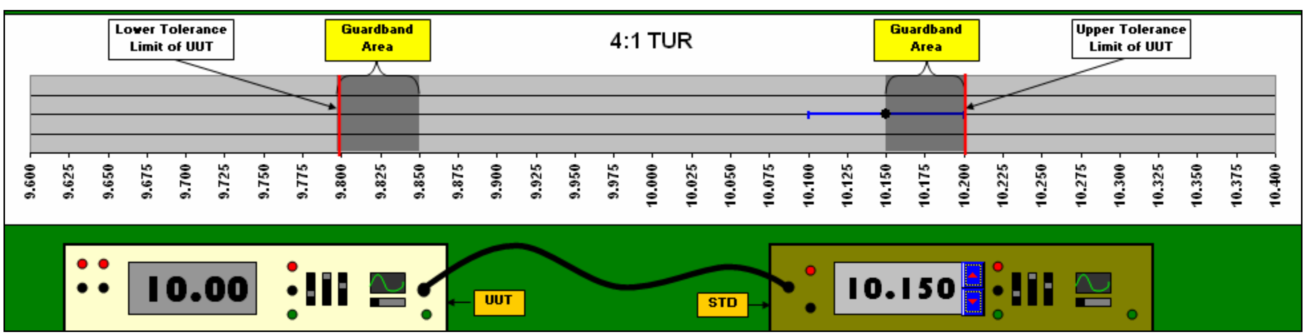


Figure 10: Guardband areas identified: Just entering the guardband.

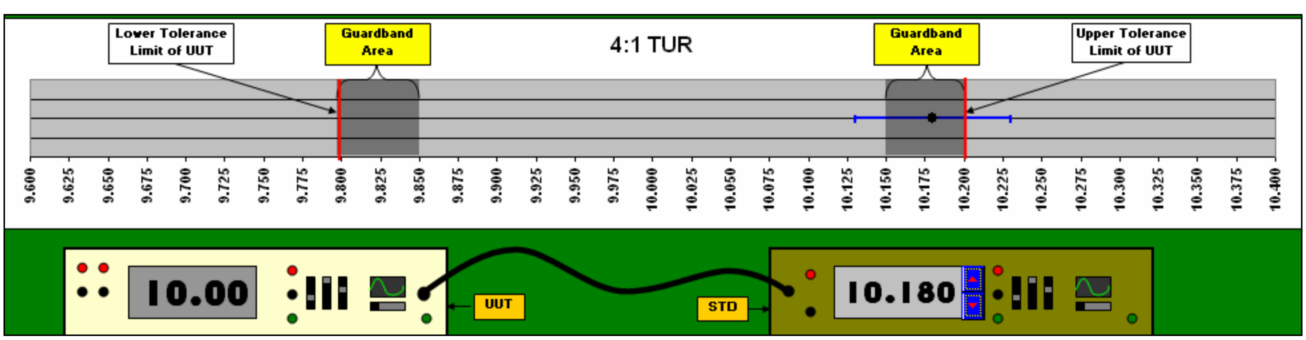


Figure 11: Guardband areas identified: Value within the guardband.

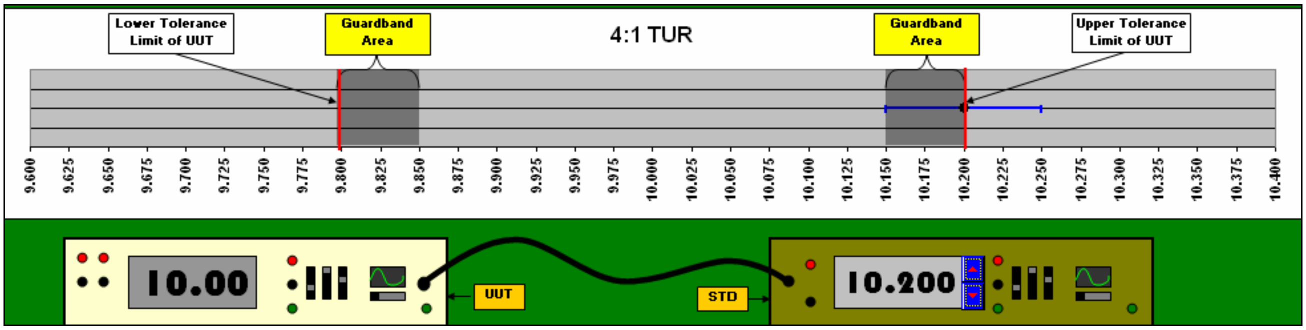


Figure 12: Guardband areas identified: Right at the UUT's upper tolerance limit.

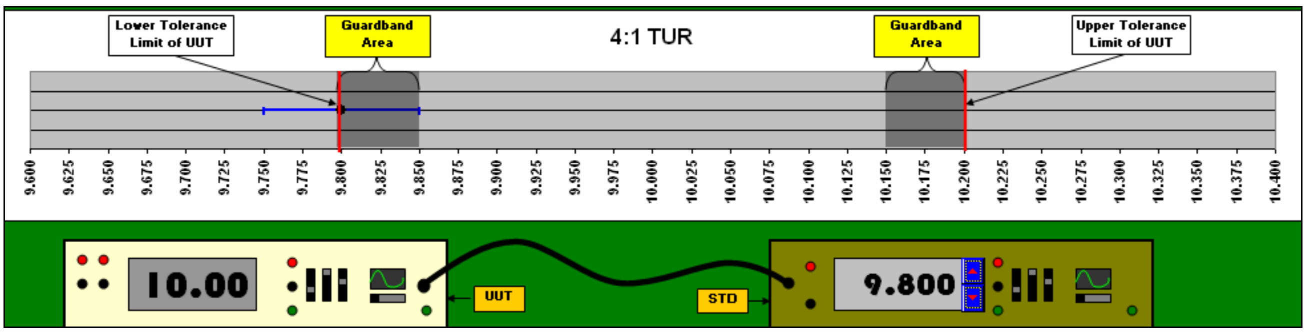


Figure 13: Guardband areas identified: Right at the UUT's lower tolerance limit.

Any outcome in which the standard indicates a value is within a guardband region implies there is a probability that the UUT is actually OOT.

Consider also that this Indeterminate region extends beyond the edge of the UUT's tolerances (figs. 14 & 15). Just as soon as the reading on the standard exceeds the UUT's tolerance we call it an OOT condition, right? Figures 14 and 15 appear to indicate OOT conditions. However, accounting for the uncertainty of the measurement, it is quite possible that this should have been called an in-tolerance condition (left end of uncertainty bar)!

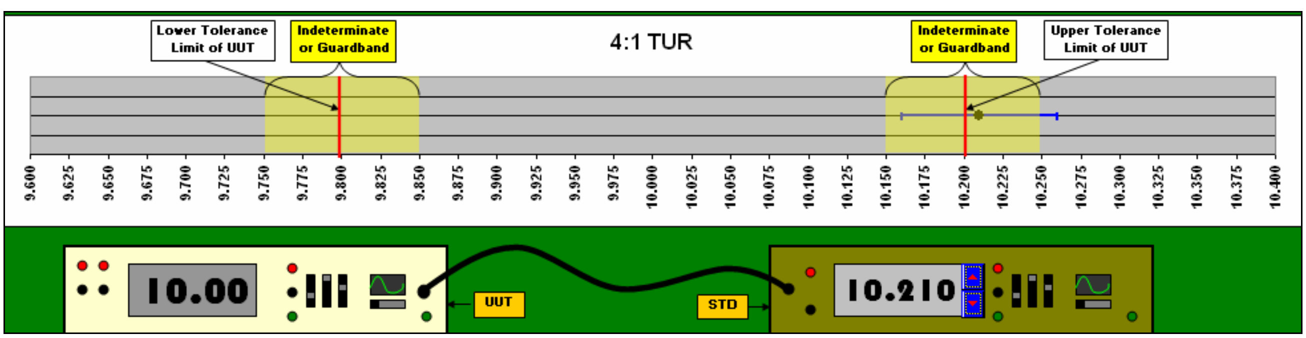


Figure 14: Indeterminate region: OOT, but the uncertainty might put the reading within tolerance.

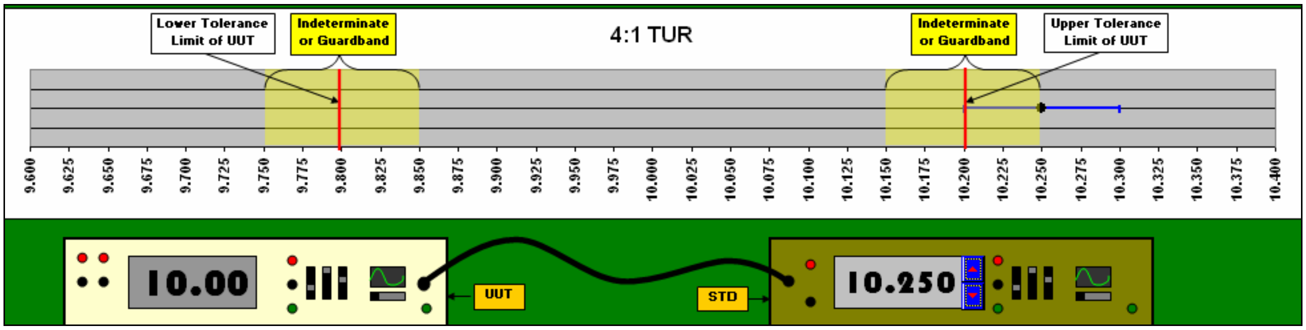


Figure 15: Indeterminate region: OOT, but uncertainty might put the reading on the UUT's upper limit.

On the flip side of that argument, it is also possible that the uncertainty of the measurement places the standard's true value further away from the indicated value (front panel), meaning that the UUT is further OOT than originally thought (right end of uncertainty range). This mental anguish of not being able to make a statement of compliance continues throughout this Indeterminate region until, finally, an outcome of the calibration occurs where the standard, with all of its surrounding uncertainty, lies fully outside of the UUT's tolerance (figs. 16 & 17)!! At this indication and beyond, we are quite certain that the UUT fails the calibration for this test point. No doubt about it! . . . maybe . . . read on.

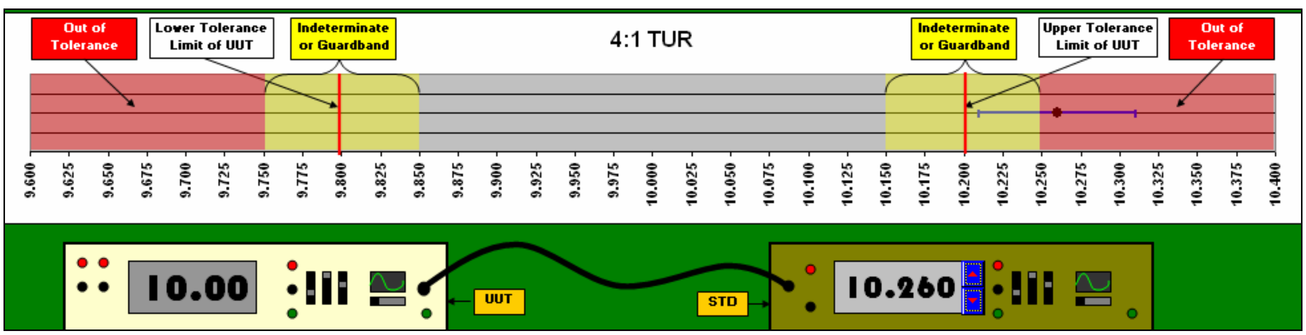


Figure 16: Mental relief? Standard indicates value lies outside of the Indeterminate region (Upper).

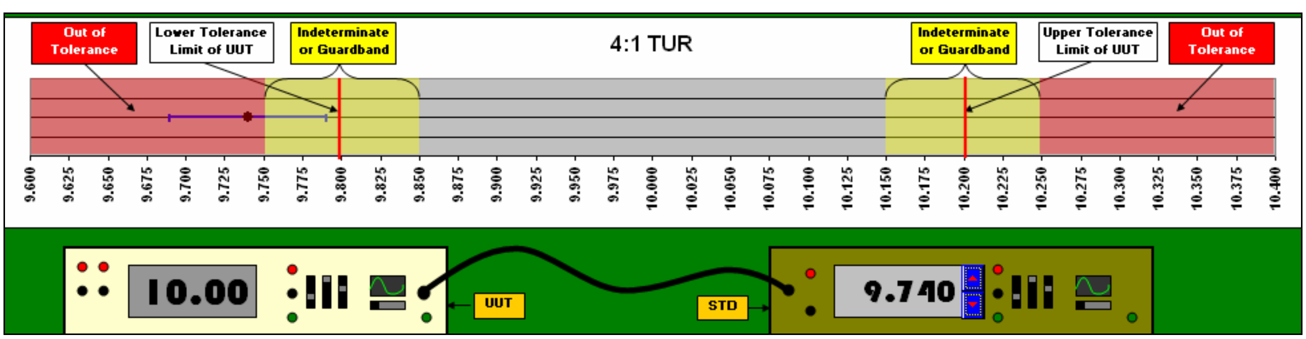


Figure 17: Mental relief? Standard indicates value lies outside of the Indeterminate region (Lower).

Something was obvious to me when I looked at these graphs. There's a gap in the middle of graph! So, what **do** you call the area in the middle? Since the red area indicates bad things will happen with the customer's process (e.g., now they have to do additional work to see if the product was affected), and the yellow area indicates that bad things might happen to the customer's process (but we are unable to quantify this), then the middle must be where 'Good Stuff' happens. I'll color this green and give it a name, indicative of the fact that we're neither in the 'Bad band' (OOT) nor the 'Guardband' regions. I'll call this the Safeband (figs. 18 through 20).

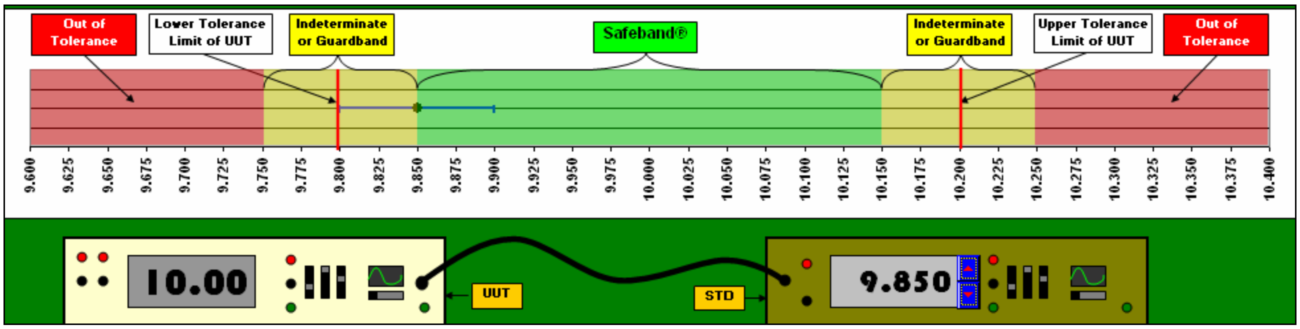


Figure 18: Safeband (Left Edge)

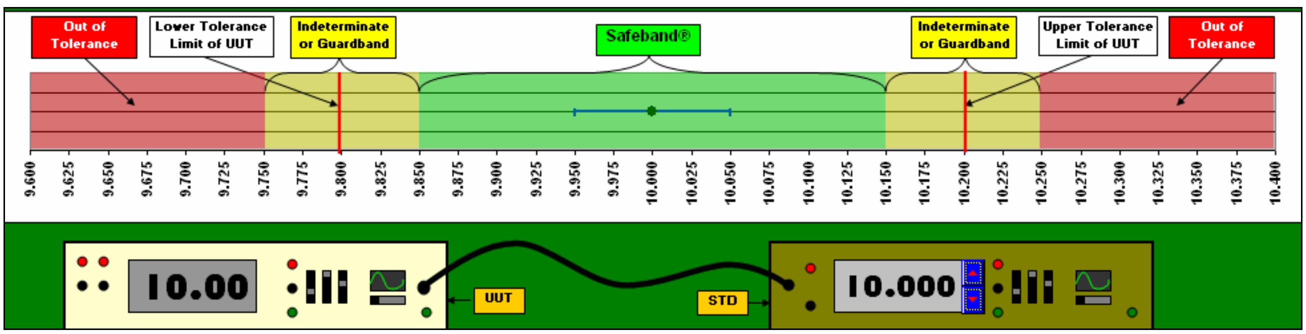


Figure 19: Safeband (Nominal)

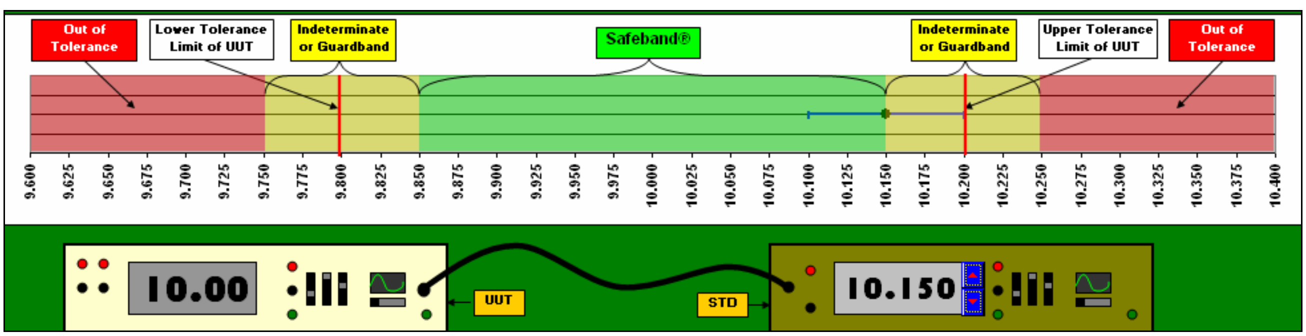


Figure 20: Safeband (Right Edge)

Are you with me so far? All of the information we've covered to this point is just the basis of what most people in the Metrology industry use to describe the application of uncertainties to the calibration process. I would like to now bring you a different perspective; one that will **literally** turn the Metrology industry's perception about statements of compliance on its end (and perhaps philosophically too)!

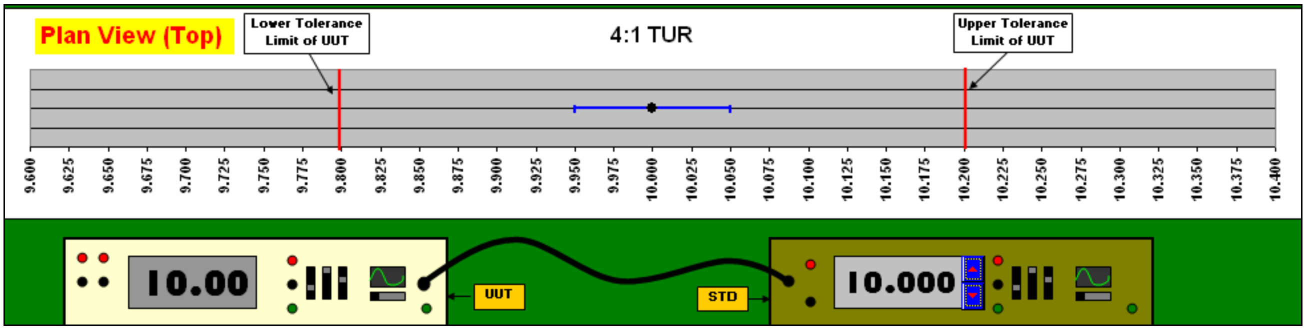


Figure 21: The traditional metrology perspective.

Figure 21 illustrates the traditional perspective that most of my colleagues have taken when contemplating measurement uncertainty. This is the one we've just finished reviewing in the previous examples. Let's refer to this in the same way that Architects and Draftsmen refer to drawings. This is the Plan View, or the view looking down from the top at the uncertainty surrounding the measurement with respect to the tolerance limits of the UUT. It's kinda like watching at a football game from a blimp flying overhead. Now let's go down to the 50-yard line and take a seat. I think we'll get an entirely different perspective on this 'uncertainty game'!

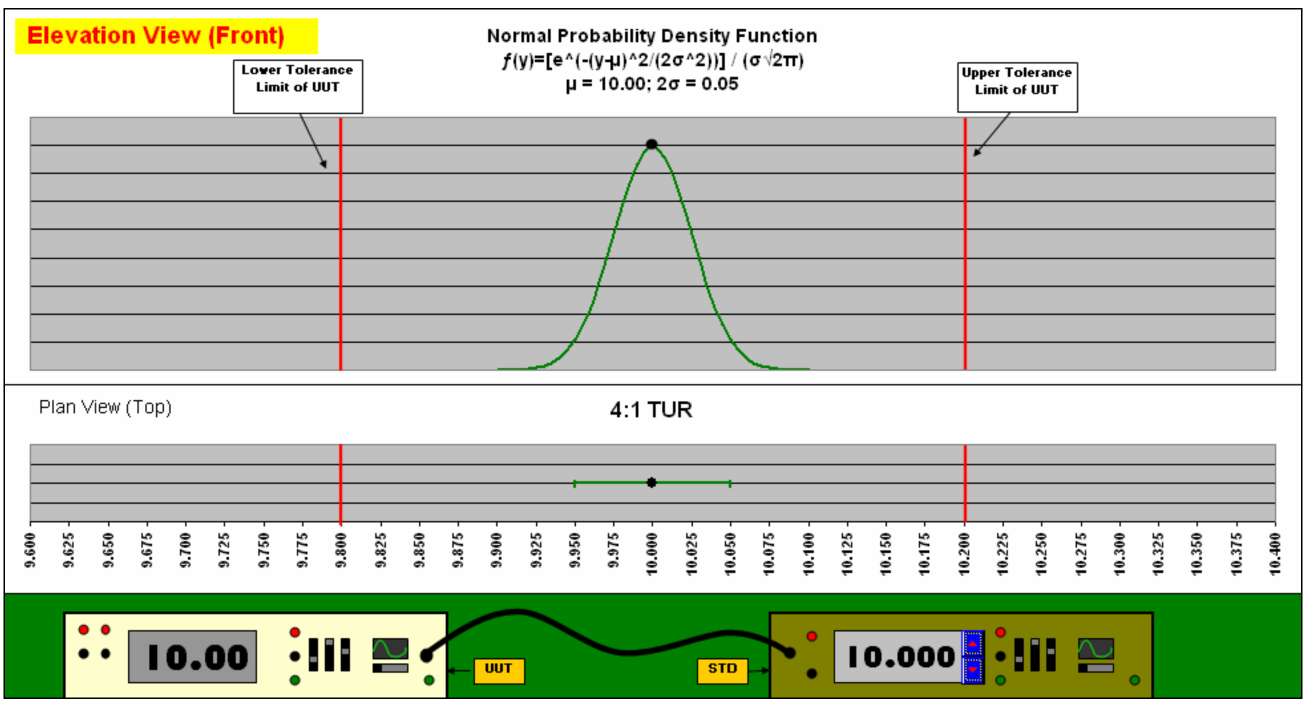


Figure 22: A new perspective: Watching the uncertainty game from the 50-yard line.

Going back to the architectural terms, I have now included an Elevation View (the 50-yard line), which turned the Plan View on its end, or backside. This allows us to see that the uncertainty bar really represents a normal distribution. We couldn't see this from the other perspective but now see it clearly from this vantage point.

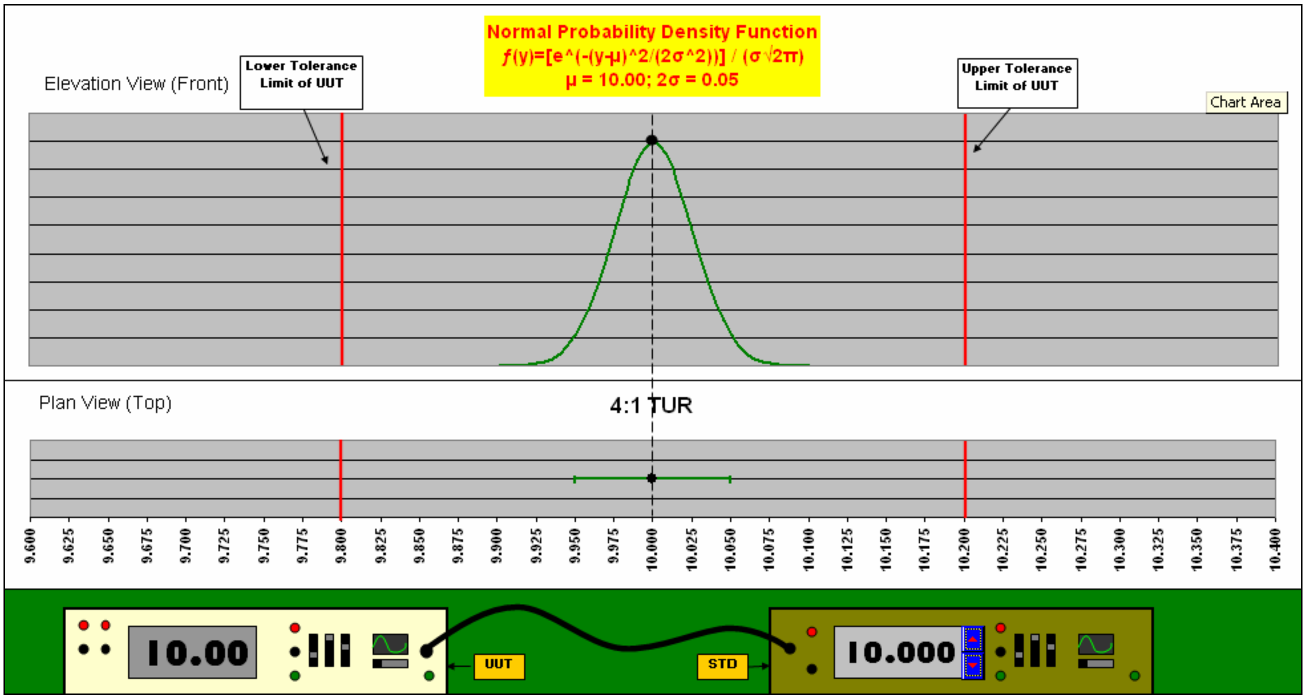


Figure 23: Plot of Normal distribution and the mean.

In figure 23, the Elevation View illustrates the graph of the Normal Probability Density Function. The formula is also shown along with the mean value.

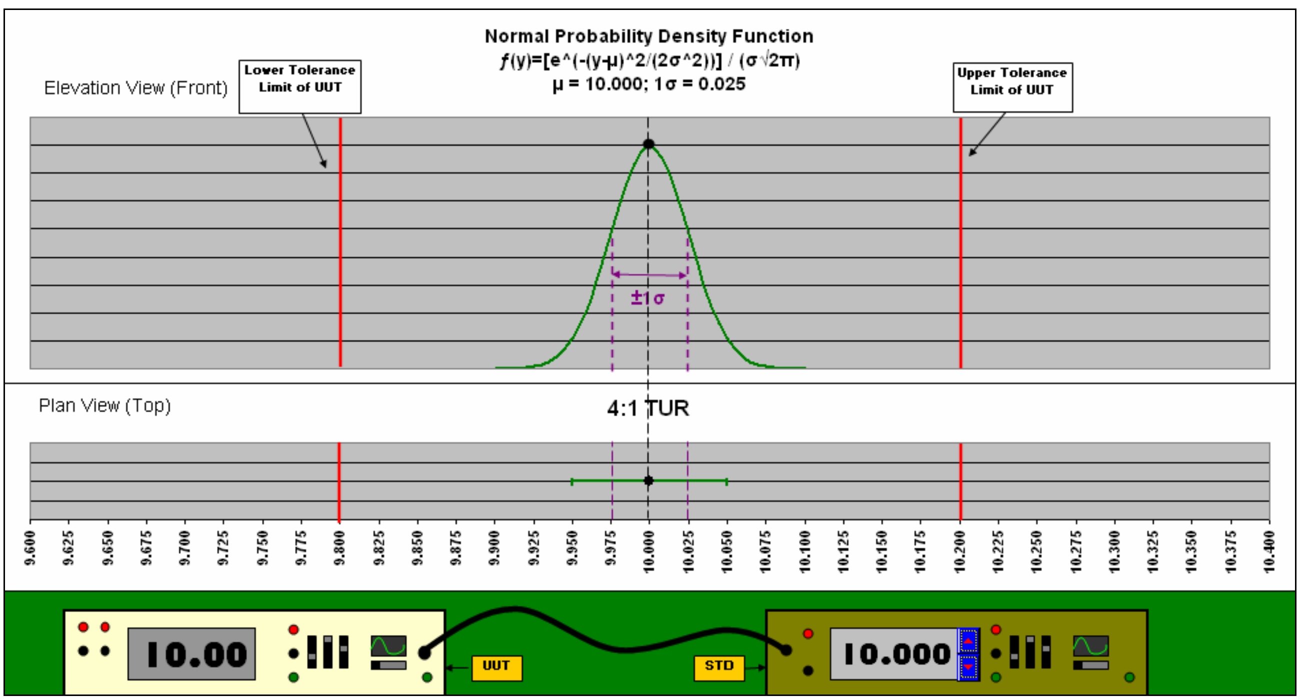


***Figure 24: 68% Confidence Interval** (±1σ or k=1)*

Figures 24 shows the 1σ value (1 sigma, or 1 standard deviation, or k=1) around the mean. Think about the area beneath the entire curve for a moment. Forget the absolute value of this area – think in terms of percentages. If the area beneath the entire curve is 100%, then the area beneath the curve that is within the ±1σ limits is something less than 100%, right? If you know statistics, then you know where I'm going with this thought process. If not, then this will build on an important premise that you'll need later, when I finally get to the point I'm trying to make.

The Empirical Rule in statistics tells us that no matter what the data is that you collect about a process, as long as its probability of events are normally distributed, then this ±1σ area around the mean will always represent approximately 68% of the total area beneath the curve. That means that there is a 68% probability that my standard reads within ±1σ of the mean, or within the values of 9.975 MHz and 10.025 MHz. This is usually written as, "The 68% confidence interval is: [9.975, 10.025]". Is that cool, or what?!

Moving on to the 2σ value (k=2) around the mean (fig. 25), the area beneath the curve that lies within these ±2σ limits is about 95%. Following the Empirical Rule we can make the following statement: The 95% confidence interval is: [9.950, 10.050]. Notice that this is the value (k=2) that we used for the uncertainty bar in the Plan View examples.

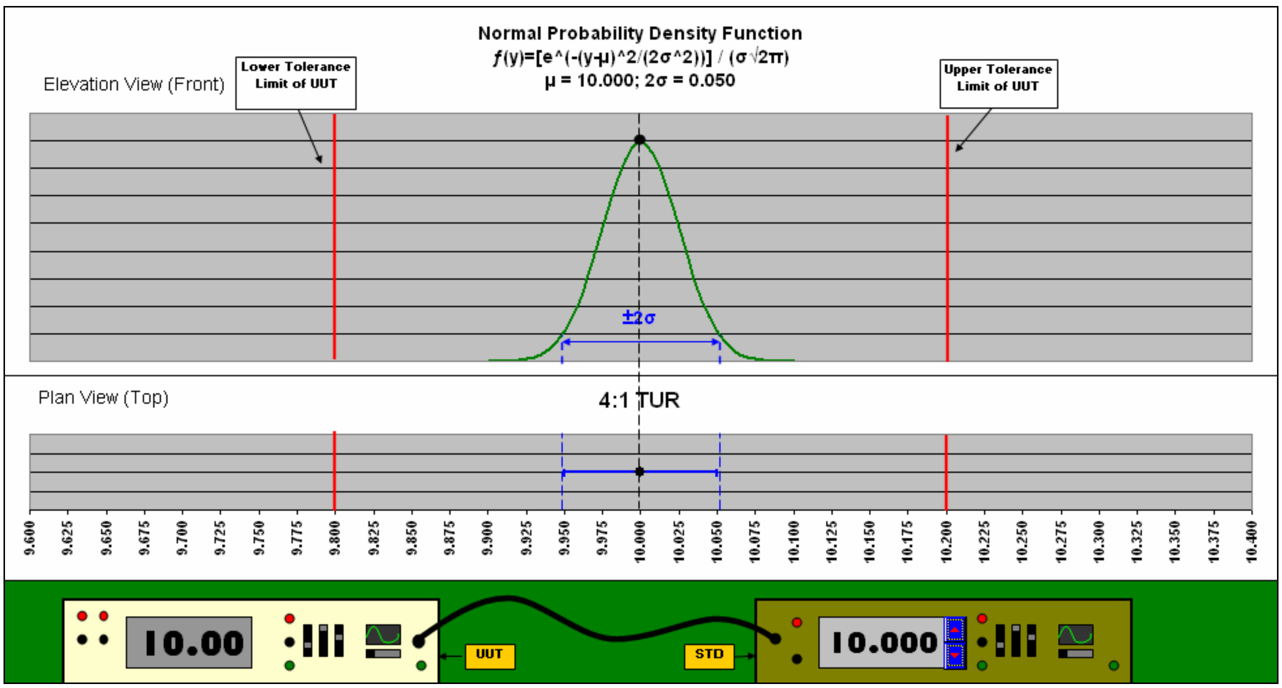


Figure 25: 95% Confidence Interval *(±2σ or k=2)*

The Metrology industry decided some years ago to standardize on reporting uncertainty values at a 95% confidence interval, or k=2, but it doesn't really matter. Why? Because whether you state the uncertainty value at a 1σ (1 standard deviation) value and report this as k=1 for a 68% confidence level, or state the uncertainty value at a 2σ (2 standard deviations) value and report this as k=2 for a 95% confidence level, they are both speaking about the same curve, just referring to the different values of standard deviation around the mean that yield different area percentages. In other words, change the confidence level and the number of standard deviations (k-value) that you report must change with it. Change the k-value and the confidence level must change with it – in order to maintain the same relationship that speaks to the curve from whence they came. Changing the k-value moves the brackets under the curve, which changes the area (confidence level) that it represents. They all correspond to the same distribution. Standardizing this at k=2 was a good way to remove some of the confusion for those who do not yet understand this concept.

In that respect, it should not surprise you that the 3σ value (k=3) around the mean (fig. 26) represents a 99% confidence level whose limits are [9.925, 10.075]. There is a greater confidence at k=3 because this larger area beneath the curve includes more of the probable events associated with this measurement process.

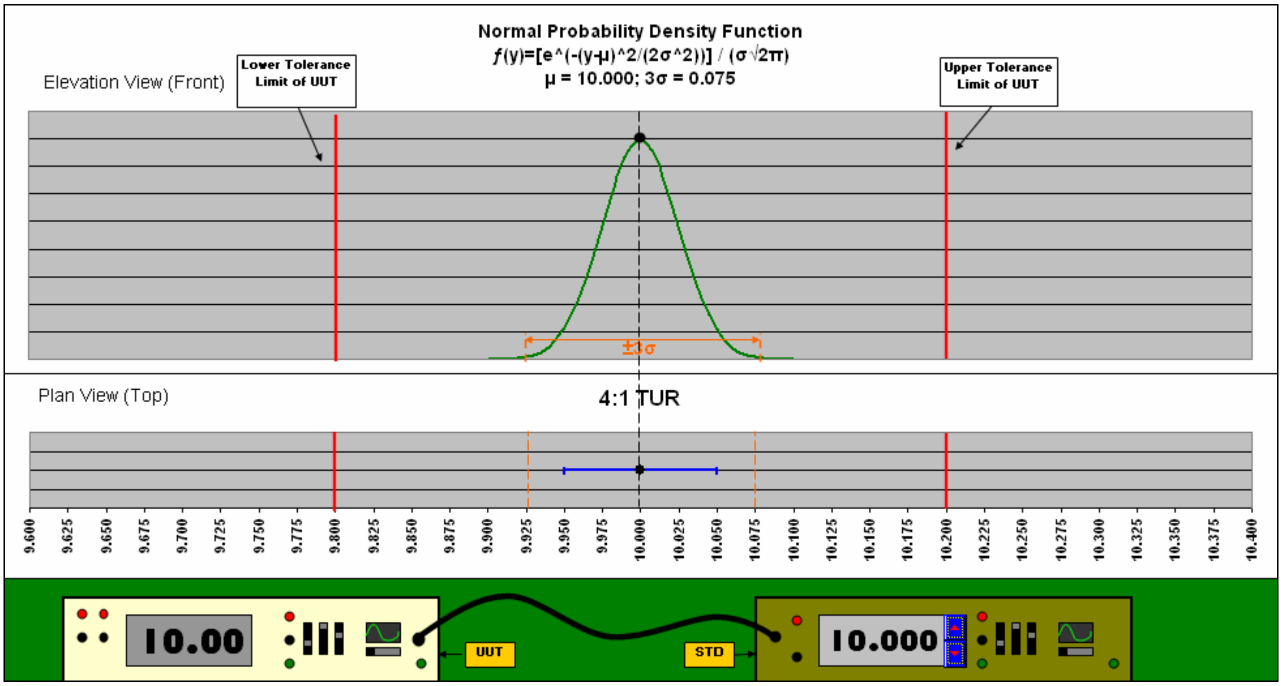


Figure 26: 99% Confidence Interval *(±3σ or k=3)*

Once we extend this to ±3.9σ (k=3.9) around the mean, we've pretty much covered the entire area beneath the curve (100%). Going any further to consider what happens at k=5 or k=6 (6σ) yields such minor changes in probability that any benefit we might receive for this application is insignificant. There are other applications in the world where 6-sigma is necessary. However, for this application, we'll focus on the significant benefit between k=2 and k=3.9.

Since the industry has agreed to standardize on reporting at k=2, we'll follow that thought for a moment to illustrate a point. In figure 27, we have returned to the representation of the three regions in the Uncertainty Game: The main playing field (green), the end zones (yellow), and out-of-bounds (red). Unlike football, we **do not** want to be in the end zone or out of bounds. This anti-football approach would have us keep the ball in the middle of the field in order to win.

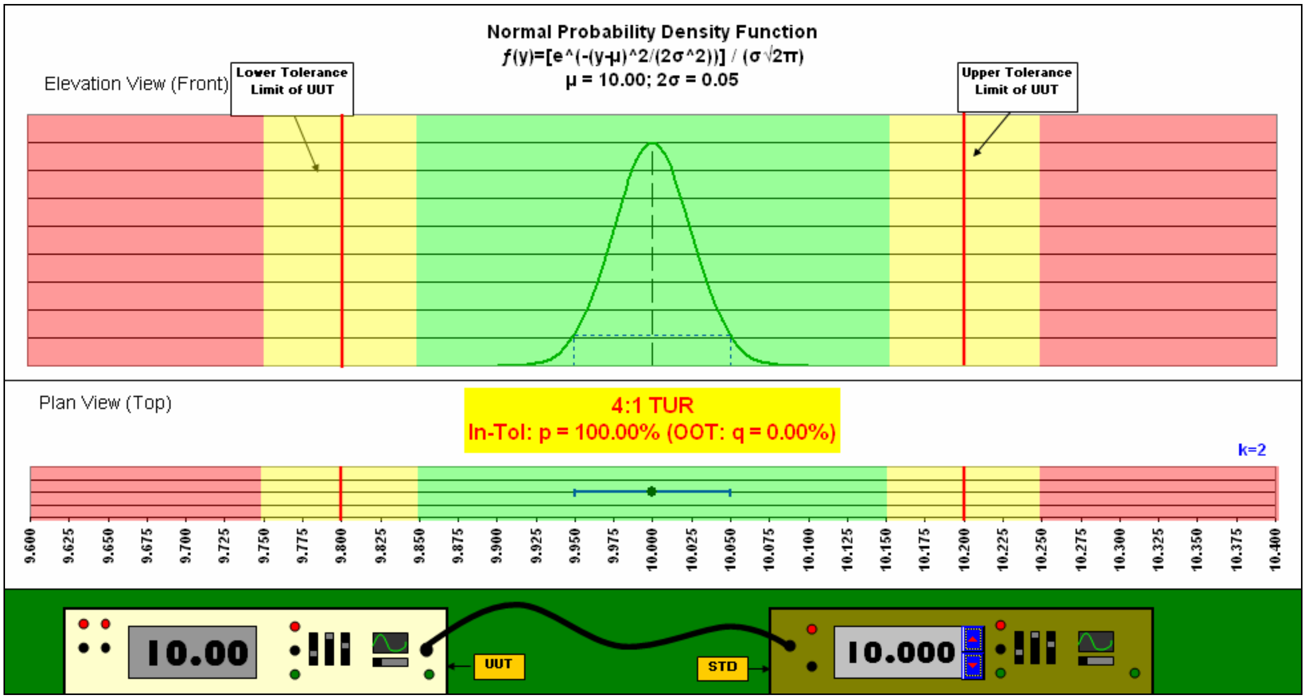


Figure 27: A new feature: In-Tolerance and OOT percentages

I have indicated our commitment to reporting at k=2 by indicating the limits of this area with a dotted line; the vertical lines indicating the limits at k=2 and the horizontal line showing the distance between the two lines. This also corresponds to the uncertainty bar shown in the Plan View, or blimp view, whichever you prefer. I have also added another feature (highlighted in yellow): the chance that the reading is In-Tolerance (p) and, conversely, the chance that the reading is OOT (q), both shown in percentage. More about this feature in a moment.

Let's go back to one of the outcomes we previously considered, but look at it from this Elevation View (fig. 28). In this situation, the uncertainty of the measurement (k=2) could mean that the UUT is actually at its upper tolerance limit (right end of k=2 limit) – or it might mean that the UUT is well within its tolerances (left end of k=2 limit). Either way, at k=2, we are assured that the UUT is not OOT. Hmmmm . . . is that right? If we only consider k=2, as many have done using the traditional Plan View, then we believe that the uncertainty has not yet crossed the UUT's upper tolerance threshold.

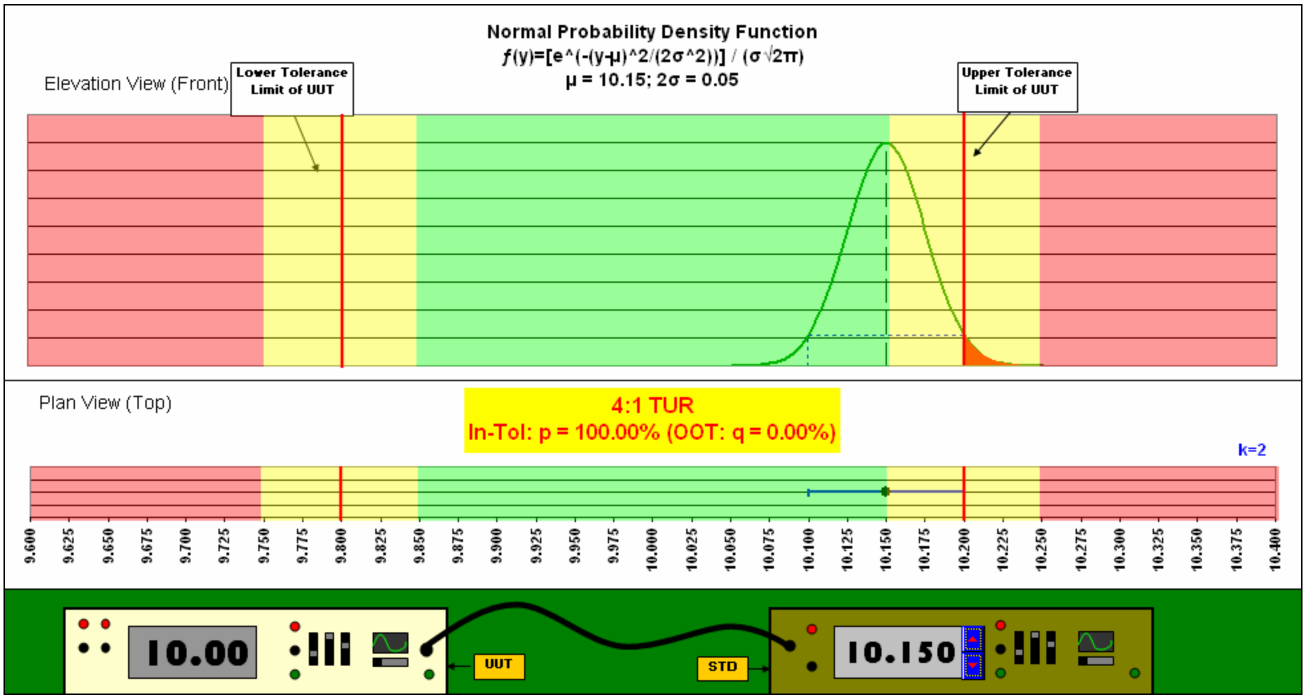


Figure 28: k=2 – Just entering the Indeterminate region?

As we view different outcomes in which the standard indicates values within the Indeterminate region, we see (from the Elevation View) that the area under the curve that is beyond the UUT's upper limit is shaded in red.

In the situation where the standard indicates the UUT is right on the edge of its upper tolerance limit (fig. 29), our 50-yard line vantage point indicates that there is a 50% chance the uncertainty could make the UUT value actually OOT as well as a 50% chance that the UUT value is In Tolerance. I used this example because it is easy to see that half of the curve extends beyond the upper tolerance of the UUT and half extends within the limits of the UUT.

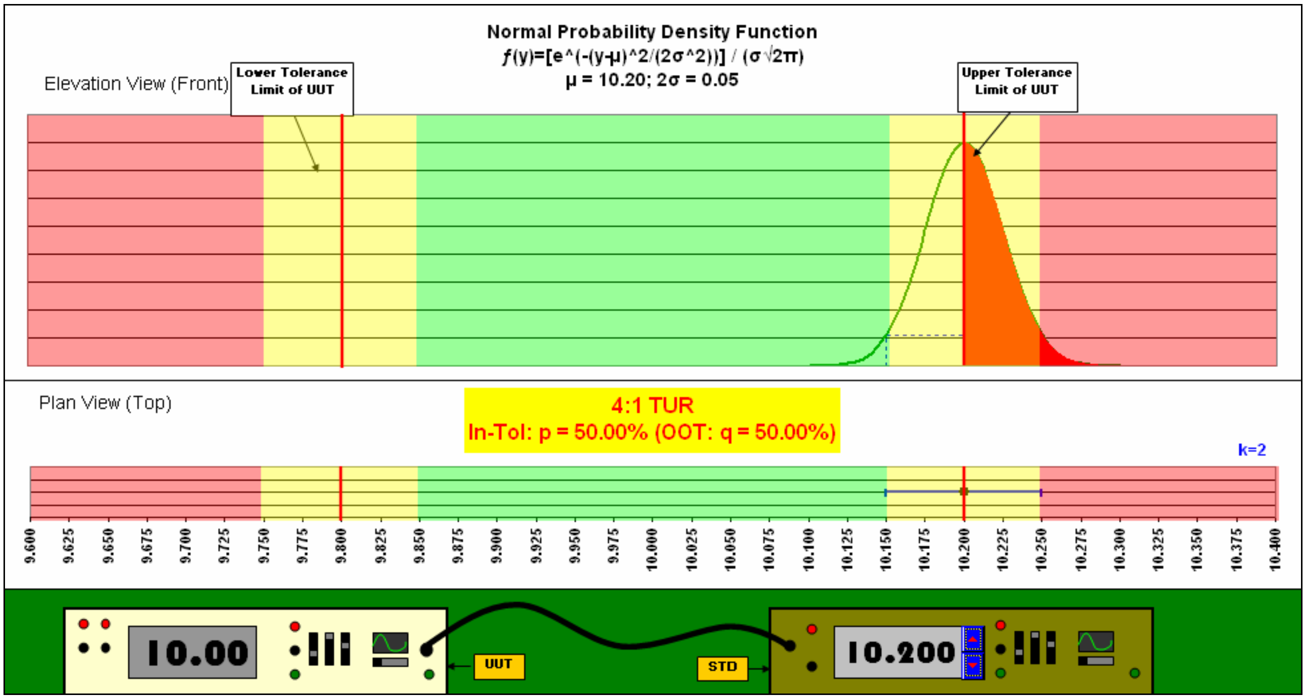


Figure 29: Standard indicates UUT's value is at the edge of its tolerance.

So, how do we calculate values in between? Standard-Normal tables (or z-Tables) are used to calculate the area under the curve of a Normal distribution, allowing us to slice the distribution many different ways. In figure 29, the area of the curve that lies beyond the upper tolerance limit of the UUT is obviously 50%. The z-Tables indicate this also. But what about the outcome shown in figure 30?

Here, the standard indicates that the signal being generated by the UUT is 10.180 MHz. The 50-yard view shows that there is a 20% chance that this value is really OOT, based on the uncertainty surrounding the measurement process. These z-Tables make it possible to calculate the area of the curve that lies to the right of the upper tolerance limit. On the other hand, we can state that there is an 80% chance that the UUT is within tolerance! But wait, there is a problem with this, and it is unintentionally caused by the k=2 approach to reporting uncertainties.

Looking back at figure 28, you'll notice that the right tail of the normal distribution lies beyond the UUT's upper tolerance limit, yet there is no indication of a chance that this is out of tolerance (p=100%; q=0%). That's because the traditional Metrology view (the Plan View) of this situation focuses on the uncertainty bar using a value of k=2. This view inadvertently disguises the truth of the matter in that there exists a probability, small as it may be, that an event could occur in which the standard's true value really lies somewhere in this tail area. So why didn't the z-Tables reflect this in the In Tolerance and OOT percentages? Is there something wrong with them? Did I make a mistake in the formulas I used? No, the z-Tables and formulas are fine. However, in the preceding illustrations I had purposely *normalized* the z-Table values and formulas to work solely within the k=2 region, rather than the entire area under the curve, so that it would match the traditional practice of using only k=2. This was done to underscore the point I am making about this practice and the natural assumptions that go with it.

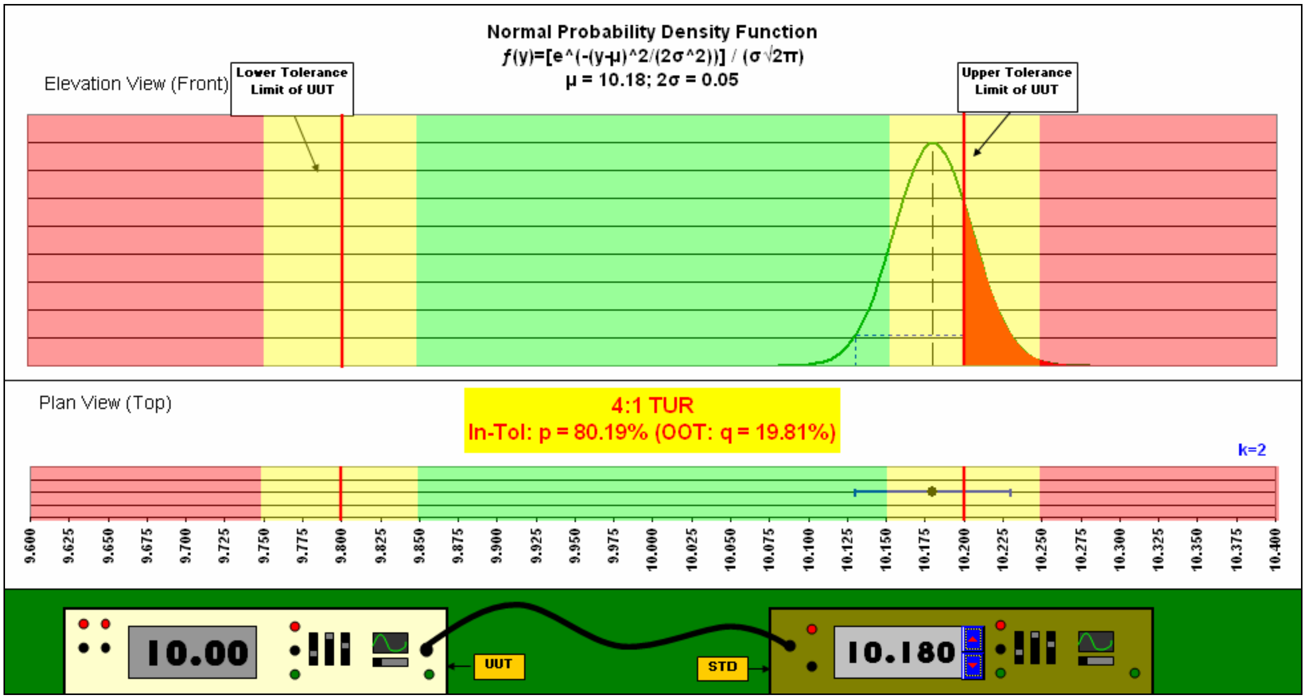


Figure 30: Uncertainty effect on OOT probability at 10.180 MHz.

Not considering all events of probability is like rounding off results at each step in a series of equations. It's generally not a good practice to do until you've arrived at the final step because rounding too soon will often give you erroneous results. The natural occurrence of people in the industry has been to use the uncertainty bar at k=2 to establish guardband limits, yet there are probabilities beyond these limits that could put the UUT's value out of tolerance. Isn't this an informal, and unintentional, rounding approach? So, I reset the z-Tables and formulas back to their original values to indicate all probable events, thereby showing the indications that lie within the Indeterminate region and their true In Tolerance probabilities for all the remaining graphs in this paper.

In figure 31, you see that I've maintained the k=2 zone for comparison (in blue), but have added a k=3.9 zone (in orange) to indicate that, this time, we're going to consider the probability of 99.999% of all events, not just the probability of events within the k=2 range. Note also that the Safeband has been reduced and the Indeterminate regions have grown. This is representative of the k=3.9 zone.

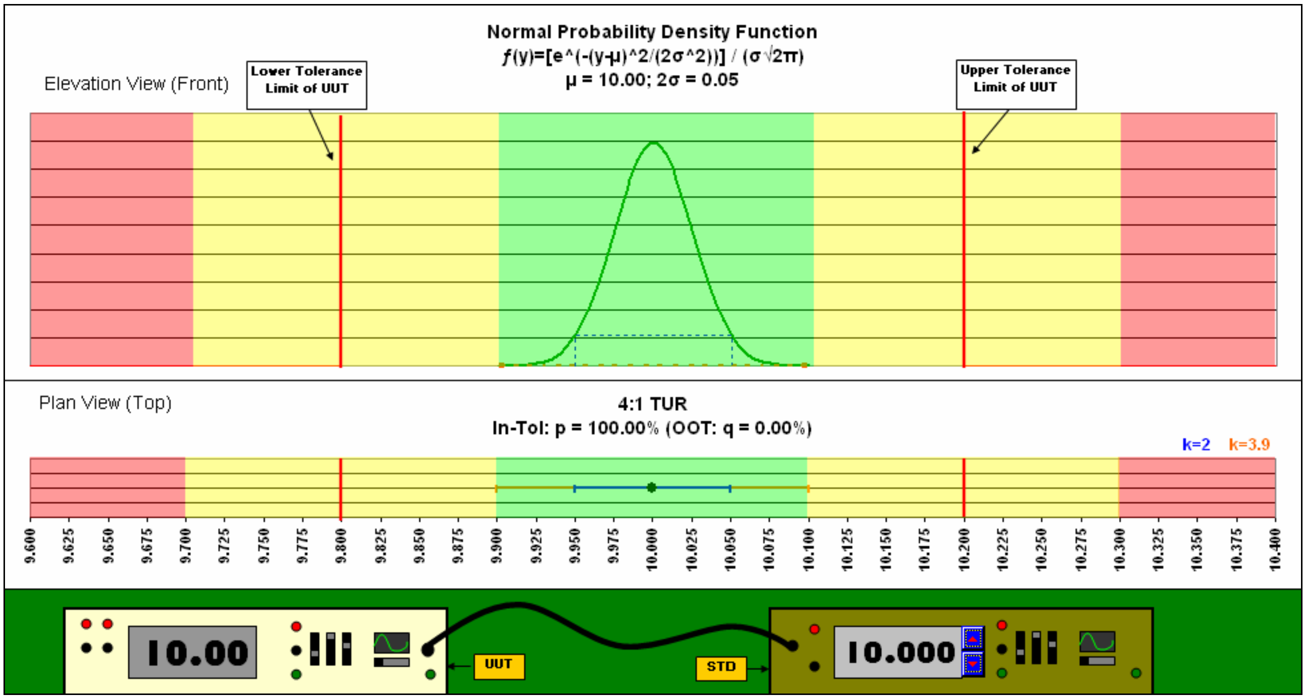


Figure 31: Contrasting k=2 vs. k=3.9 and the effect on reporting OOT conditions.

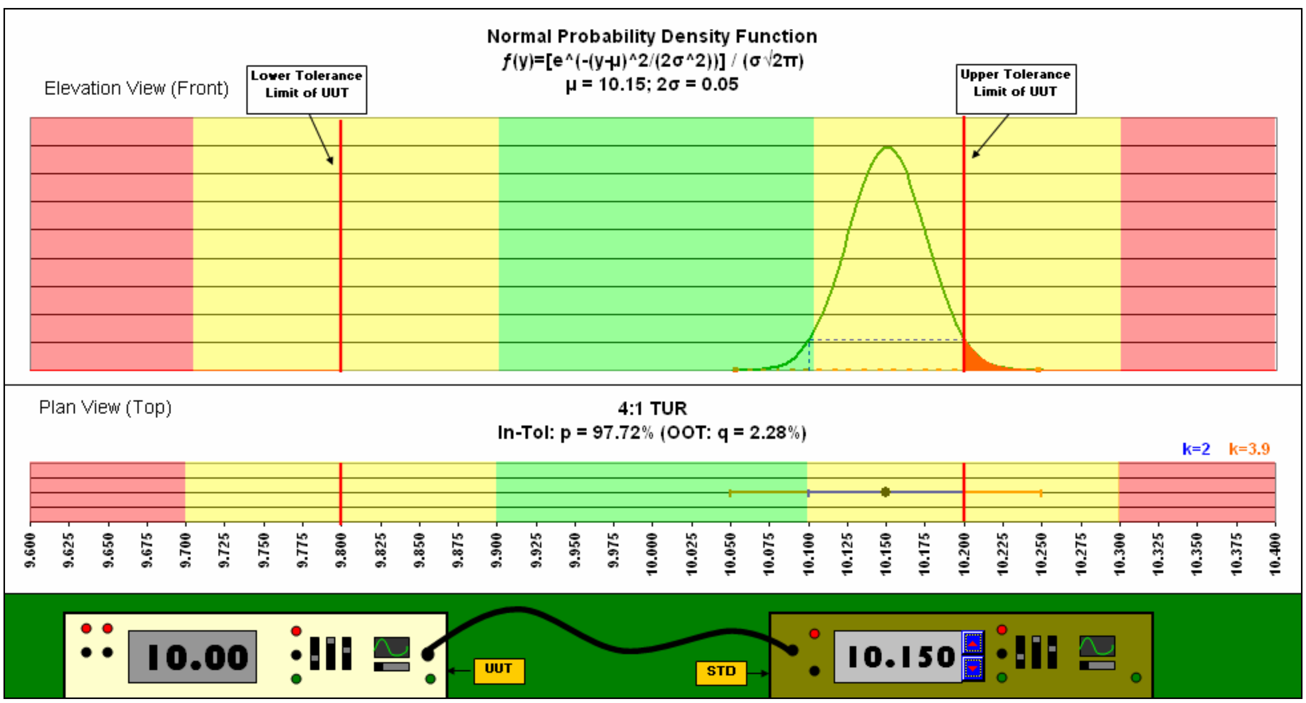


Figure 32: k=2 vs.k=3.9: Right-end of k=2 uncertainty bar at UUT's upper tolerance limit.

Figure 32 is a repeat of fig.28, but with the correct z-Tables and formulas. Here, the k=2 range indicates that the right end of the uncertainty is at the upper tolerance limit. This implies that there is a 100% chance of an In-Tolerance reading and a 0% chance of an OOT condition. However, the k=3.9 range shows that there is really a 97.7% chance that this reading is actually In Tolerance and a 2.3% chance that this reading is actually OOT, when considering all probable events that represent the uncertainty of the measurement process.

Looking at an outcome on the other side of the upper limit, figure 33 shows that, when the standard indicates that the UUT's value is 10.250 MHz, considering the measurement uncertainty there is a 2.3% chance that the UUT is actually In Tolerance (97.7% chance of OOT).

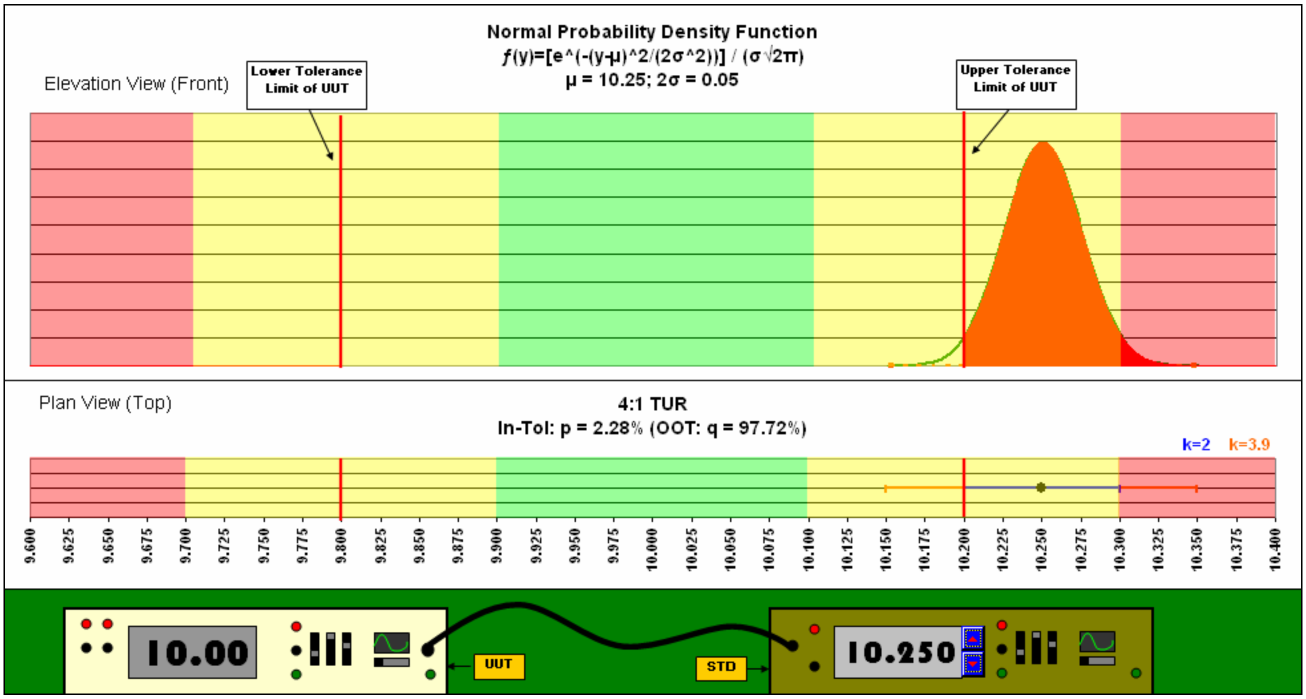


Figure 33: k=2 vs.k=3.9: Left-end of k=2 uncertainty bar at UUT's upper tolerance limit.

The traditional viewpoint looked not at percentages, but at whether or not the value lies within the Indeterminate region. This inability to determine an In or Out of Tolerance condition has led some companies to remove statements of compliance from their certificates of calibration (i.e., they have removed the warning flags that the customer should be able to rely upon). I suggest that, by utilizing the Elevation View (50-Yard Line) the Metrology industry can make compliance statements, using k=3.9, as follows:

1. If the reading is within the Safeband, then the probability of an In-Tolerance reading is 100%.
2. If the reading is within the Indeterminate region, then the probability of an In-Tolerance reading can be determined using Standard-Normal tables (z-Tables).
3. If the reading is within the OOT band, then the probability of an In-Tolerance reading is 0%.

In all three situations, there is an In Tolerance (p) and OOT (q) probability that can easily be reported for every measurement included on a calibration certificate. This value will slide from p=100% (q=0%) to p=0% (q=100%), depending on where the measurement lies with respect to the UUT's tolerances and also dependent upon the uncertainty of the measurement at k=3.9. The effect on the calibration report will be that, along with the reported measurement of the UUT's error and the uncertainty of the measurement, there would also be a statement of compliance to the accuracy specification given as a percentage of compliance. I will refer to this as the Probability of Compliance to the Specification, or PCS. Since we are talking about the compliance to the spec (and not the non-compliance), then we would report on the "p" value, or In-Tolerance probability (and not the "q" value, or OOT probability). It might look something like the data shown in figure 34:

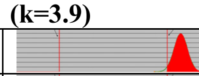
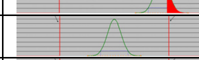
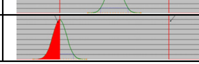
Nominal (UUT)	Lower Tolerance	Upper Tolerance	"As Found"	Error	Uncertainty (k=2)	TUR (k=2)	PCS* (k=3.9)	
10.00 MHz	9.80 MHz	10.20 MHz	10.25 MHz	-0.25 MHz	0.050 MHz	4:1	2%	

* PCS = Probability of Compliance to the Specification (see attachment for details)

Figure 34: Adding a PCS statement to calibration reports.

Notice that we can still report the uncertainty at k=2, and we can still calculate the TUR using k=2, but the PCS would be determined using k=3.9 in order to compare all probabilities associated with the measurement process to the tolerances of the UUT. I had originally thought that the industry might want to change the standardized reporting format from k=2 to k=3.9. But, as long as the PCS is calculated using k=3.9, there is really no need to change this, especially since there are an inestimable number of quality documents that would need to be changed to reflect this. So I suggest that, unless someone can give a very good reason to change standardized uncertainty reporting, we leave this alone and focus on the addition of the PCS statement to calibration reports. Since this is a new measure of confidence in the reported reading, we may also want to include a document that explains the PCS and what it means to the customer.

From the customer's standpoint, if the PCS is 100%, there is no cause for alarm (i.e., no warning flag). Any indication other than this would give them the sign that a reverse traceability to their product is required in order to minimize risk in their quality system (which addresses both Producer and Consumer risk). A number of measurements included in the report might look something like fig. 35:

Nominal (UUT)	Lower Tolerance	Upper Tolerance	"As Found"	Error	Uncertainty (k=2)	TUR (k=2)	PCS* (k=3.9)	
10.00 MHz	9.80 MHz	10.20 MHz	10.25 MHz	-0.25 MHz	0.050 MHz	4:1	2%	
50.00 MHz	49.80 MHz	50.20 MHz	50.18 MHz	-0.18 MHz	0.050 MHz	4:1	79%	
100.0 MHz	99.8 MHz	100.2 MHz	100.0 MHz	0.0 MHz	0.050 MHz	4:1	100%	
200.0 MHz	199.8 MHz	200.2 MHz	199.8 MHz	+0.2 MHz	0.050 MHz	4:1	50%	

* PCS = Probability of Compliance to the Specification (see attachment for details)

Figure 35: Calibration Report – Multiple test points with PCS statement added.

This application of uncertainty and statistics puts a finishing touch, or capstone, on the traceable chain of measurement and its meaning to commerce. But before you sign up to this proposal, or reject it, let's consider what happens to other levels of TUR.

To illustrate a 3:1 TUR situation, we'll assume that the best standard we could find has an uncertainty (at k=2) of 0.067 MHz. The TUR is (0.2 ÷ 0.067 = 3), or 3:1. As you can see in figure 36, this causes the Safeband® to reduce even further and the Indeterminate region to increase. In other words, the PCS will result in a 100% value over a shorter range than with a higher TUR. To illustrate this, a few examples are given for the 3:1 scenario in figures 36 through 50.

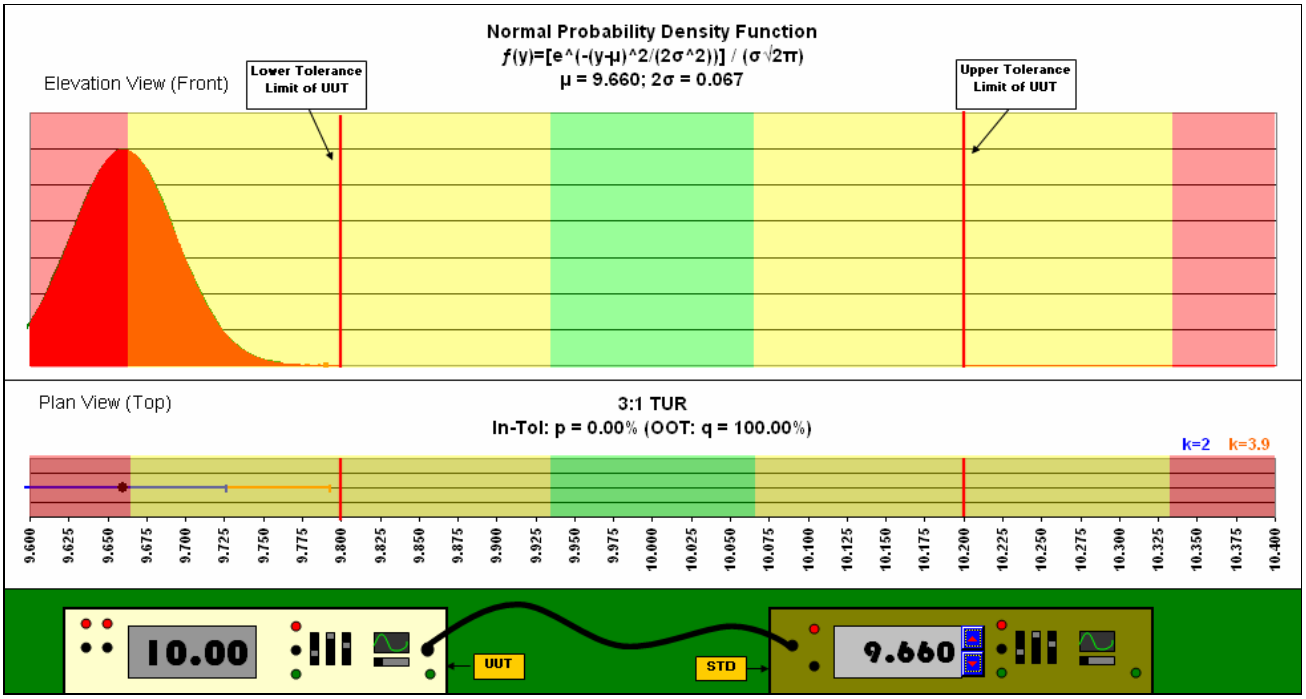


Figure 36: 3:1 TUR – Standard indicates UUT's actual value is 9.660 MHz, which is below the lower Indeterminate region (PCS = 0%).

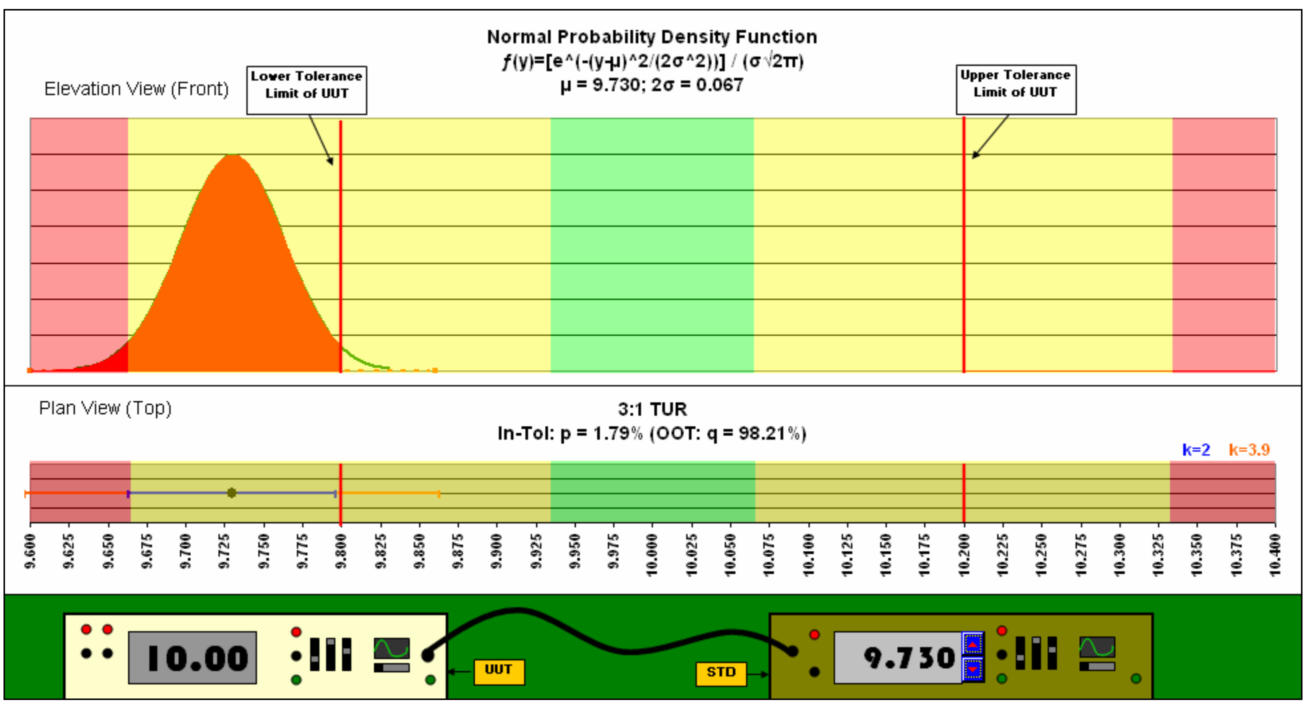


Figure 37: 3:1 TUR – Standard indicates UUT's value is 9.730 MHz, which is within the lower Indeterminate region (PCS = 1.8%).

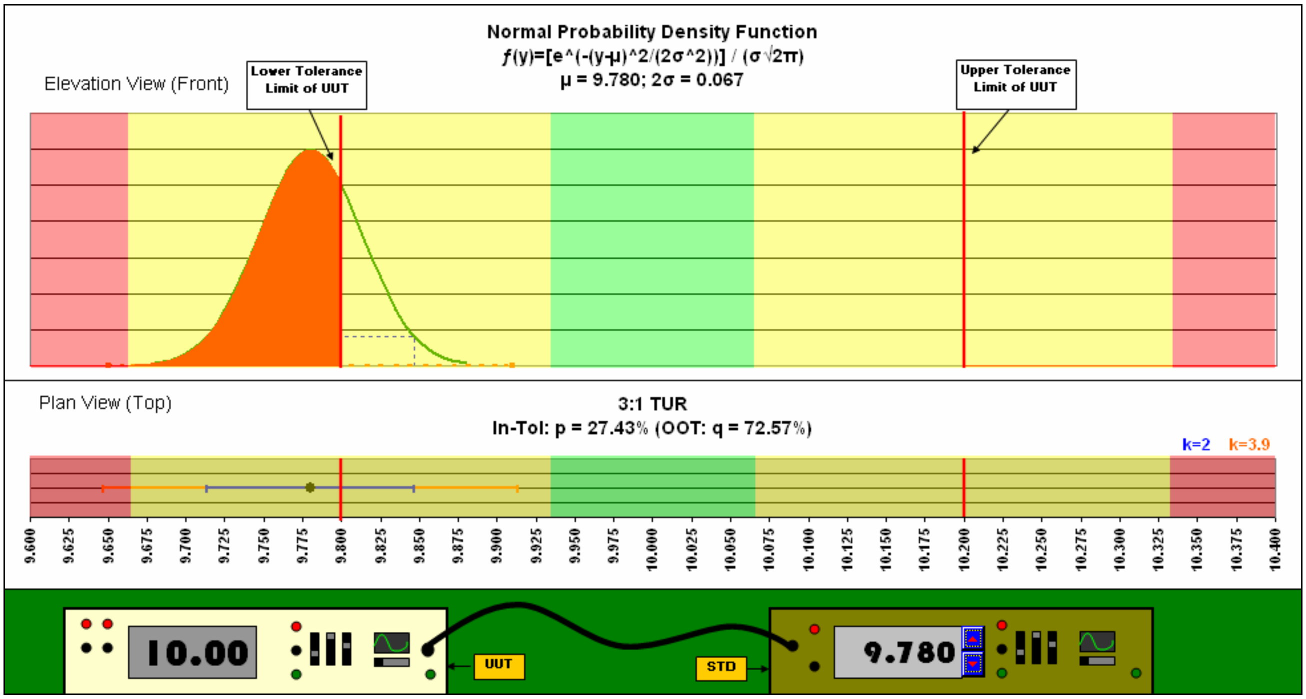


Figure 38: 3:1 TUR – Standard indicates UUT's value is 9.780 MHz, which is within the lower Indeterminate region (PCS = 27.4%).

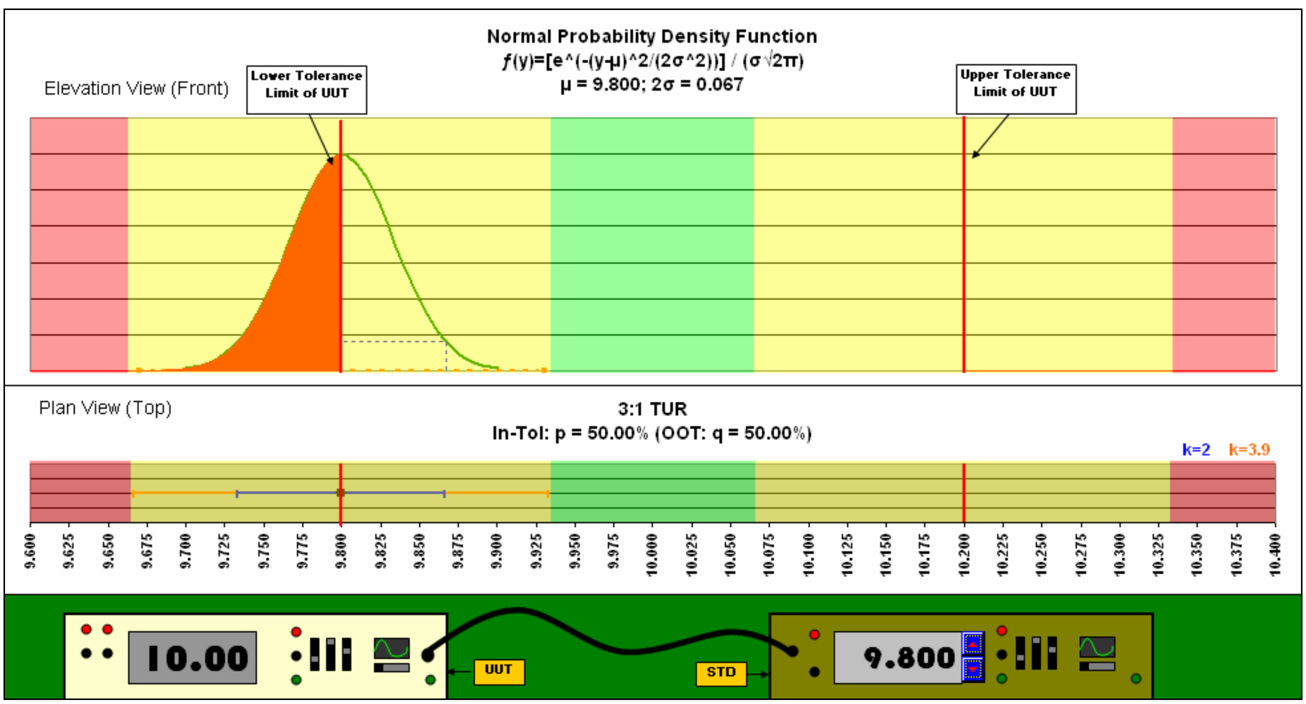


Figure 39: 3:1 TUR – Standard indicates UUT's value is 9.800 MHz, which is the lower tolerance limit (PCS = 50.0%).

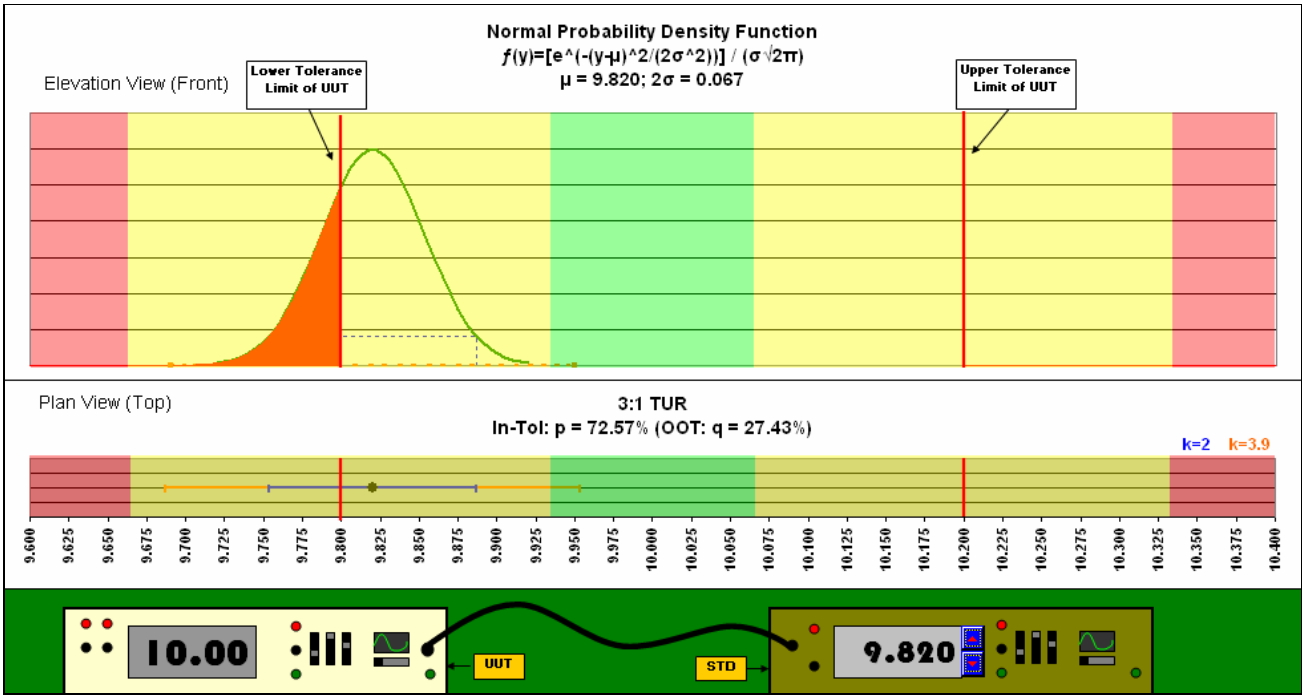


Figure 40: 3:1 TUR – Standard indicates UUT's value is 9.820 MHz, which is within the lower Indeterminate region (PCS = 72.6%).

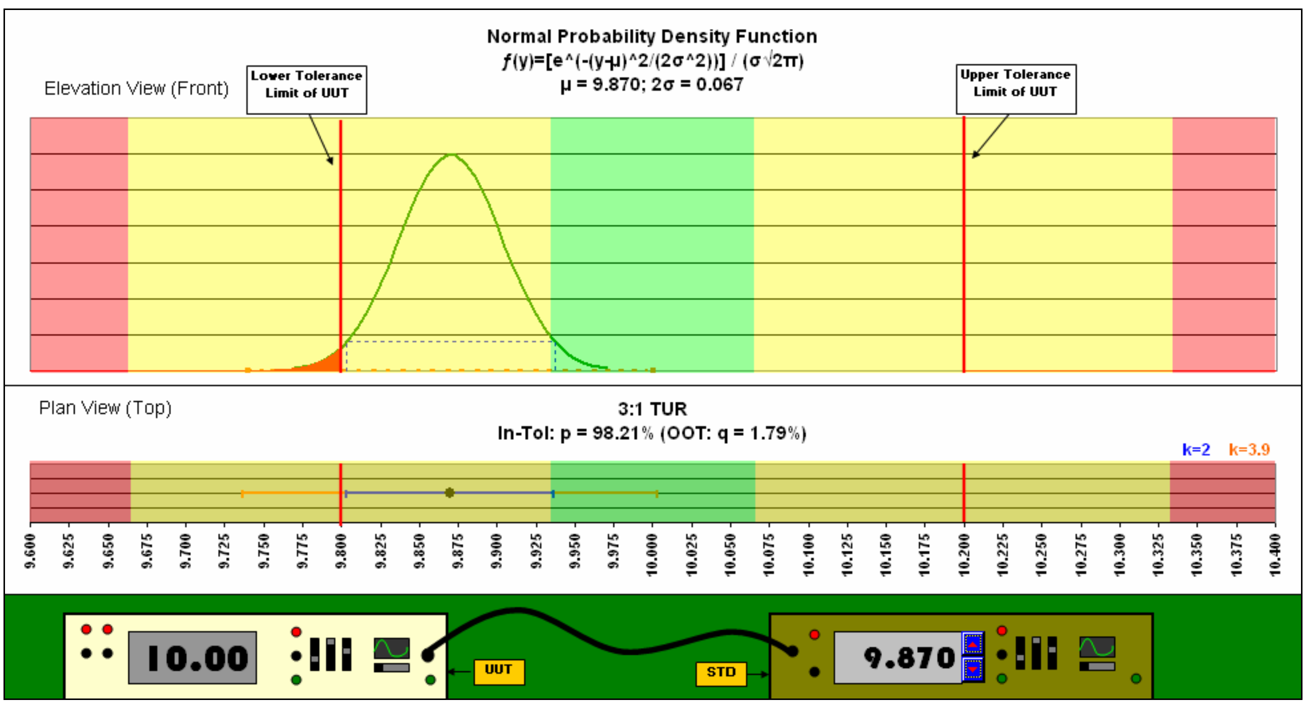


Figure 41: 3:1 TUR – Standard indicates UUT's value is 9.870 MHz, which is within the lower Indeterminate region (PCS = 98.2%).

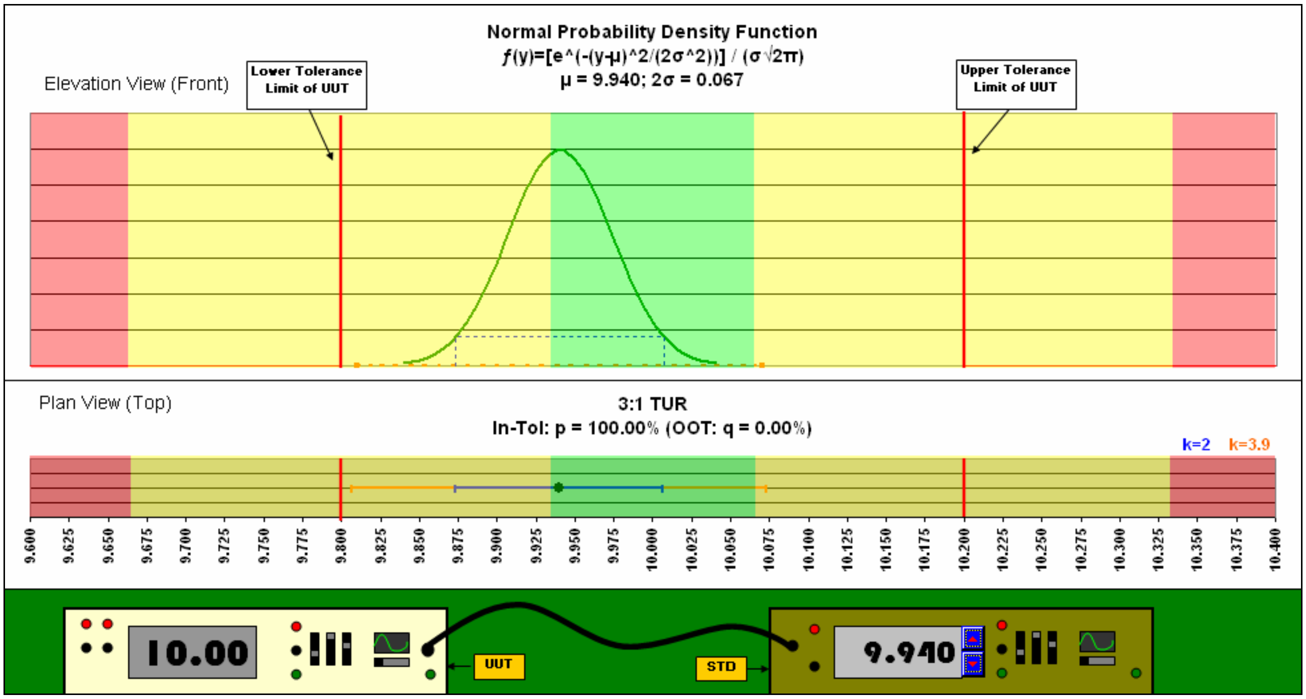


Figure 42: 3:1 TUR – Standard indicates UUT's value is 9.940 MHz, which is at the left end of the Safeband® region (PCS = 100.0%).

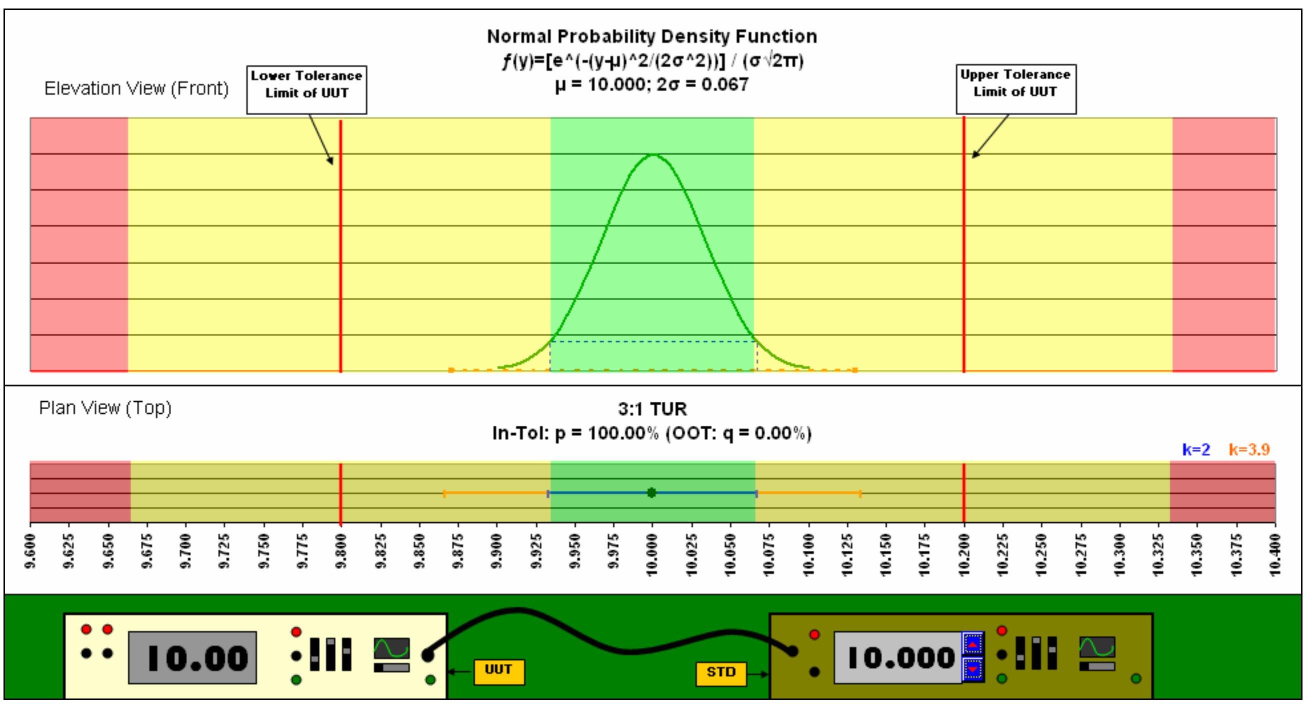


Figure 43: 3:1 TUR – Standard indicates the UUT's value is 10.000 MHz, which is at nominal (PCS = 100.0%).

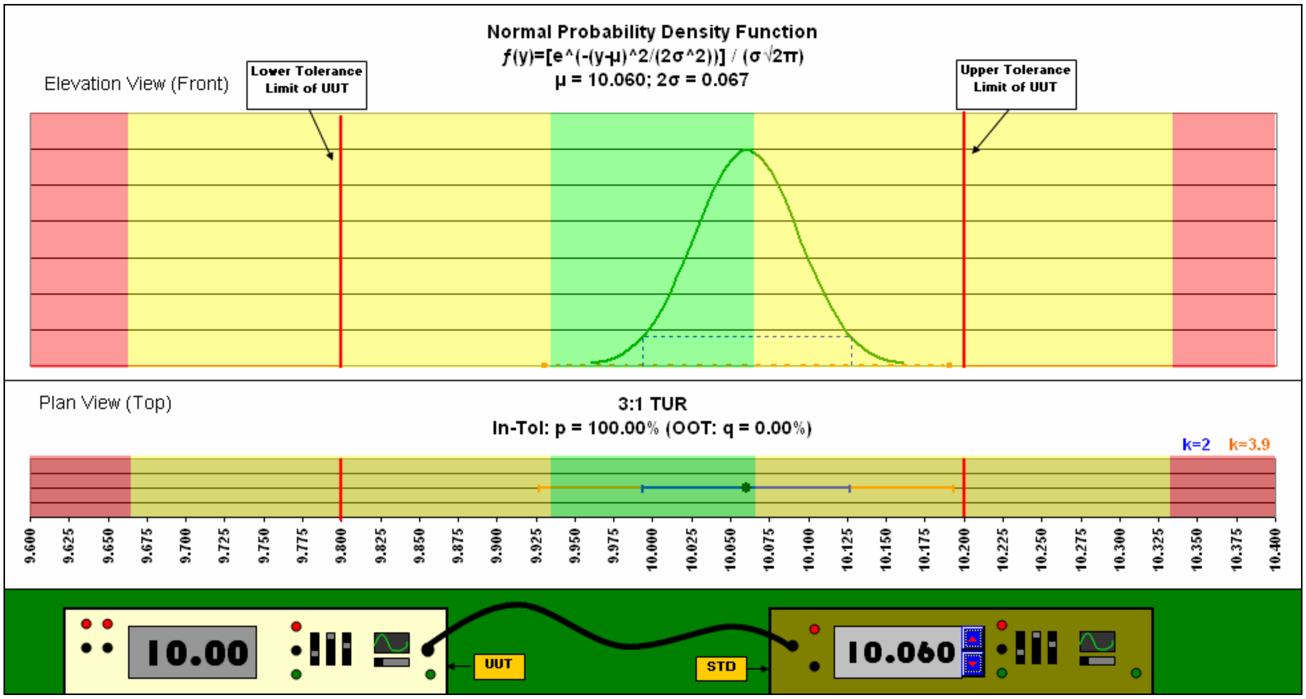


Figure 44: 3:1 TUR – Standard indicates UUT's value is 10.060 MHz, which is at the right end of the Safeband® region (PCS = 100.0%).

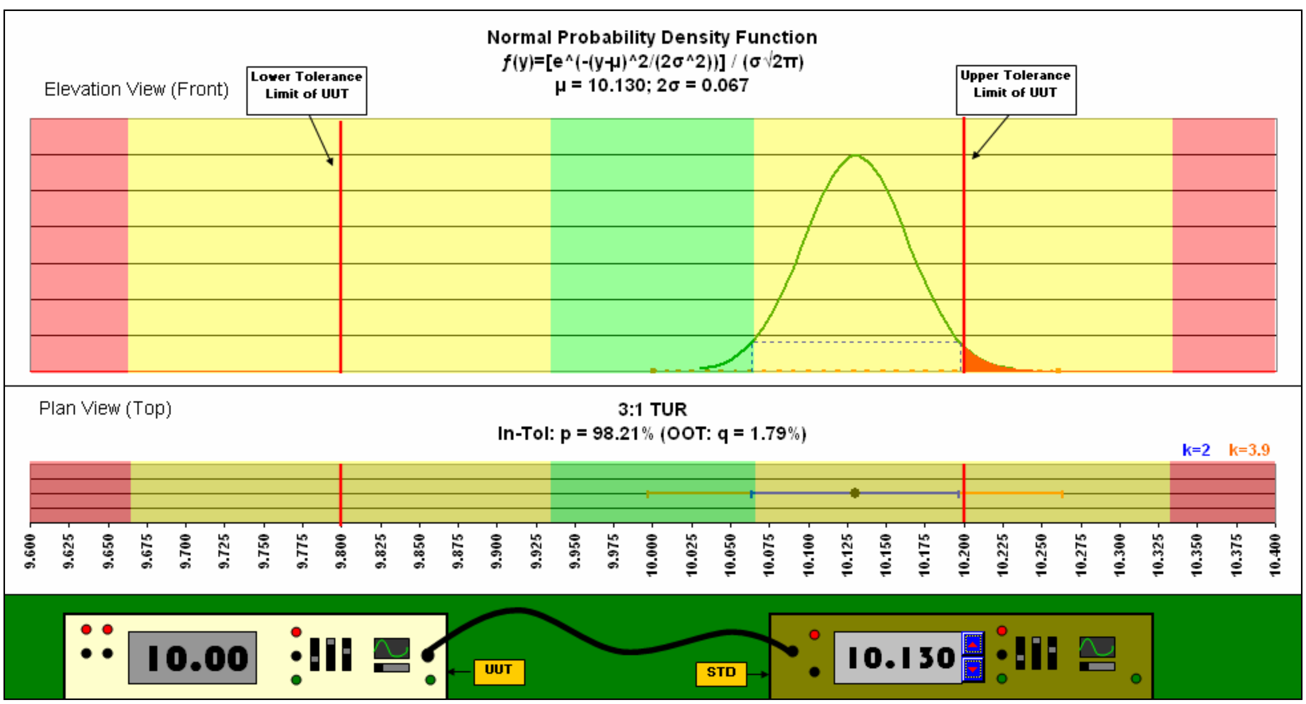


Figure 45: 3:1 TUR – Standard indicates UUT's value is 10.130 MHz, which is within the upper Indeterminate region (PCS = 98.2%).

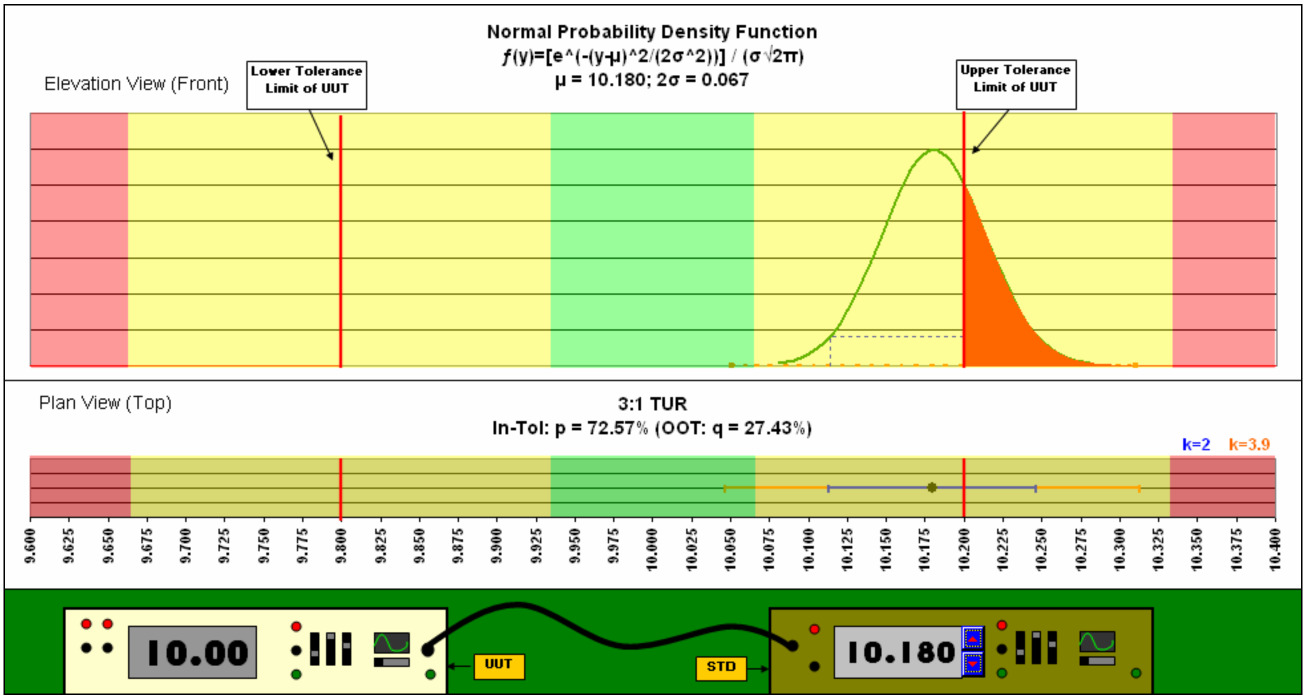


Figure 46: 3:1 TUR – Standard indicates UUT's value is 10.180 MHz, which is within the upper Indeterminate region (PCS = 72.6%).

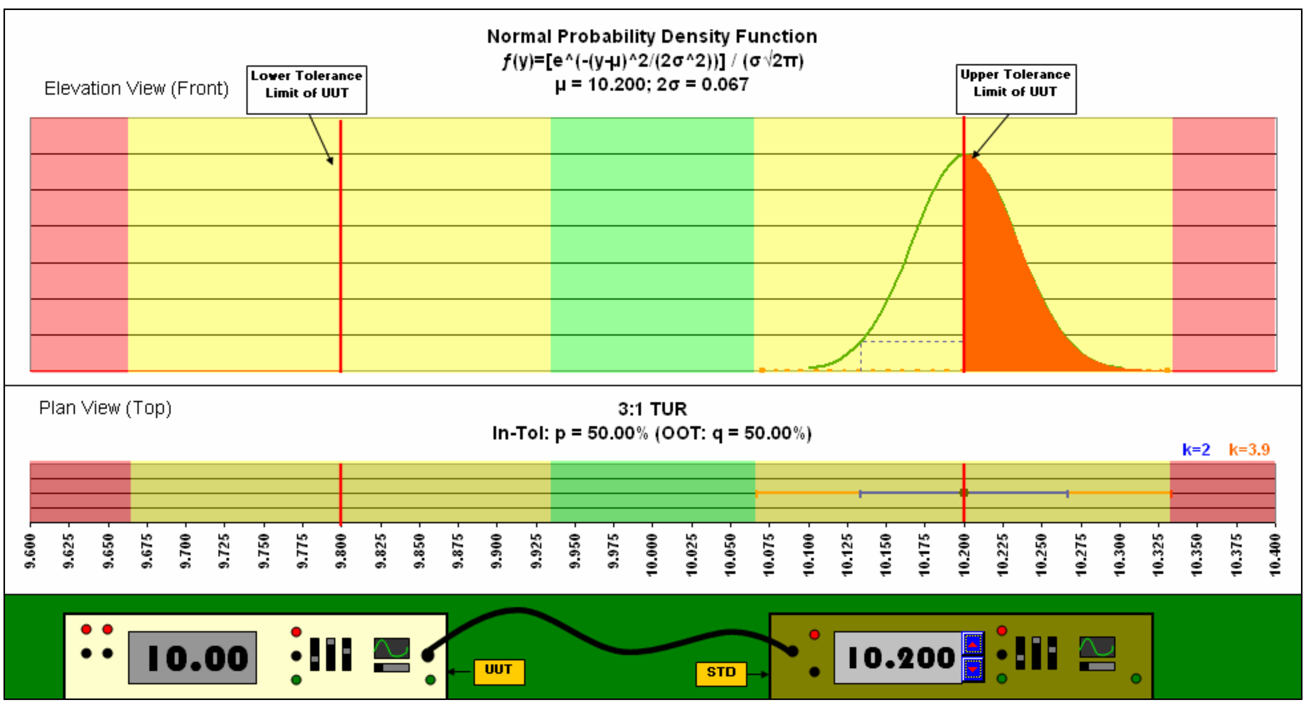


Figure 47: 3:1 TUR – Standard indicates UUT's value is 10.200 MHz, which is at the upper tolerance limit (PCS = 50.0%).

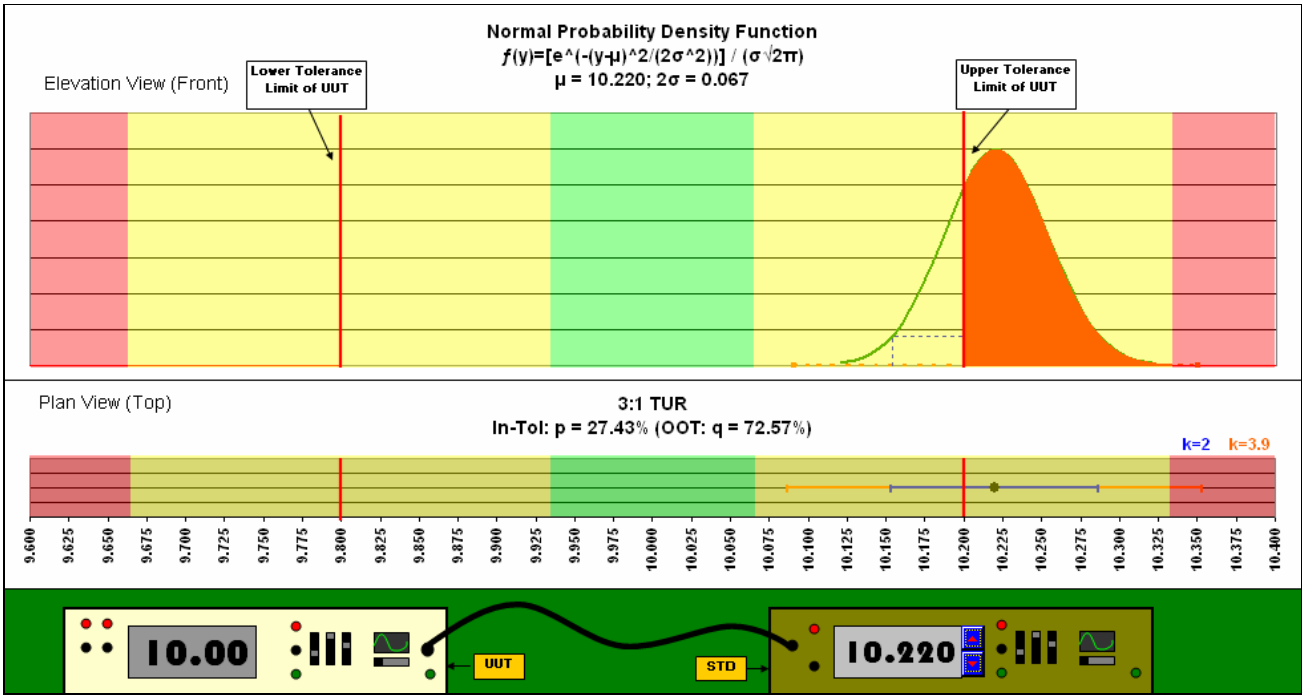


Figure 48: 3:1 TUR – Standard indicates UUT's value is 10.220 MHz, which is within the upper Indeterminate region (PCS = 27.4%).

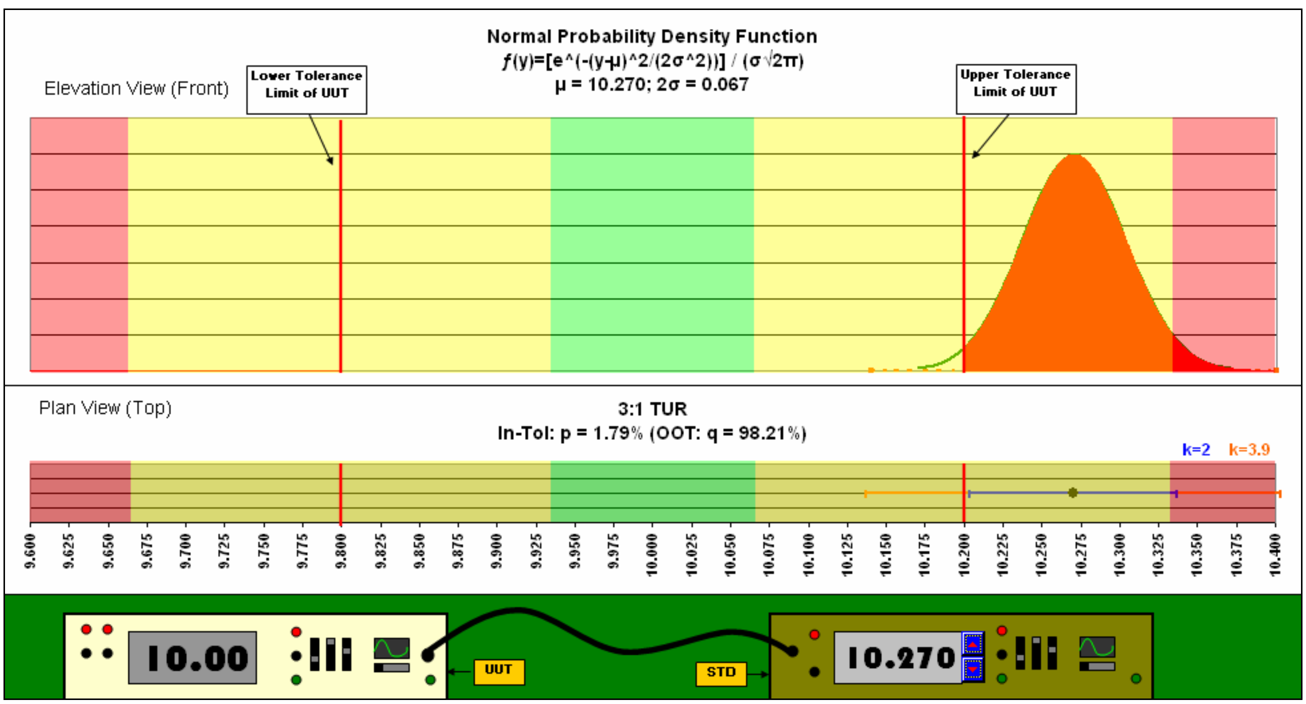


Figure 49: 3:1 TUR – Standard indicates UUT's value is 10.270 MHz, which is within the upper Indeterminate region (PCS = 1.8%).

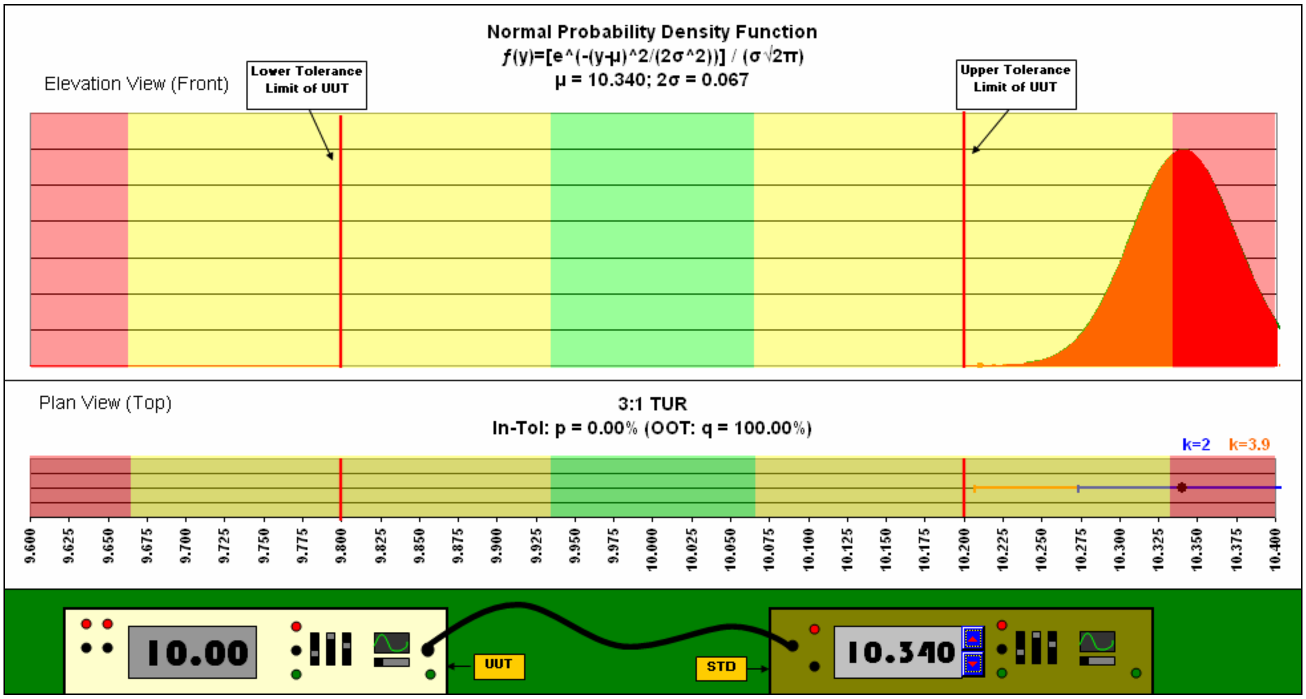


Figure 50: 3:1 TUR – Standard indicates UUT's value is 10.340 MHz, which is above the upper Indeterminate region (PCS = 0%).

As you can see from these illustrations, whenever the standard indicates that the UUT's value is on either the lower or upper tolerance limits, the PCS is always 50%, regardless of the TUR. This should makes sense though because the distribution representing the uncertainty of the measurement will always be evenly split at this point. The difference lies in the spread of the distribution; the greater the uncertainty, the further the end points of the distribution at k=3.9. This also means that the lower the TUR, the further OOT the UUT may be, which increases the risk to the customer. Again, this illuminates the need for a simpler way of telling the customer there may be trouble with their process without them having to understand the details of uncertainties and TURs in relation to the UUT's specification.

To illustrate a 2:1 TUR situation, we'll assume that the best standard we could find has an uncertainty (at k=2) of 0.100 MHz. The TUR is (0.2 ÷ 0.100 = 2), or 2:1. As you can see in figures 51 through 57, this causes the Safeband® to reduce even further and the Indeterminate region to increase. However, with a 2:1 TUR, the Safeband is now reduced to only one point: **nominal!** At a 2:1 TUR, if the UUT drifts from nominal at all, it enters the realm of the Indeterminate region where this larger uncertainty associated with the measurement chokes out the Safeband region, increasing the probability of an OOT condition. I've included only 7 graphs to illustrate this point; however the tool that I have designed to show this will indicate any of the points related to this measurement.

Because the indeterminate range is so much larger, the PCS statements are even more valid, as these values will be less than 100% at any point off of nominal over a wider range of readings. Granted, there is not any appreciable degradation of the PCS value until the UUT reads about 9.900 MHz at the lower end (97.7%) and equally at the upper end (10.100MHz; PCS = 97.7%). But this ability to quantify the statement of compliance will standardize the way in which the application of uncertainty against a UUT's tolerances is reported. This example, and the next

(1:1 TUR), just supports the reason why measurement ratios at less than 4:1 are not a good idea. The PCS statements simply quantify this concept.

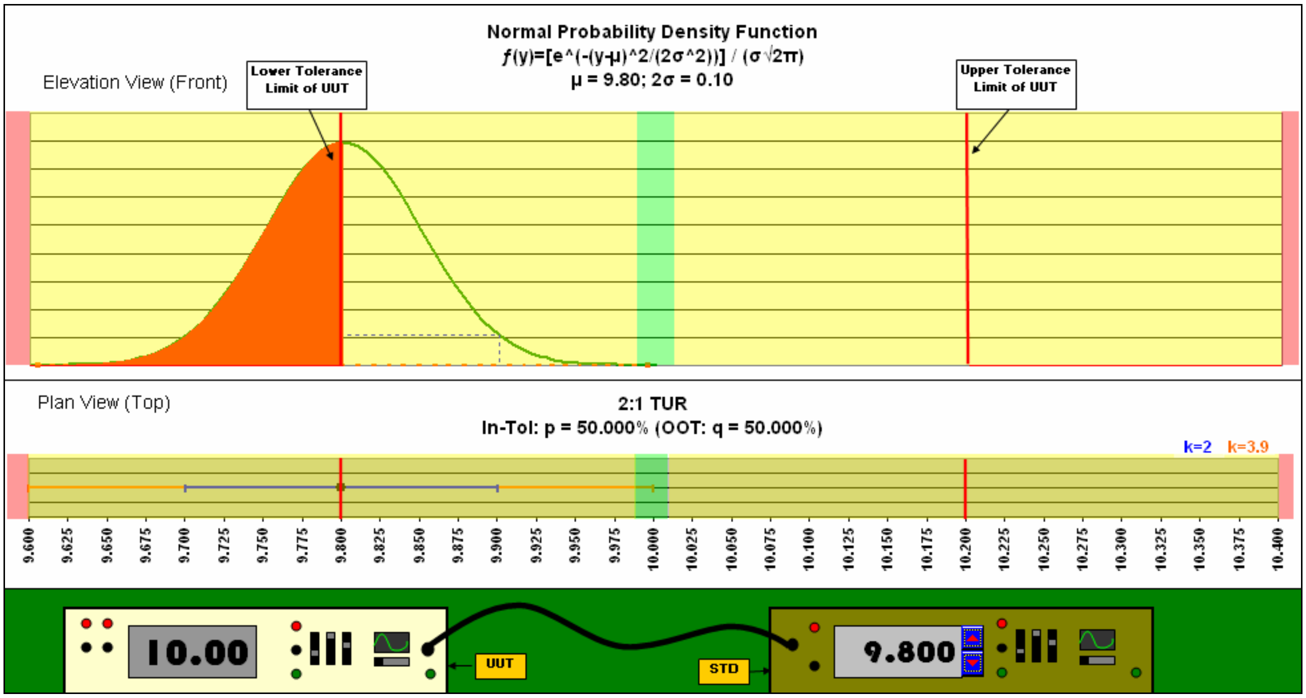


Figure 51: 2:1 TUR – Standard indicates UUT's value is 9.800 MHz, which is at the lower tolerance limit (PCS = 50.0%).

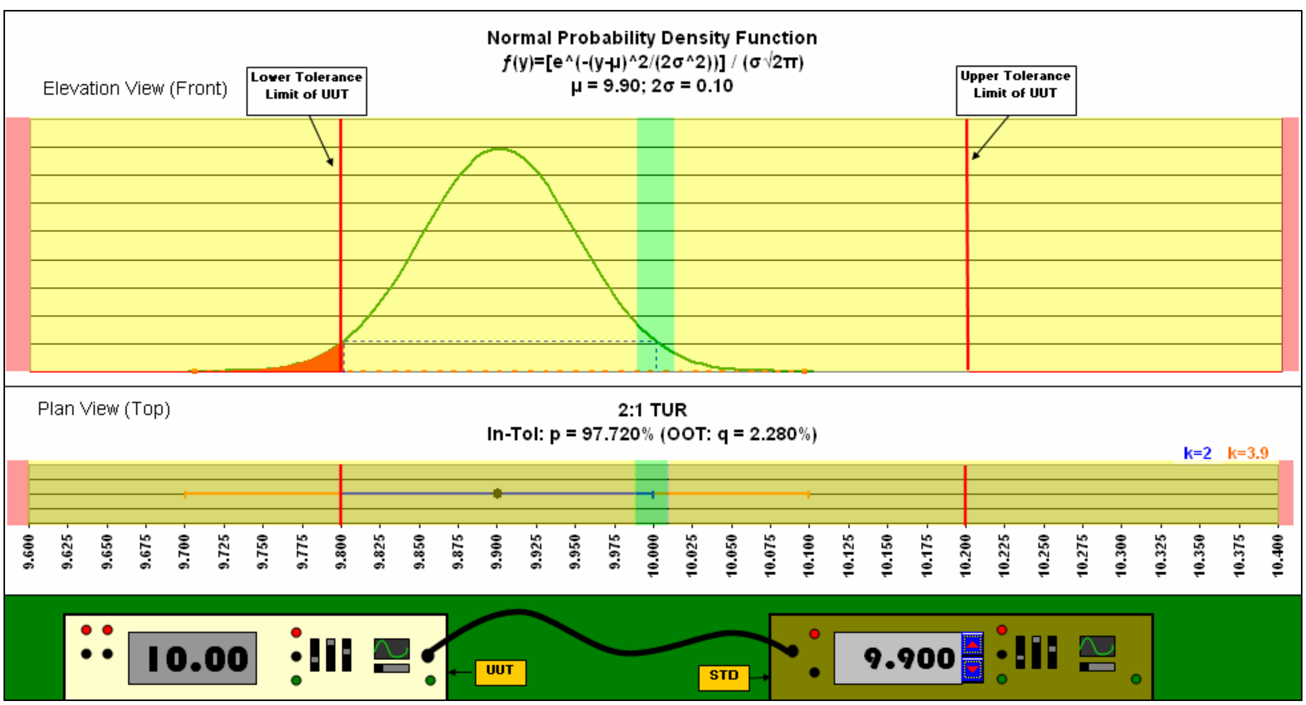


Figure 52: 2:1 TUR – Standard indicates UUT's value is 9.900 MHz, which is within the lower Indeterminate region (PCS = 97.7%).

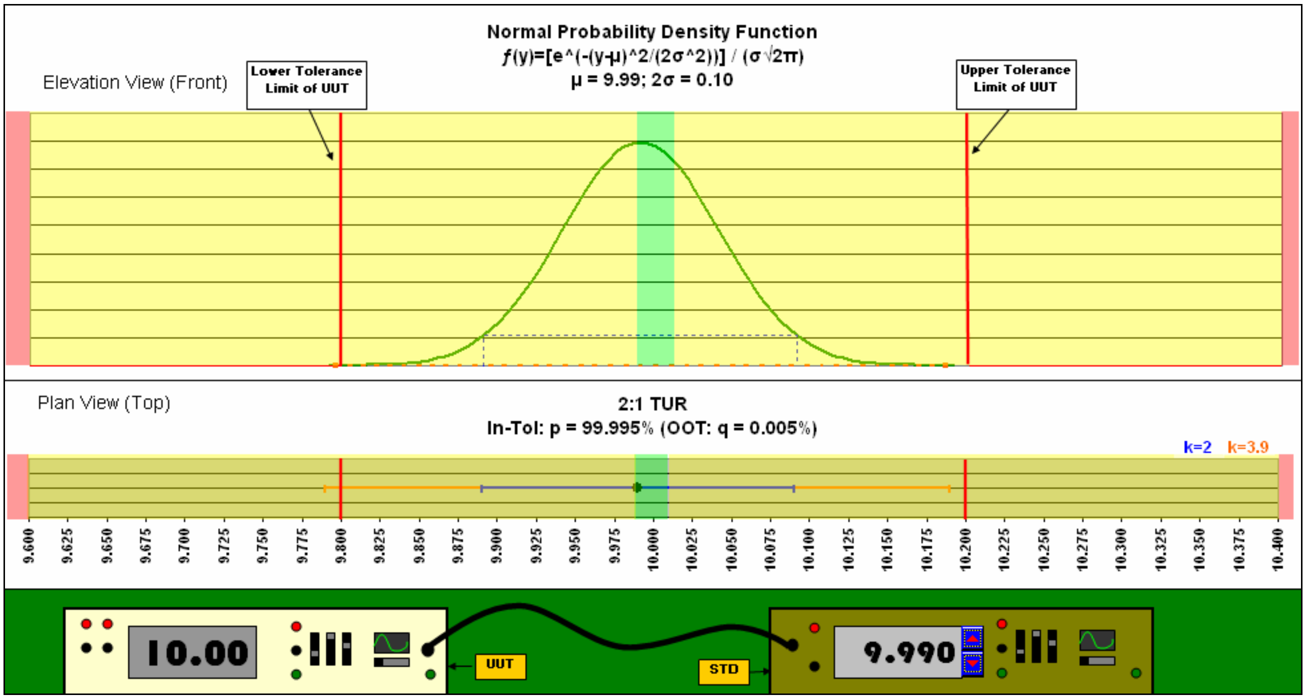


Figure 53: 2:1 TUR – Standard indicates UUT's value is 9.990 MHz, which is within the lower Indeterminate region (PCS = 99.995%).

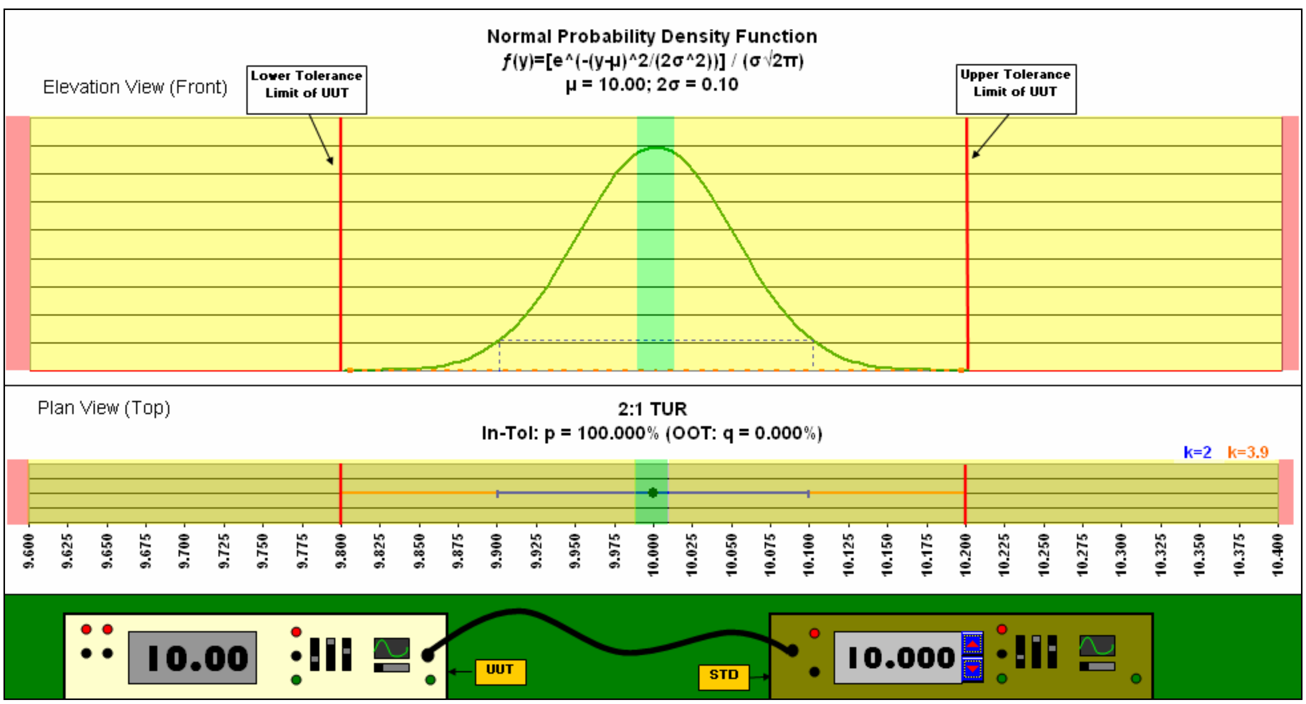


Figure 54: 2:1 TUR – Standard indicates UUT's value is 10.000 MHz, which is at nominal (PCS = 100.0%).

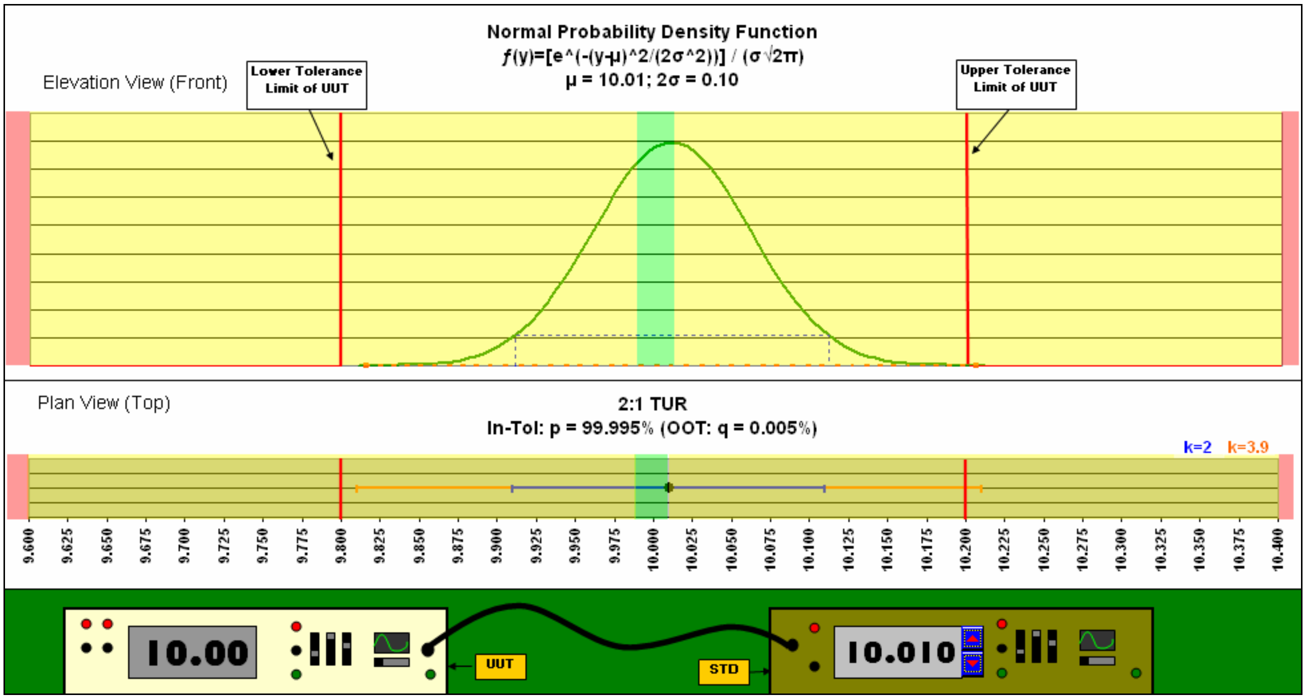


Figure 55: 2:1 TUR – Standard indicates UUT's value is 10.010 MHz, which is within the upper Indeterminate region (PCS = 99.995%).

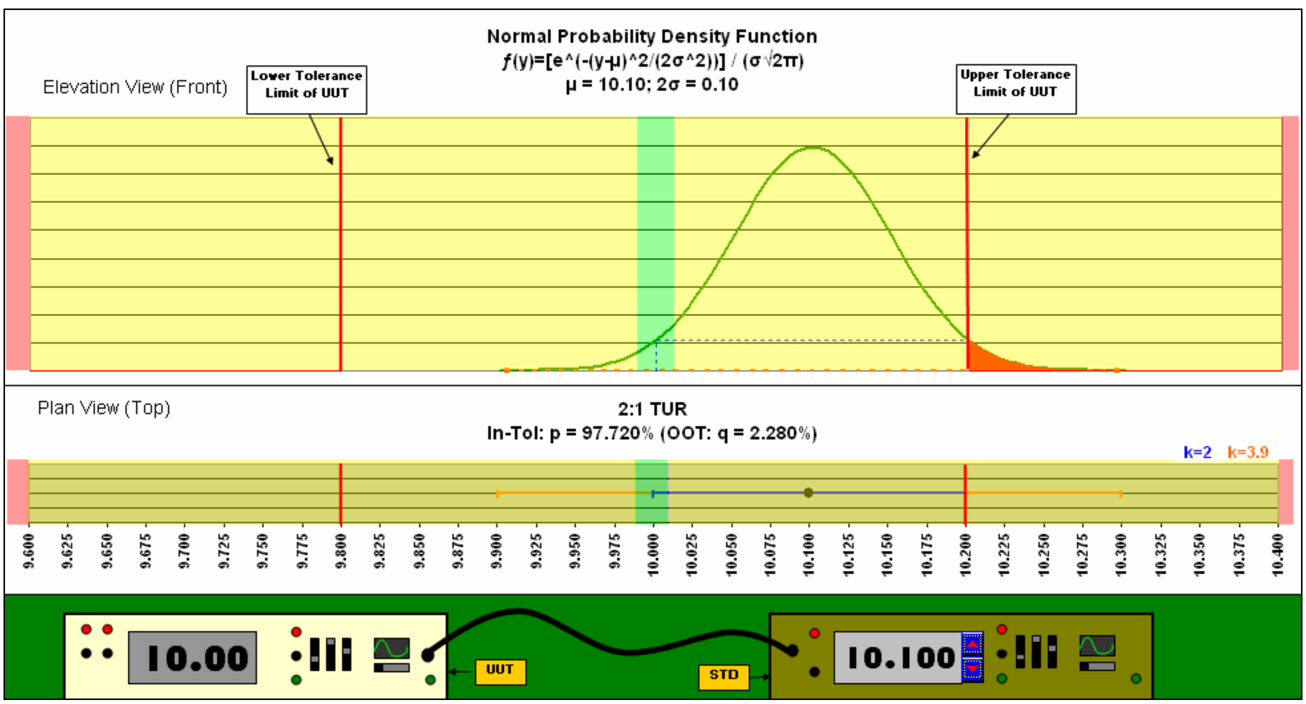


Figure 56: 2:1 TUR – Standard indicates UUT's value is 10.100 MHz, which is within the upper Indeterminate region (PCS = 97.7%).

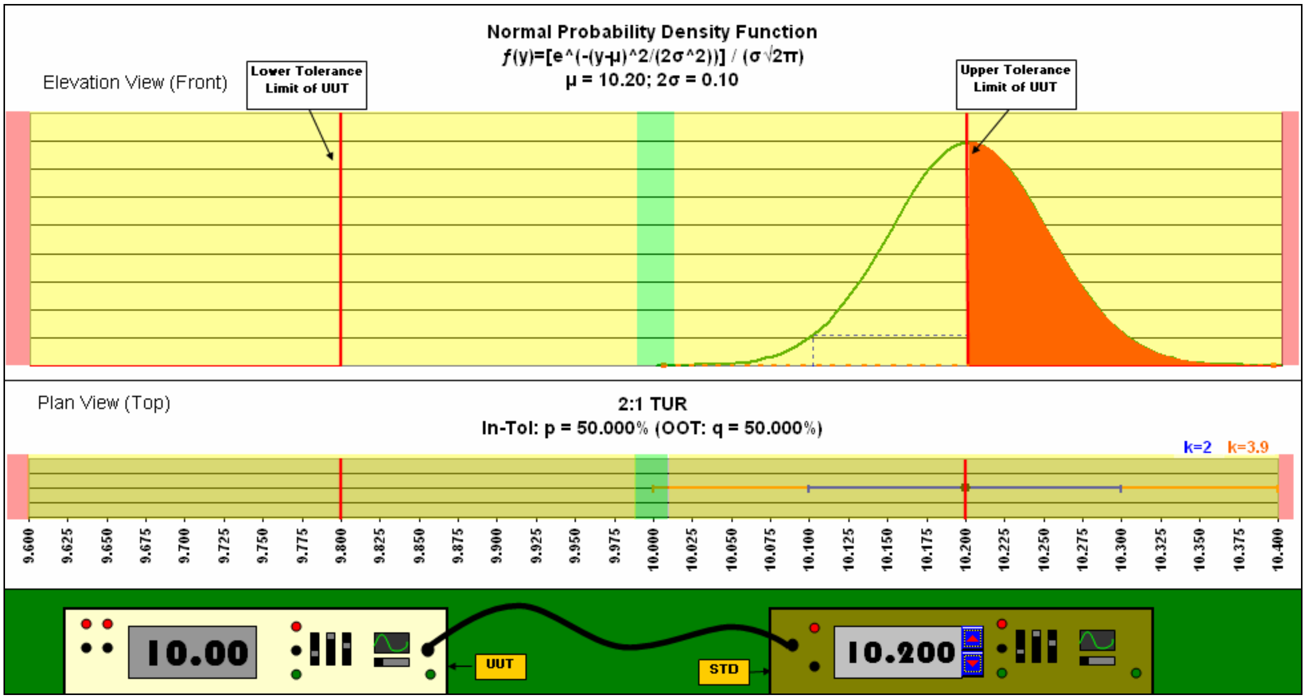


Figure 57: 2:1 TUR – Standard indicates UUT's value is 10.200 MHz, which is at the upper tolerance limit (PCS = 50.0%).

As you might expect, the 1:1 TUR scenario is even worse. Let's assume that, due to the limits of technology, we have only the ability to produce a standard whose measurement uncertainty is ±0.200MHz. The TUR is (0.2 ÷ 0.200 = 1), or 1:1, and that is a 1:1 when comparing k=2 to the UUT's tolerances. As you can see in figure 58, the Safeband has been annihilated! All that remains is one big, fat, Indetermine region!! How bad is it? Well, even when the UUT is at nominal, the uncertainty associated with the measurement causes the tail ends of the distribution to hang over **both** the lower **and** the upper tolerance limits! So, is it really a 1:1 ratio? Should we change the way TURs are calculated, so that the denominator uses a value at k=3.9 instead of k=2? This would make a 1:1 TUR more representative of its title. I'm not so sure we should change the ways TURs are calculated. If they remain as they are today, then they will speak to the relative 'goodness' of the measurement. From this paper though, they appear to have become a relative measure rather than an absolute measure. This will be a topic for discussion, should this concept go to committee.

Even at nominal, there are probabilities that the reading is actually OOT!! In fact, the PCS value is only 95% - at **nominal**!!! And notice how far OOT the UUT might actually be – probabilities exist for the UUT to actually be anywhere from 9.60 MHz to 10.40 MHz! Looking at this the other way, if the standard indicates that the UUT is at nominal value, then there is still a 5% chance that the UUT is actually OOT.

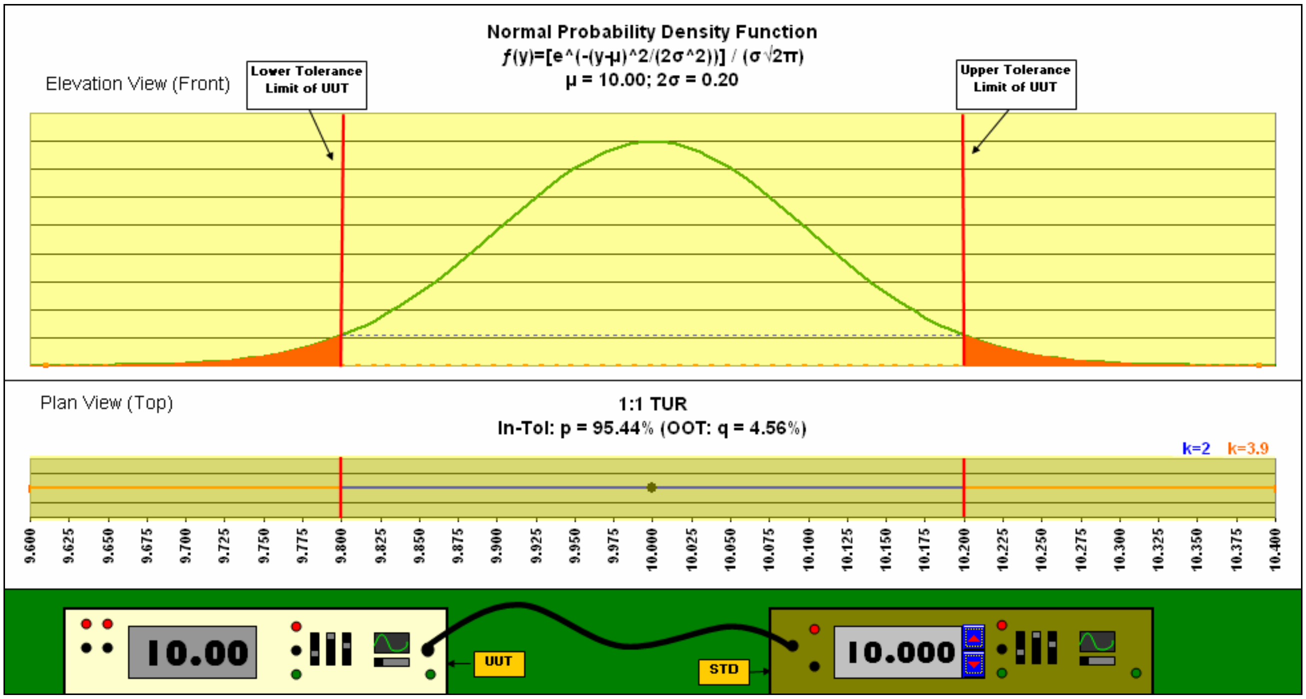


Figure 58: 1:1 TUR – Standard indicates UUT's value is 10.000 MHz, which is at nominal (PCS = 95.4%).

Regardless of how bad this looks, we can still make an accurate statement of the "Probability of Compliance to the Specification". Take a look at what happens to the PCS value as the nominal point moves throughout the tolerance limits (figures 59-69). Since the distribution at a 1:1 TUR has become so wide, I have rescaled the range of these graphs so that you can clearly see what happens at both ends.

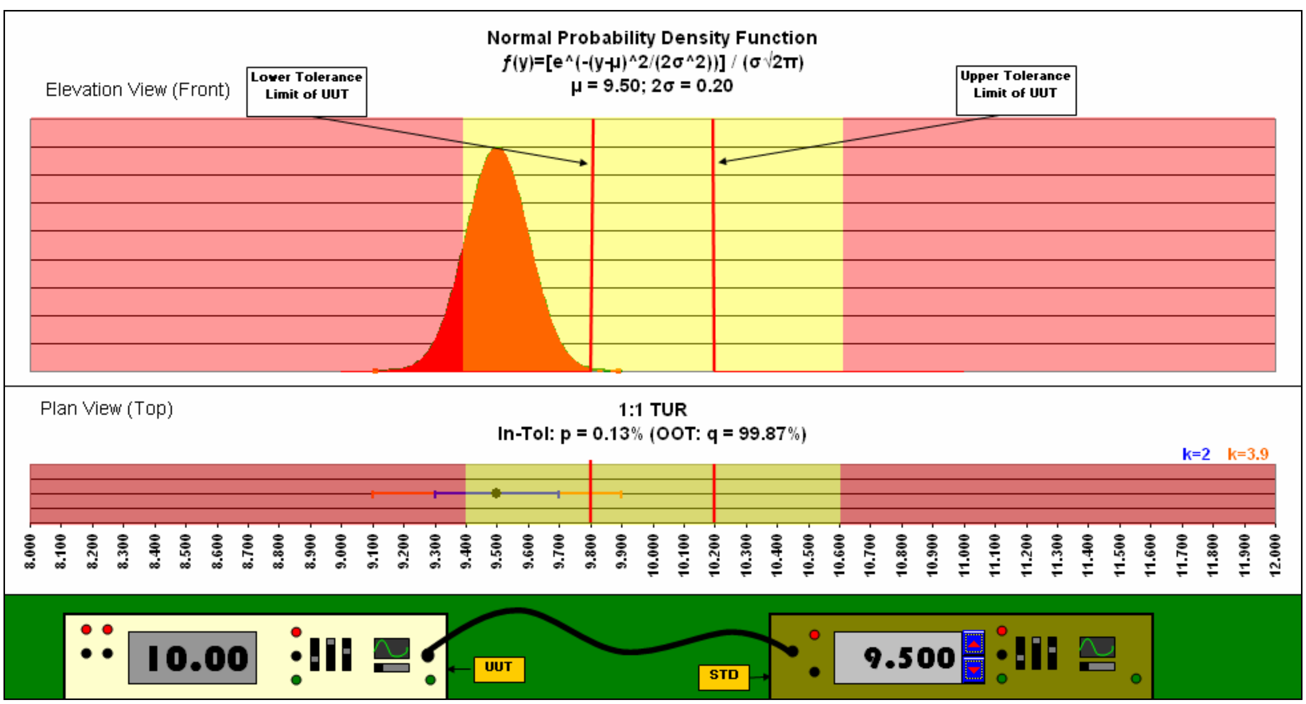


Figure 59: 1:1 TUR – Standard indicates UUT's value is 9.500 MHz, which is within the lower Indeterminate region (PCS = 0.13%).

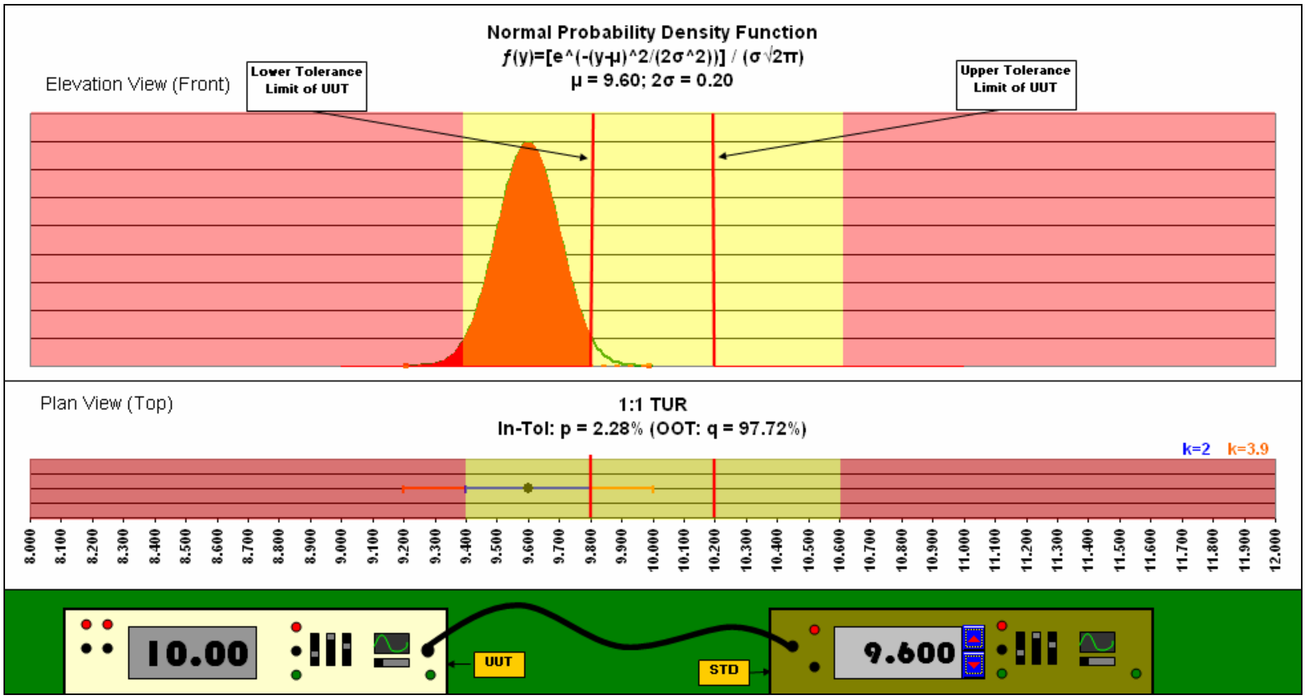


Figure 60: 1:1 TUR – Standard indicates UUT's value is 9.600 MHz, which is within the lower Indeterminate region (PCS = 2.3%).

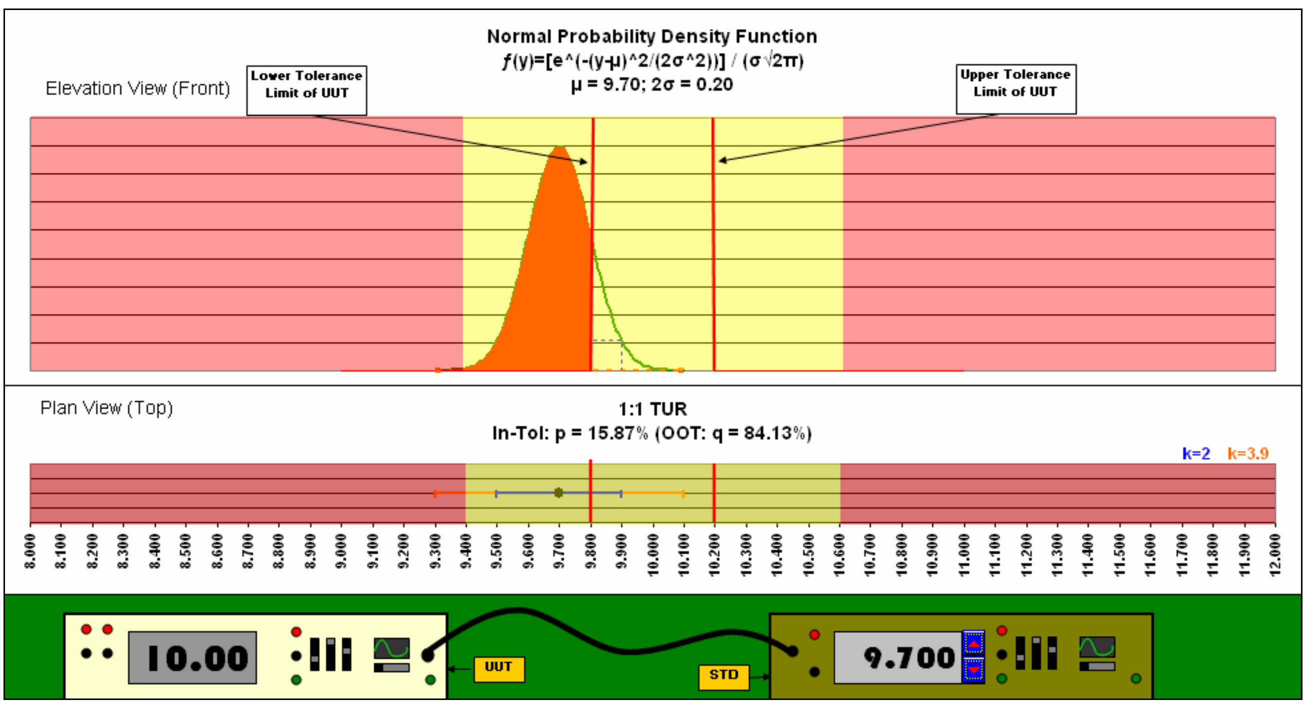


Figure 61: 1:1 TUR – Standard indicates UUT's value is 9.700 MHz, which is within the lower Indeterminate region (PCS = 15.9%).

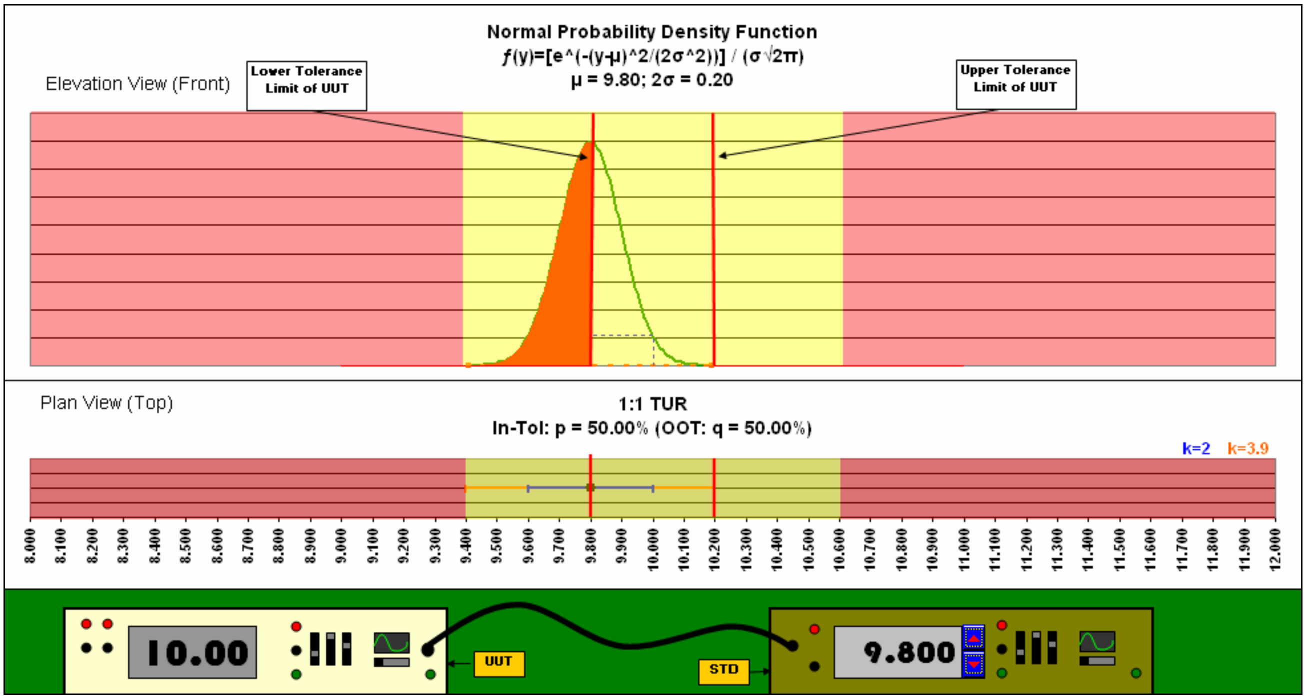


Figure 62: 1:1 TUR – Standard indicates UUT's value is 9.800 MHz, which is at the lower tolerance limit (PCS = 50.0%).

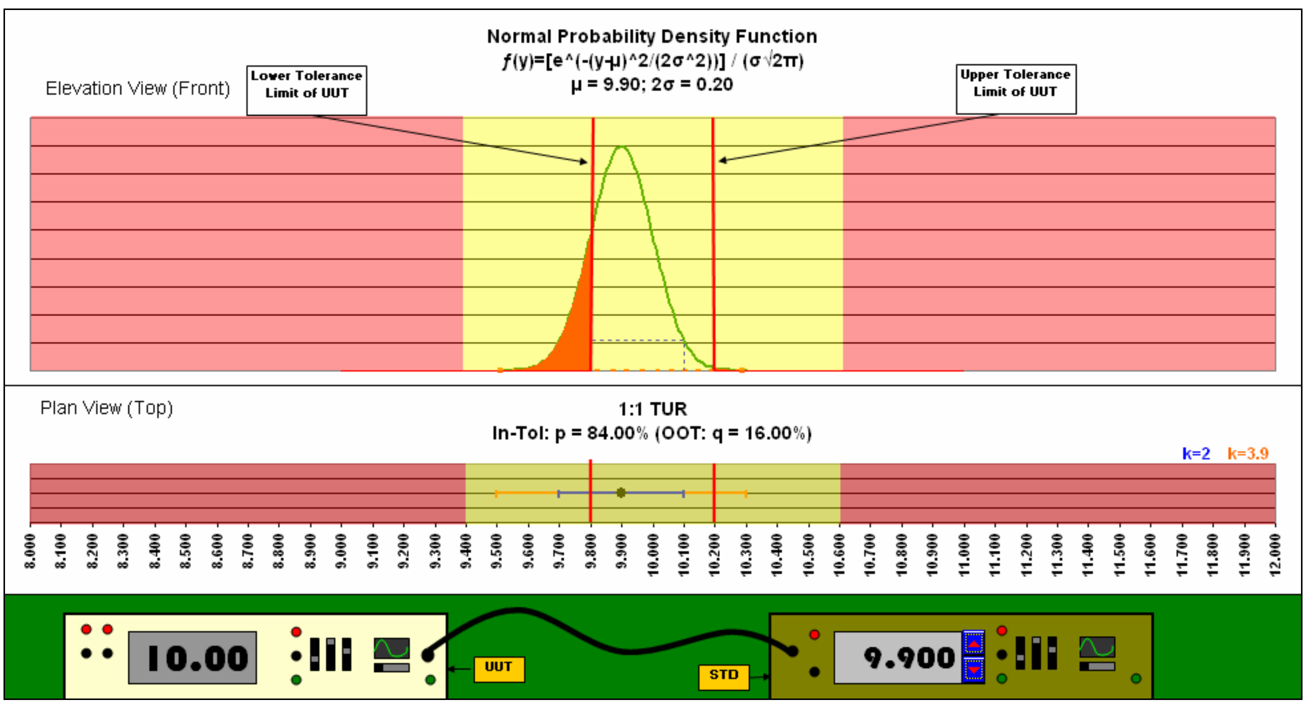


Figure 63: 1:1 TUR – Standard indicates UUT's value is 9.900 MHz, which is within the lower Indeterminate region (PCS = 84.0%).

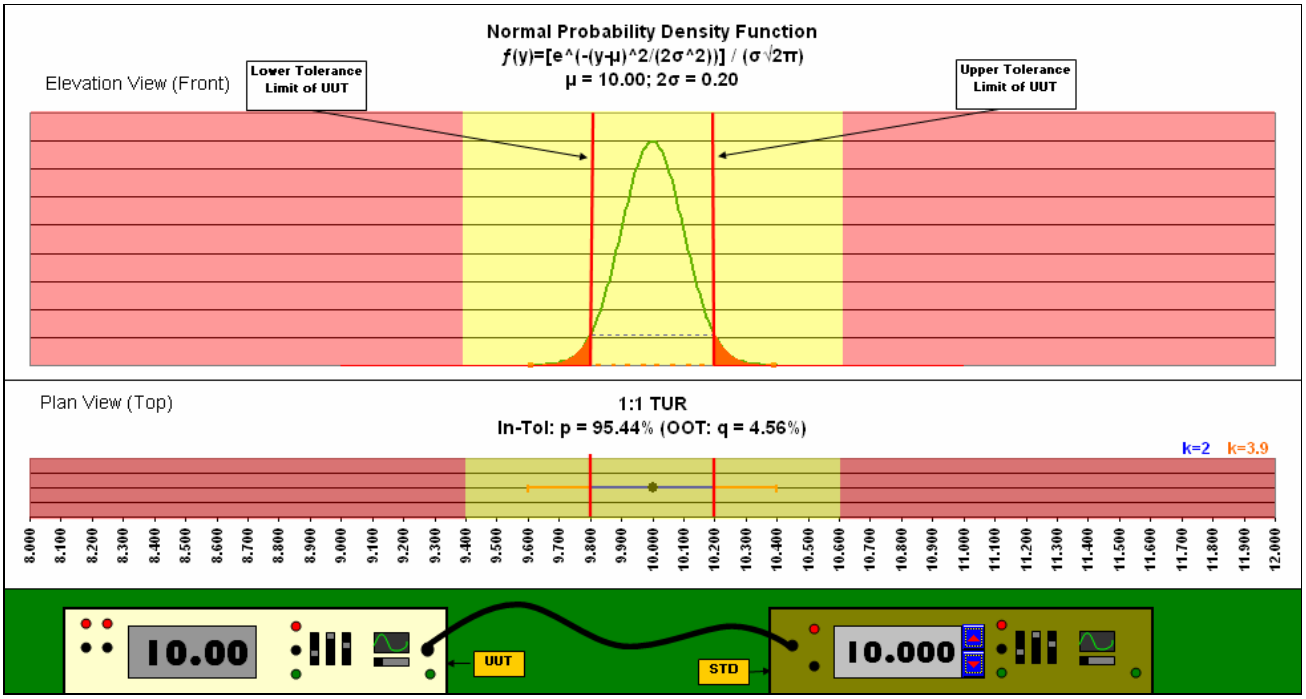


Figure 64: 1:1 TUR – Standard indicates UUT's value is 10.000 MHz, which is at nominal (PCS = 95.4%).

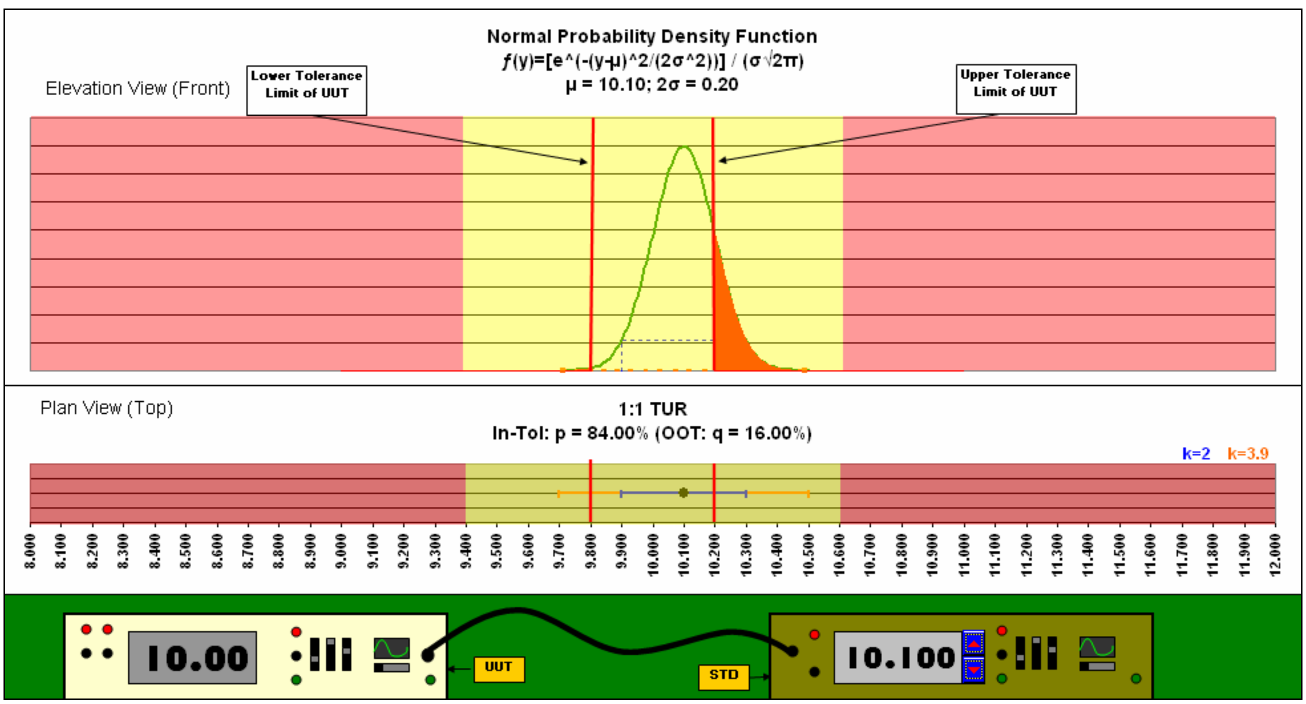


Figure 65: 1:1 TUR – Standard indicates UUT's value is 10.100 MHz, which is within the upper Indeterminate region (PCS = 84.0%).

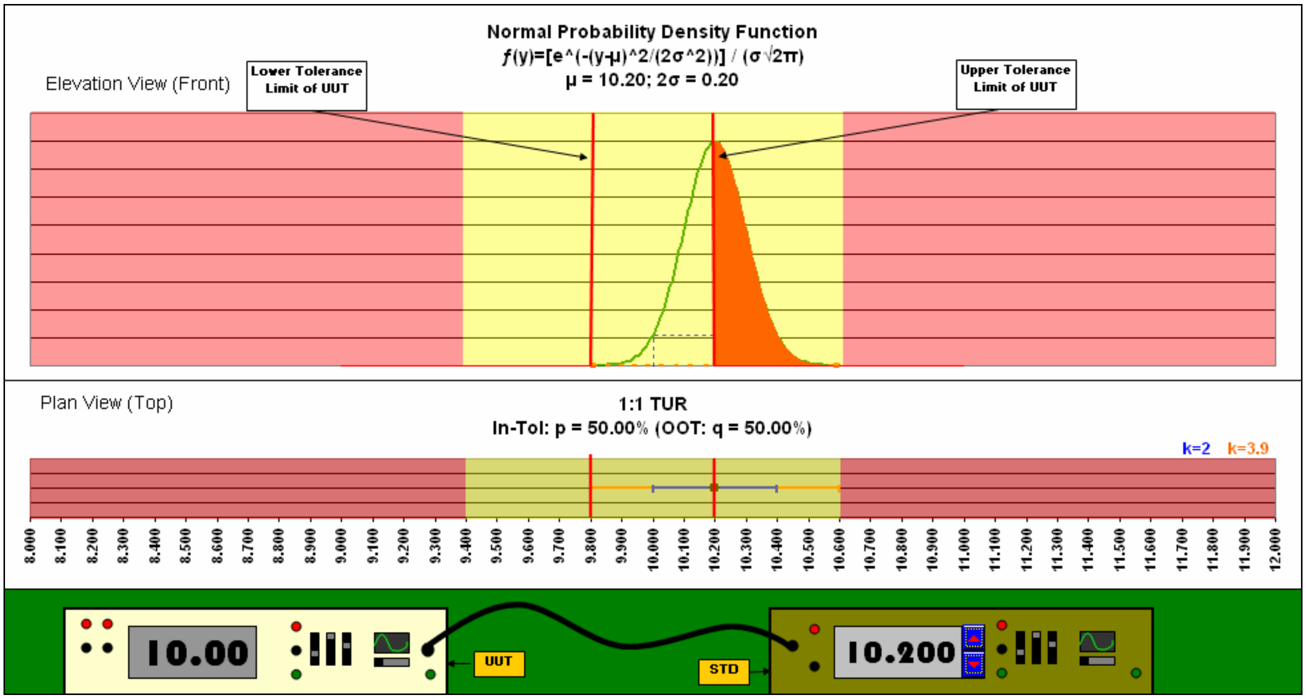


Figure 66: 1:1 TUR – Standard indicates UUT's value is 10.200 MHz, which is at the upper tolerance limit (PCS = 50.0%).

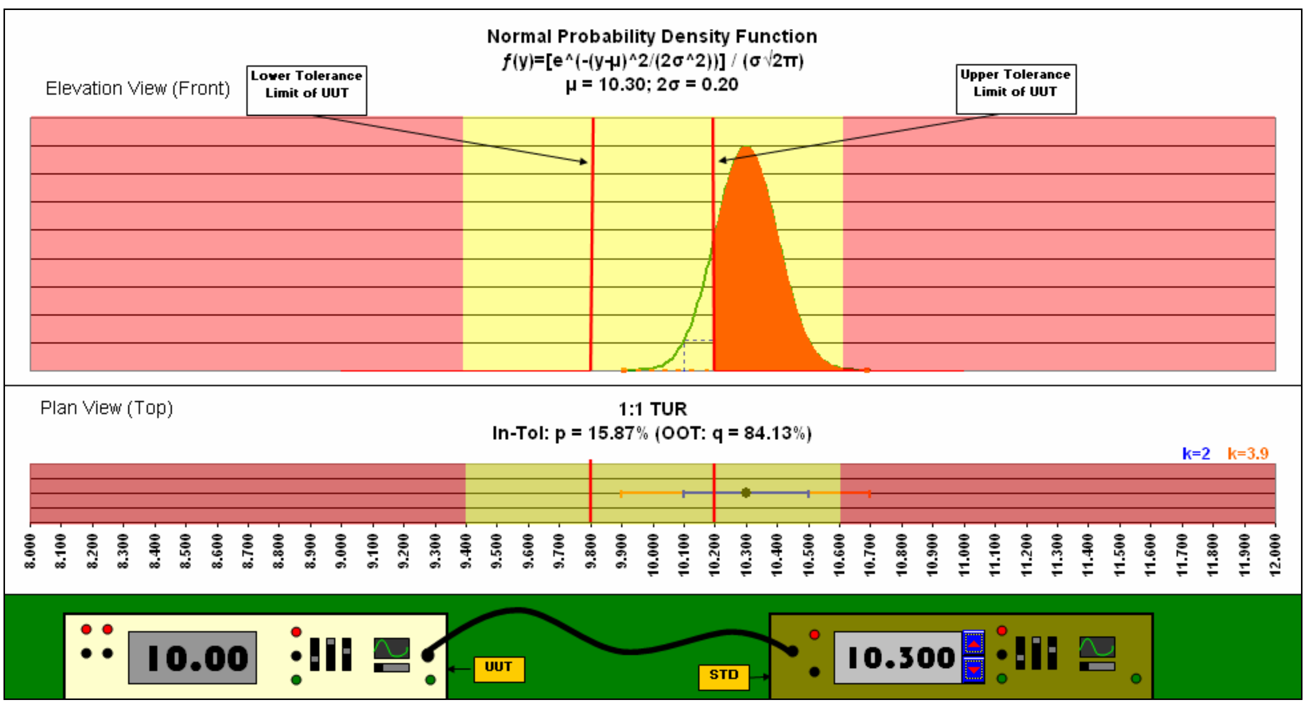


Figure 67: 1:1 TUR – Standard indicates UUT's value is 10.300 MHz, which is within the upper Indeterminate region (PCS = 15.9%).

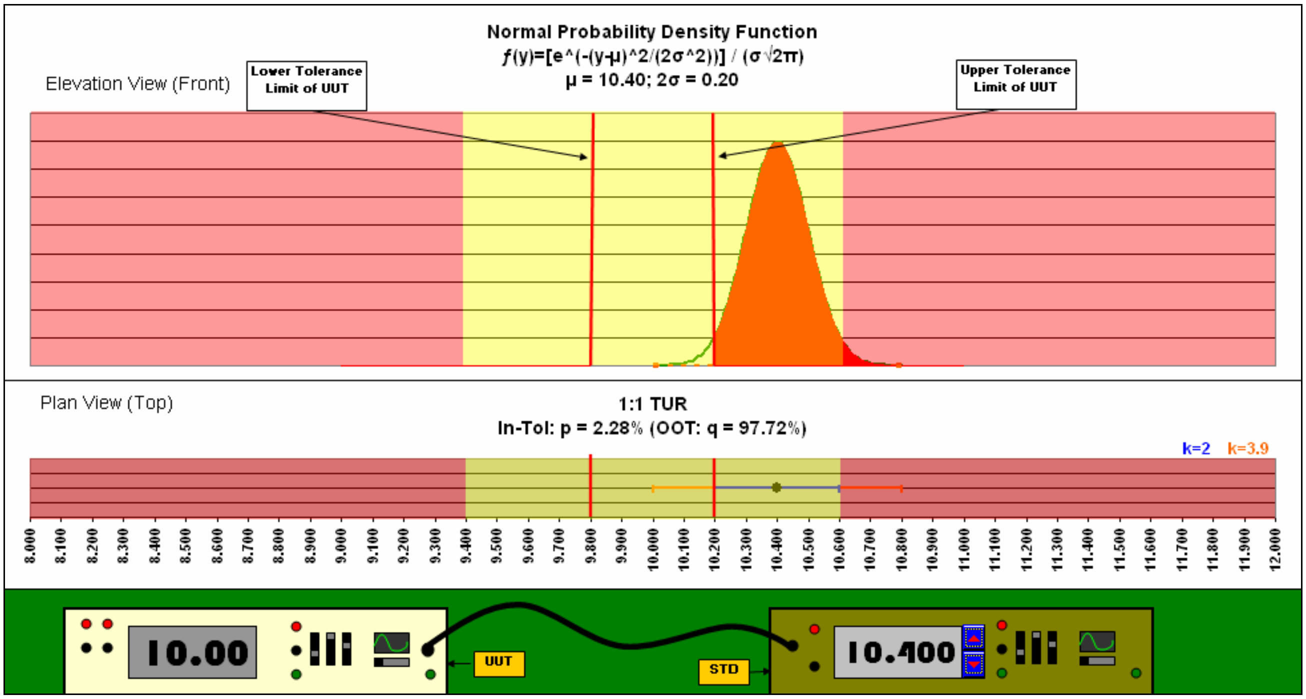


Figure 68: 1:1 TUR – Standard indicates UUT's value is 10.400 MHz, which is within the upper Indeterminate region (PCS = 2.3%).

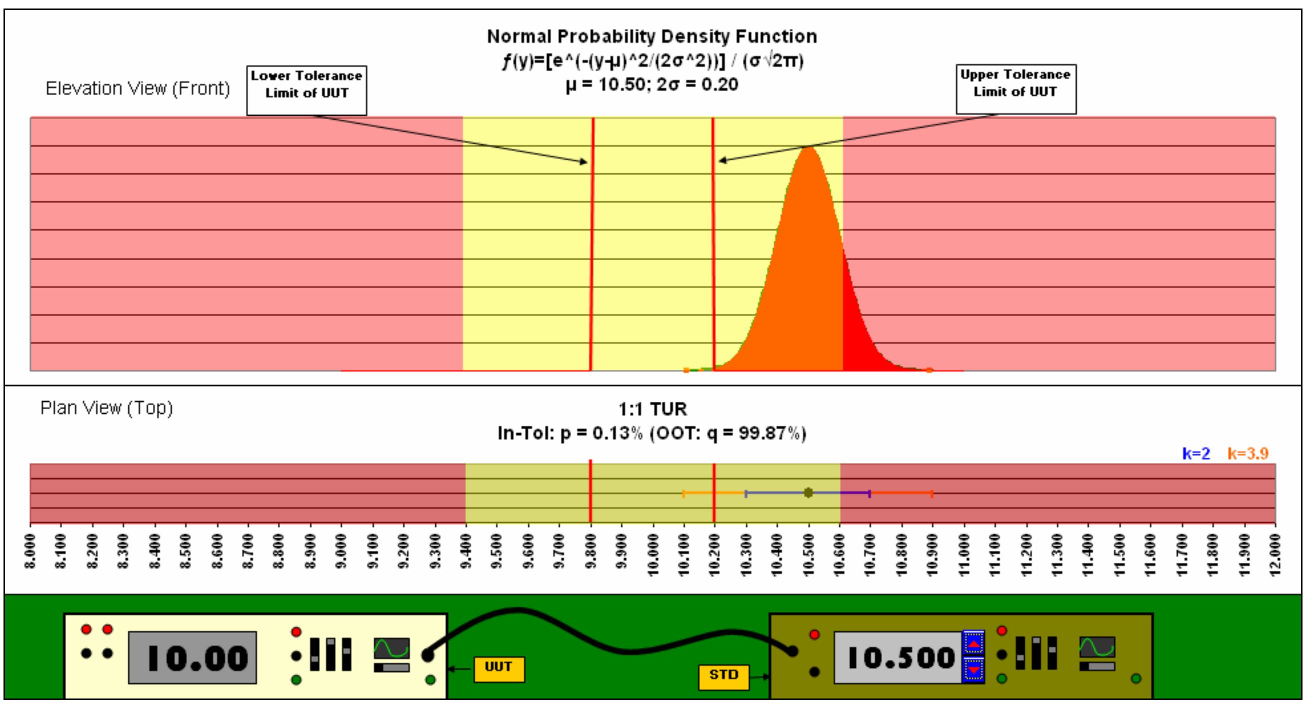


Figure 69: 1:1 TUR – Standard indicates UUT's value is 10.500 MHz, which is within the upper Indeterminate region (PCS = 0.13%).

So, this reveals what happens with the relationship of a 1:1 TUR, when the UUT's specifications are equivalent to the measurement uncertainty at k=2.

Now, let's take one last look at a series of readings for a 10:1 TUR, just for something fun to do! Say we have calculated the uncertainty budget for a brand new standard and the expanded uncertainty comes out to ±0.020MHz. The TUR is (0.2 ÷ 0.020 = 10), or 10:1. Figures 70 through 83 indicate many of the different calibration outcomes that we might come across.

Wow – look at the fields of green!! Obviously 10:1 TUR is not always achievable, either because the technology is not available, or because the cost is not justifiable. But this really increases the number of PCS values that come out to 100%.

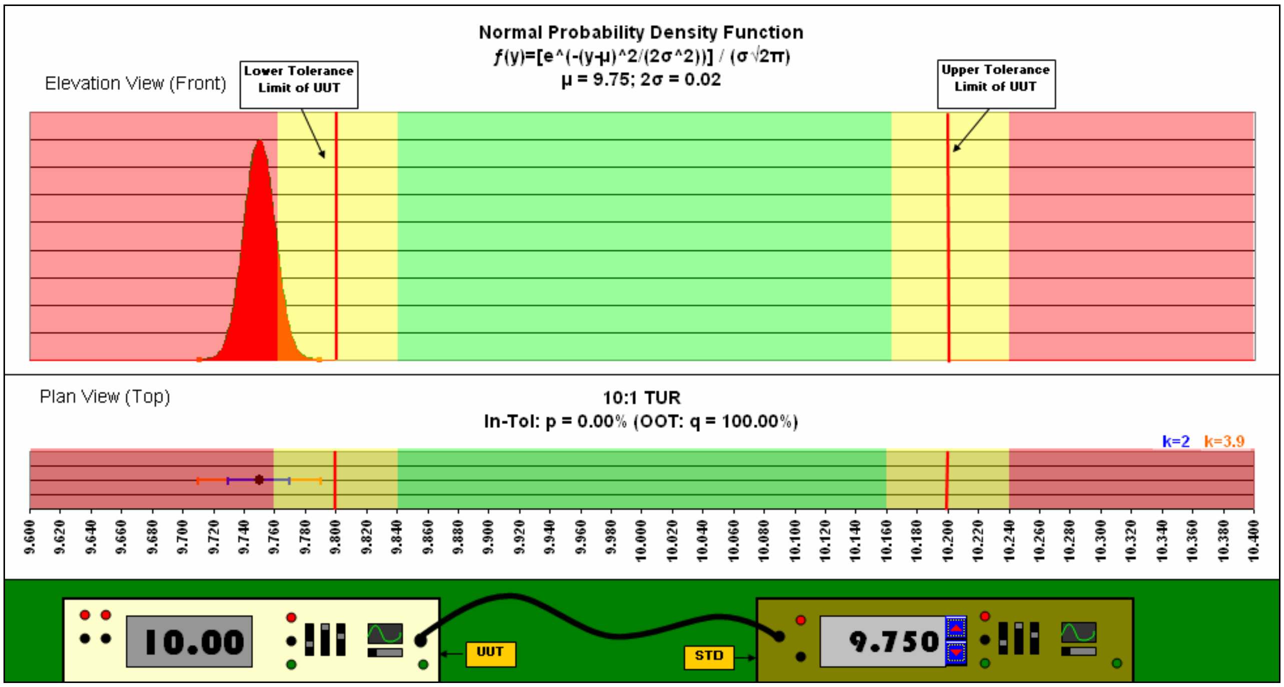


Figure 70: 10:1 TUR – Standard indicates UUT's value is 9.750 MHz, which is below the lower Indeterminate region (PCS = 0.0%).

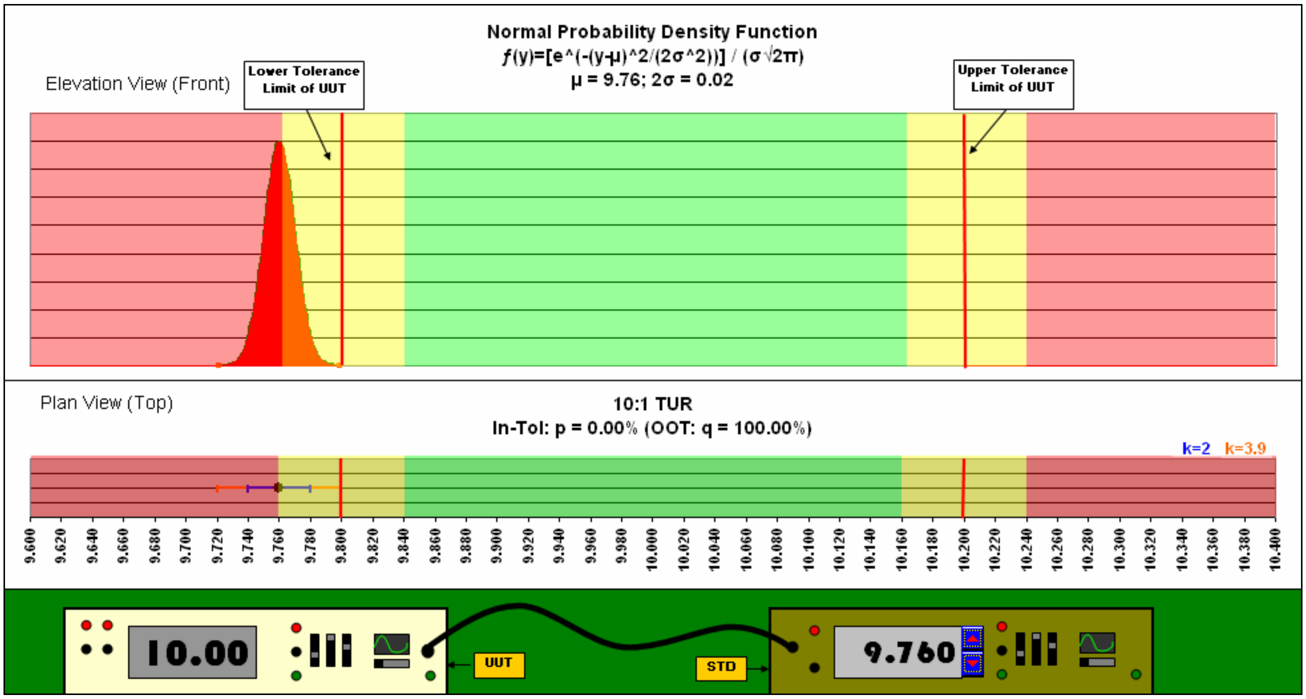


Figure 71: 10:1 TUR – Standard indicates UUT's value is 9.760 MHz, which is at the edge of the lower Indeterminate region (PCS = 0.0%).

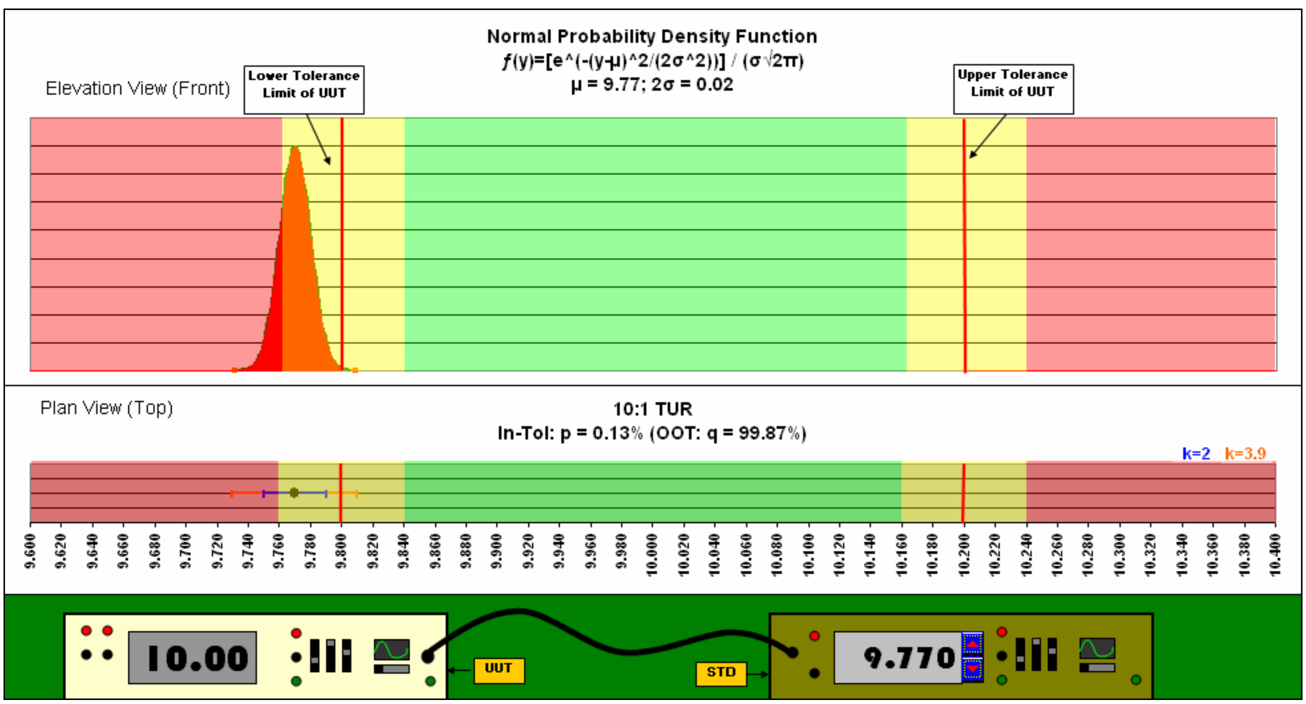


Figure 72: 10:1 TUR – Standard indicates UUT's value is 9.770 MHz, which is within the lower Indeterminate region (PCS = 0.1%).

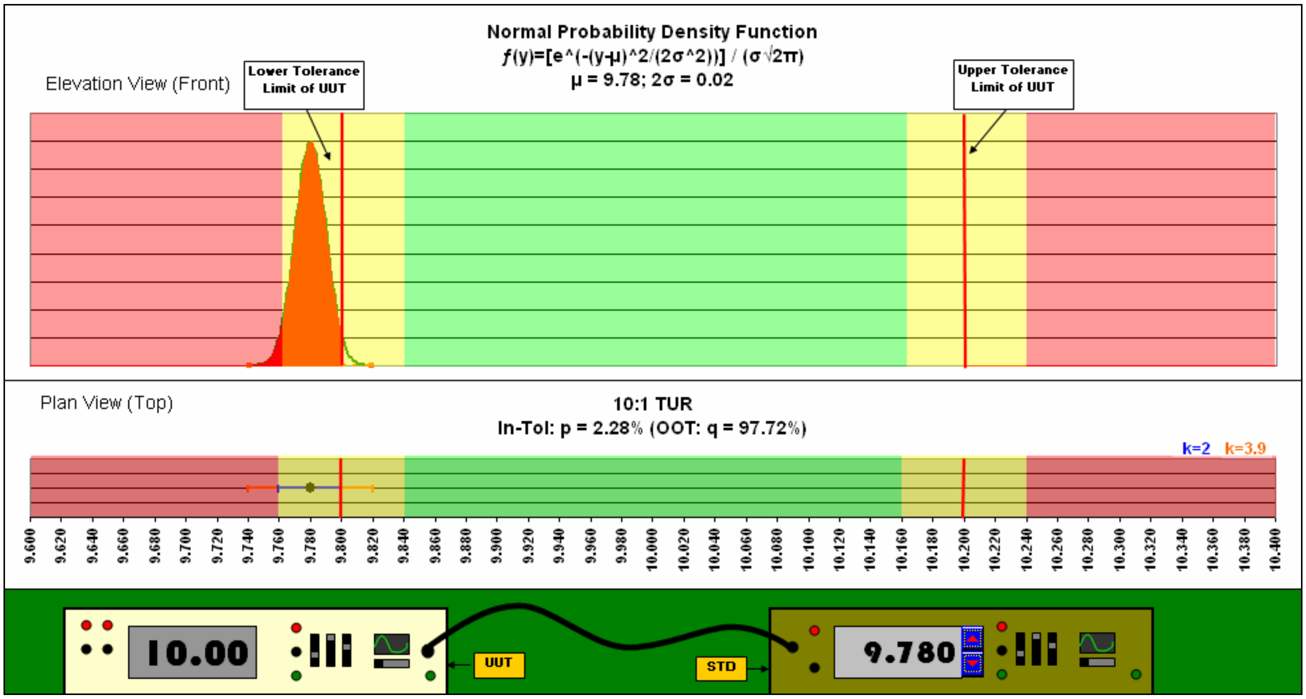


Figure 73: 10:1 TUR – Standard indicates UUT's value is 9.780 MHz, which is within the lower Indeterminate region (PCS = 2.3%).

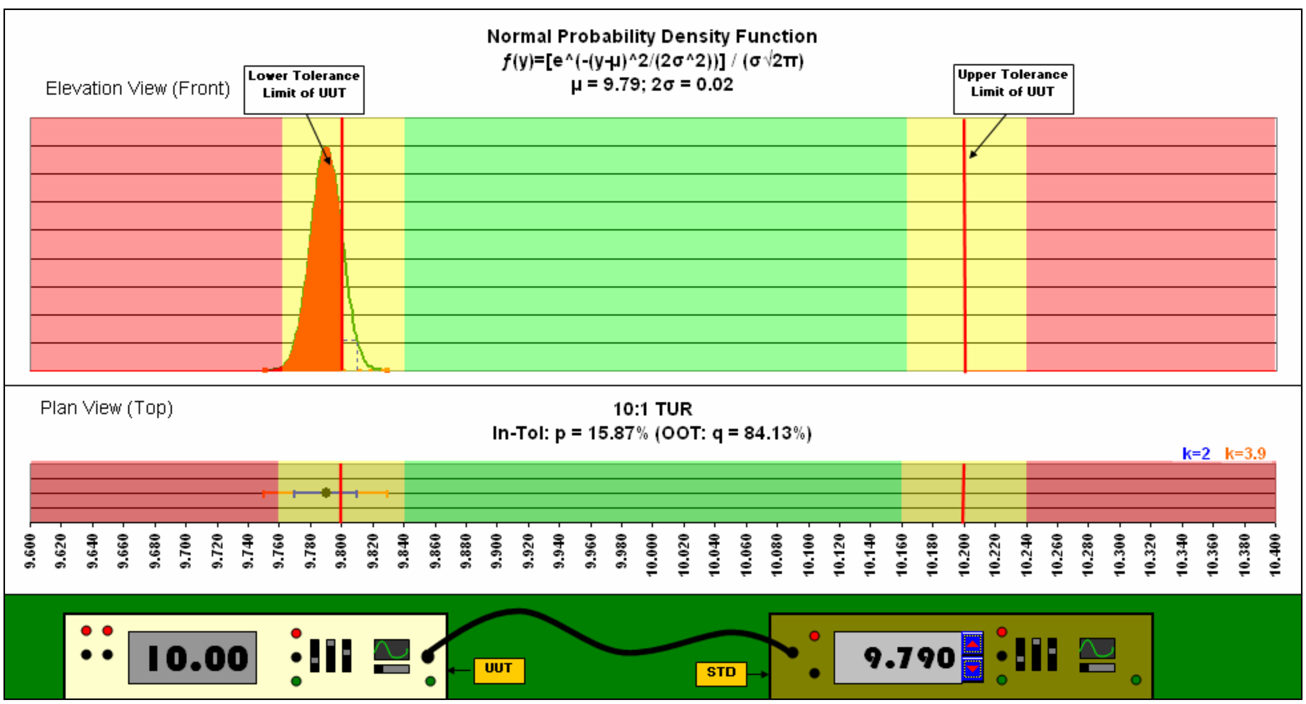


Figure 74: 10:1 TUR – Standard indicates UUT's value is 9.790 MHz, which is within the lower Indeterminate region (PCS = 15.6%).

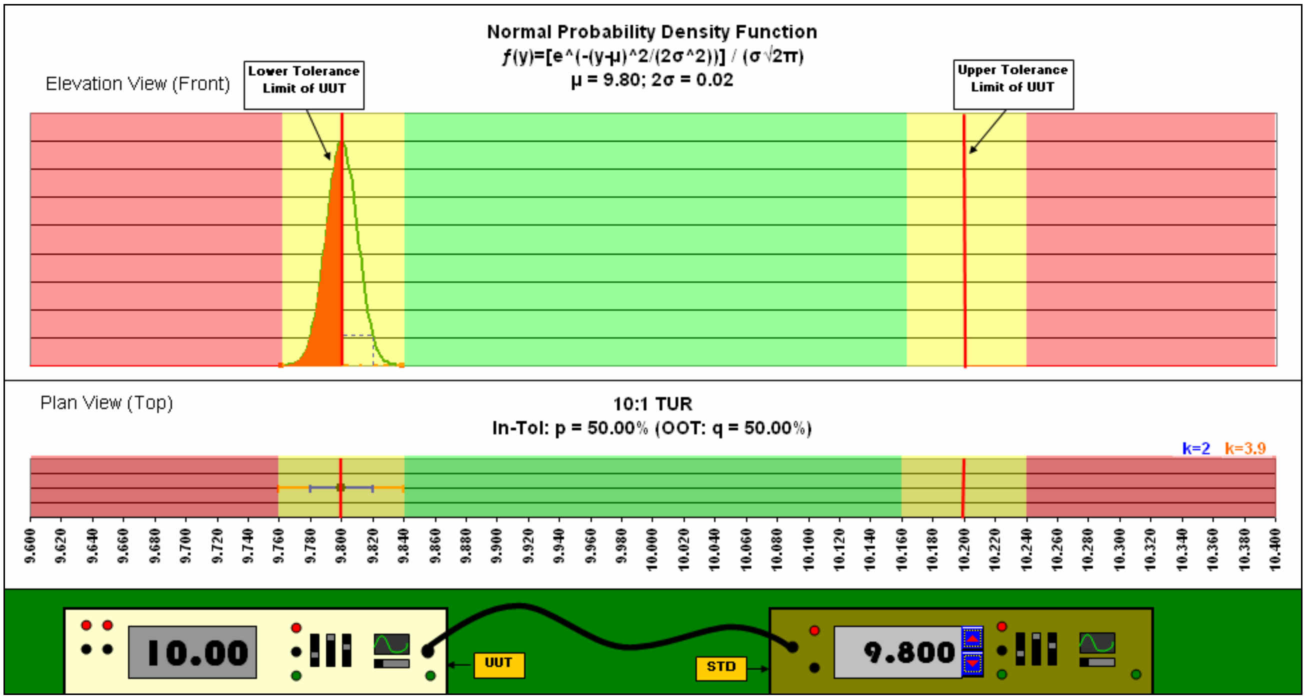


Figure 75: 10:1 TUR – Standard indicates UUT's value is 9.800 MHz, which is at the lower tolerance limit (PCS = 50.0%).

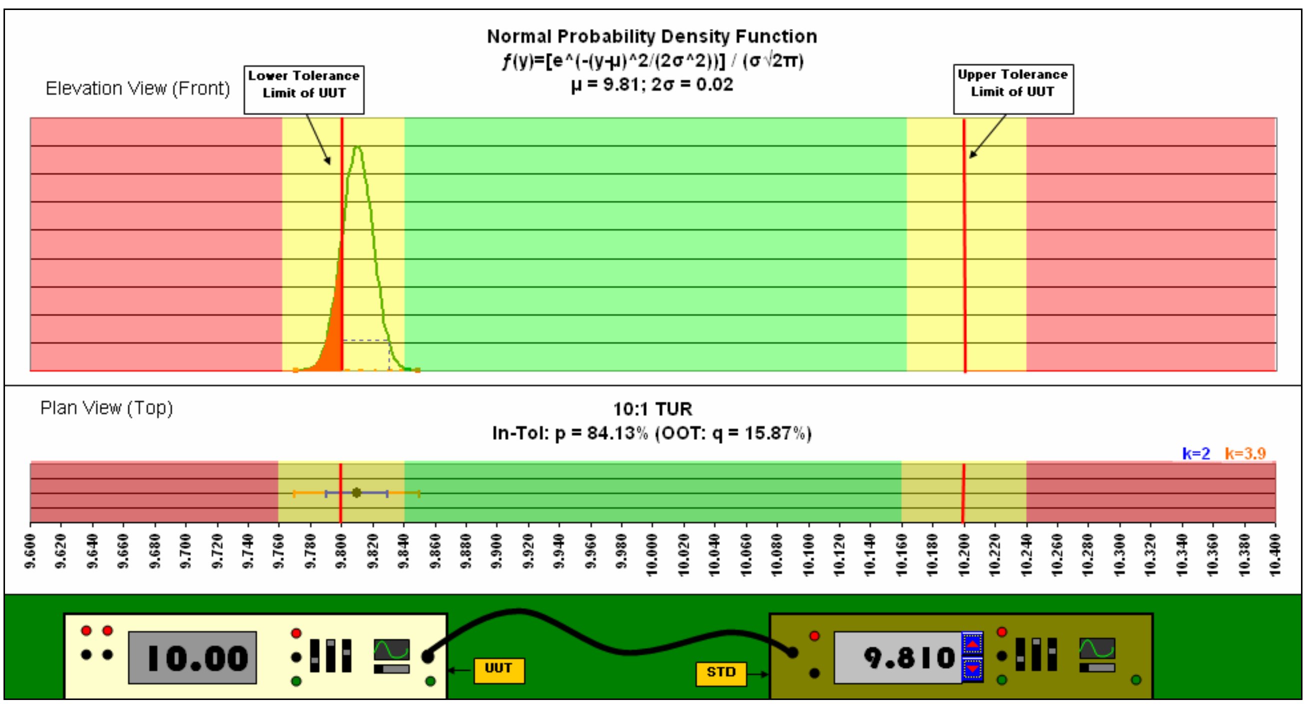


Figure 76: 10:1 TUR – Standard indicates UUT's value is 9.810 MHz, which is within the lower Indeterminate region (PCS = 84.1%).

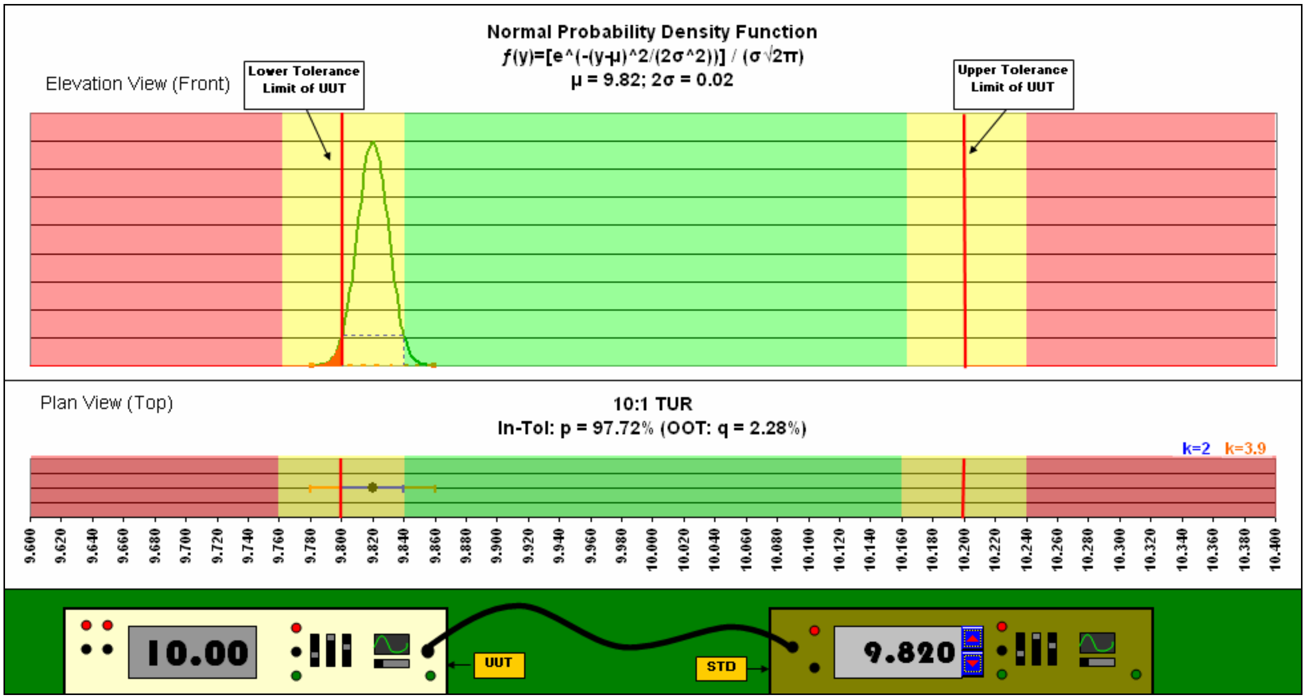


Figure 77: 10:1 TUR – Standard indicates UUT's value is 9.820 MHz, which is within the lower Indeterminate region (PCS = 97.7%).

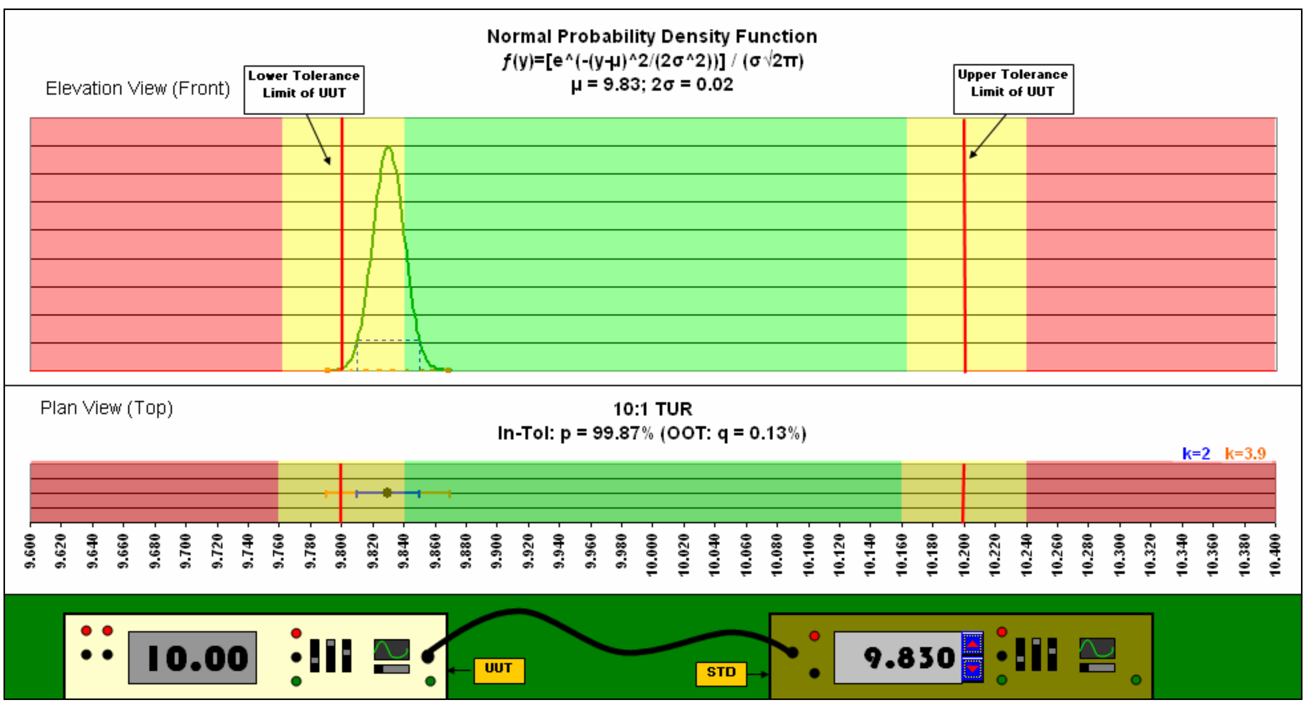


Figure 78: 10:1 TUR – Standard indicates UUT's value is 9.830 MHz, which is within the lower Indeterminate region (PCS = 99.9%).

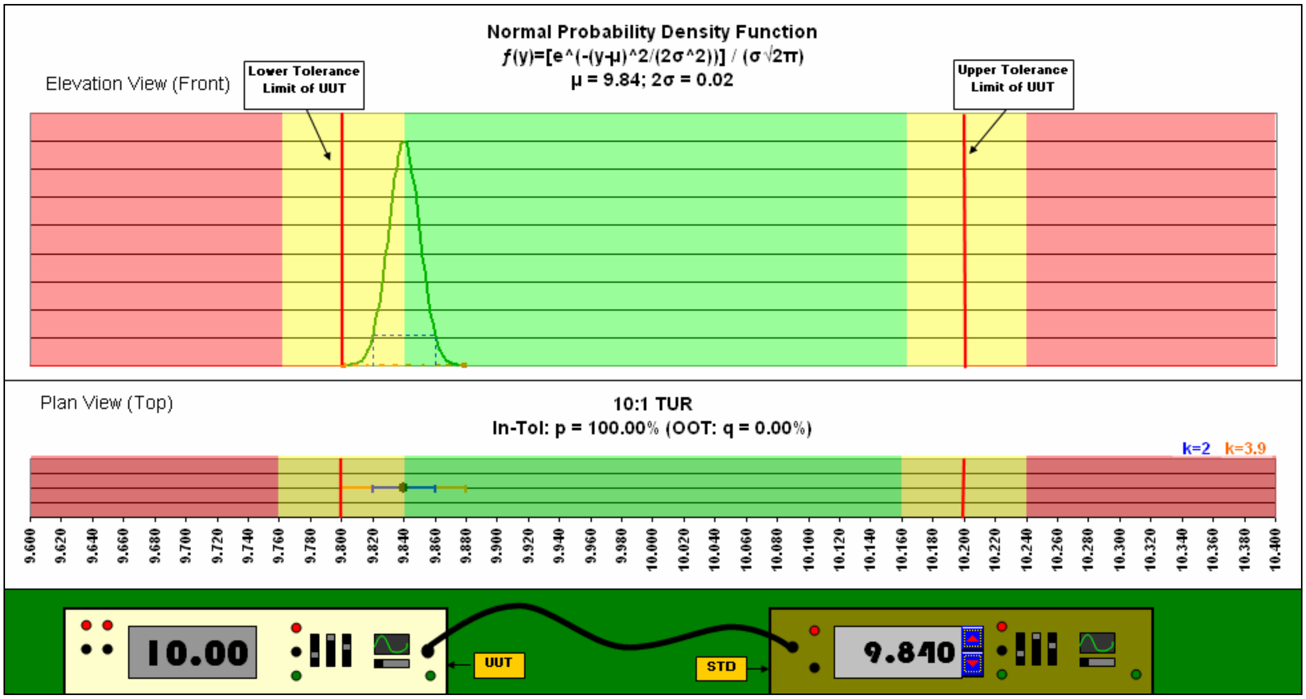


Figure 79: 10:1 TUR – Standard indicates UUT's value is 9.840 MHz, which is at the edge of the lower Indeterminate region (PCS = 100.0%).

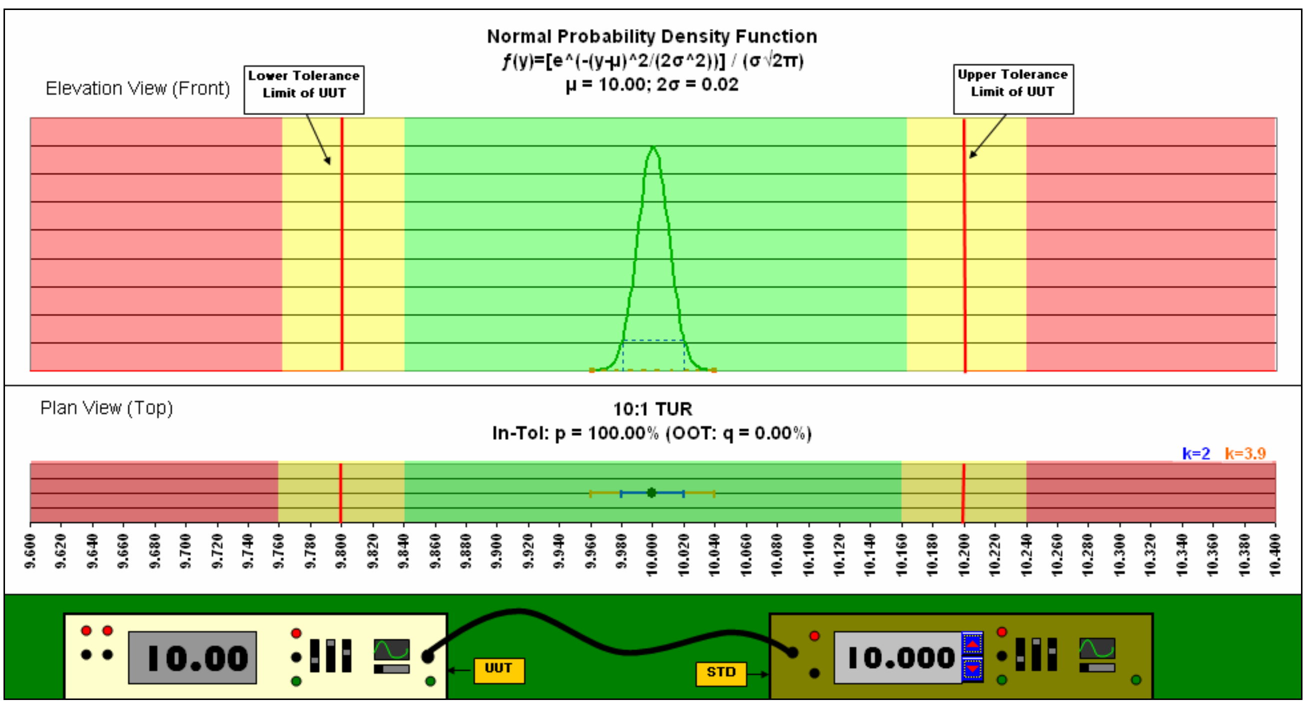


Figure 80: 10:1 TUR – Standard indicates UUT's value is 10.000 MHz, which is at nominal (PCS = 100.0%).

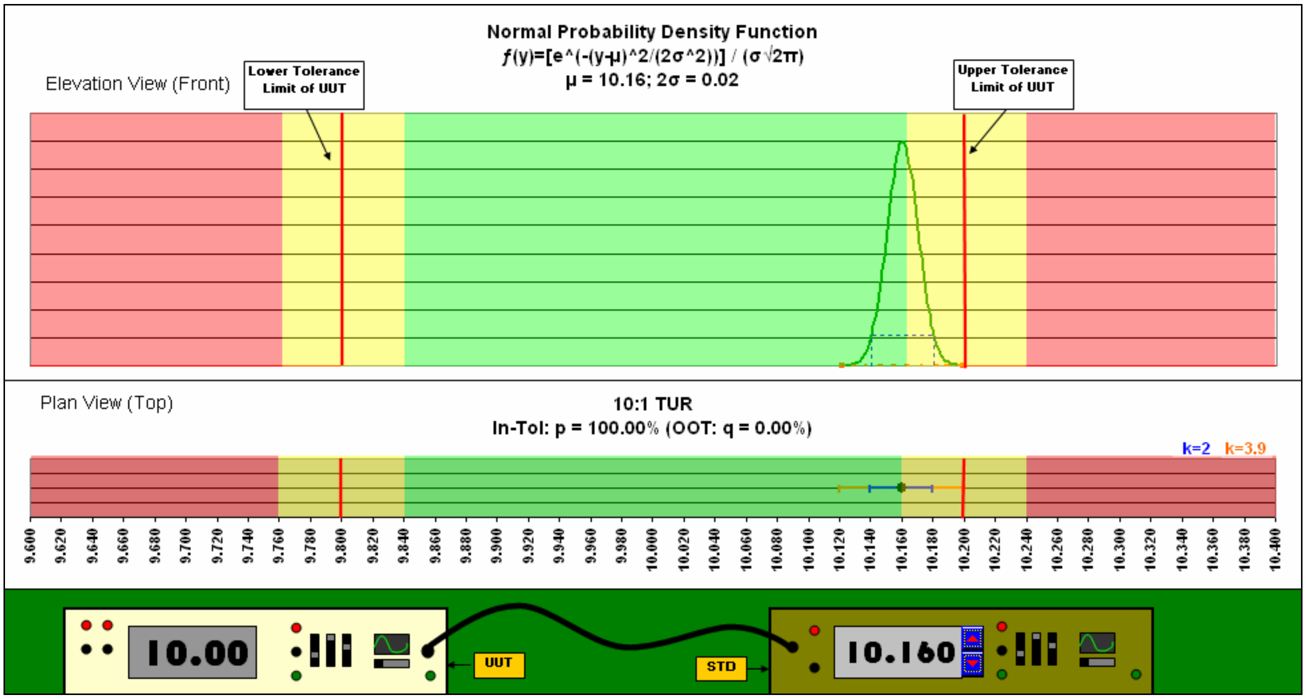


Figure 81: 10:1 TUR – Standard indicates UUT's value is 10.160 MHz, which is at the edge of the upper Indeterminate region (PCS = 100.0%).

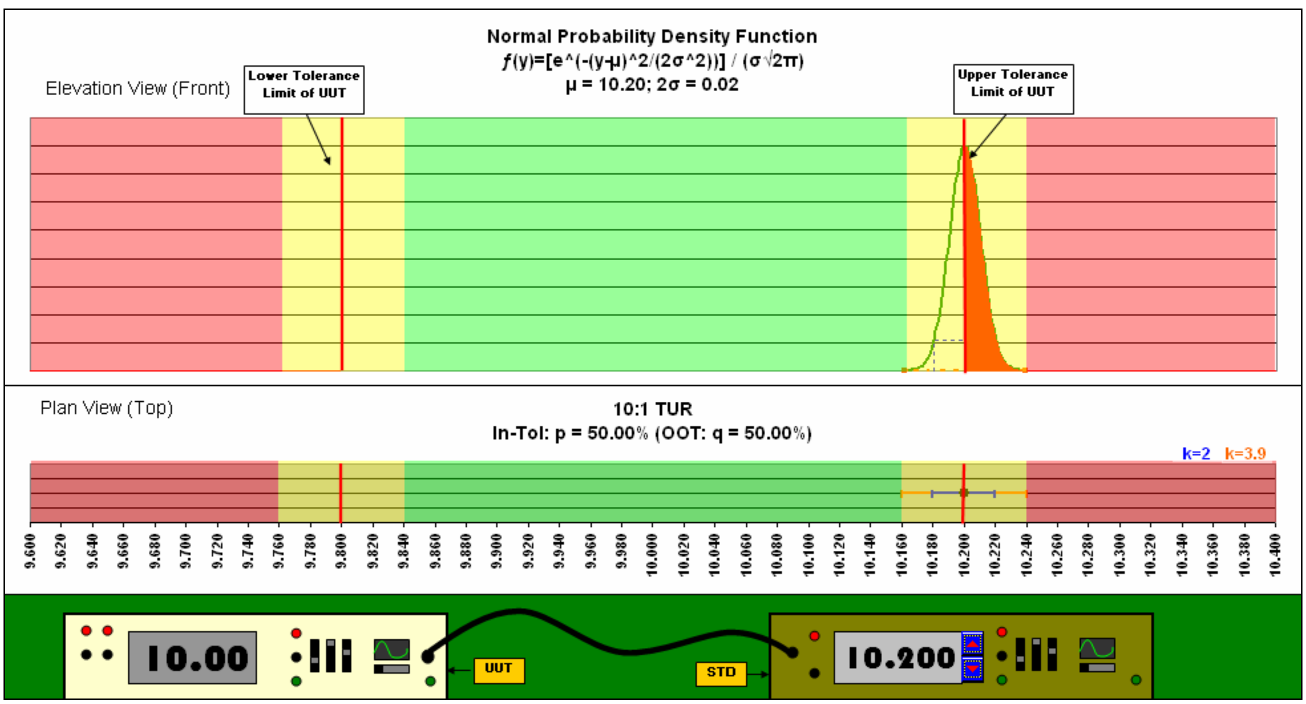


Figure 82: 10:1 TUR – Standard indicates UUT's value is 10.200 MHz, which is at the upper tolerance limit (PCS = 50.0%).

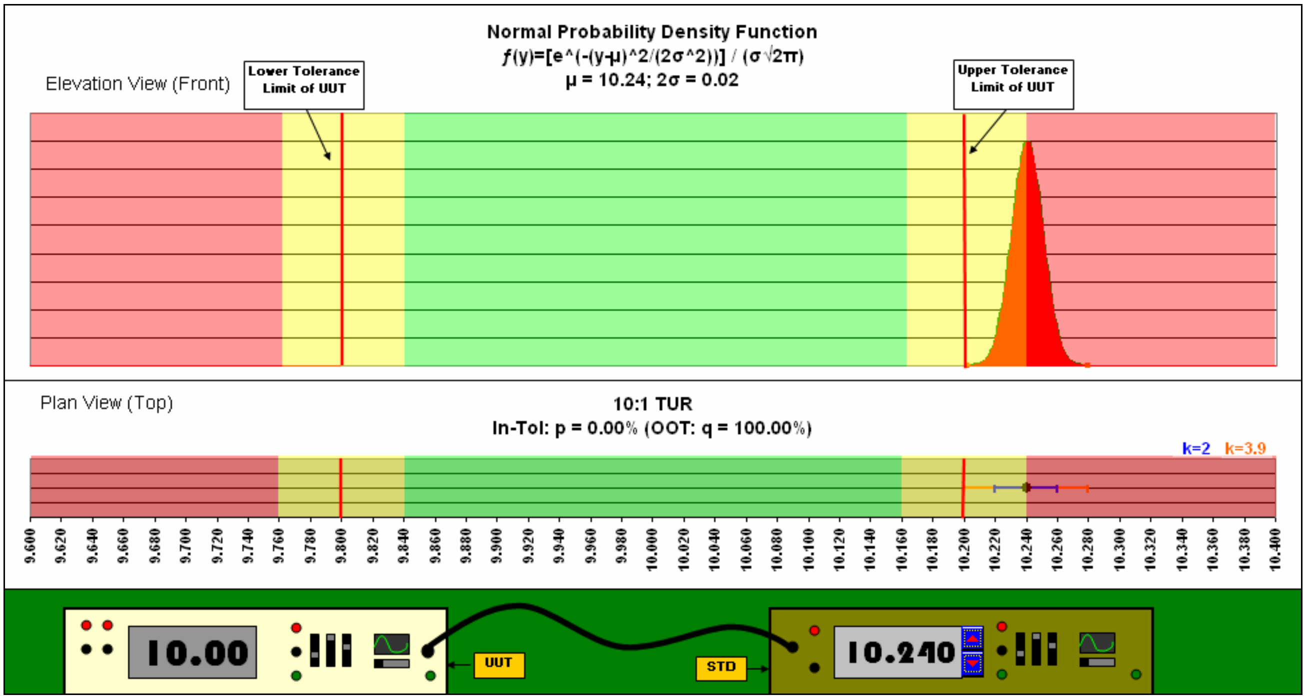


Figure 83: 10:1 TUR – Standard indicates UUT's value is 10.240 MHz, which is at the edge of the upper Indeterminate region (PCS = 0.0%).

Summary

Finally, the relationship between Metrology and commerce can be connected through one simple (yet complex behind the scenes) number: the Probability of Compliance to the Specification (PCS) statement is either 100% or it's something less than 100%. All relationships between the measuring process and the UUT, regardless of the TUR or the UUT's specifications, can be reduced to this easy-to-use indicator that all customers can come to understand. No more confusion about what the uncertainty means to the customer or how to translate TUR to some effective use in their process, unless they have a specific need to apply uncertainty, in which case this is available on the report. The PCS indicator takes all of these factors into account and translates it to a useful number the customer can rely upon. Either the customer doesn't need to worry about it (PCS=100%) or they must investigate the effect on their process (PCS<100%)!

I'd like to test this proposal against cases that I have not yet considered. I welcome feedback from the industry to test this so that it can either be modified to encompass all situations, or validated that it works in all situations. We know that there is uncertainty in measurement at all levels in the traceability chain, and we have solidified a means of estimating this uncertainty through ISO-17025 and the GUM (and also NIST TN 1297). And, as long as there are expectations of how far the UUT can drift over a given interval, then there are tolerance limits and calibration cycles. Given these components in the measurement process, a statement of Probability of Compliance to the Specification (PCS) should be applicable to all levels in the traceability chain.

About the Author

Howard Zion is the Director of Technical Operations for Transcat, Inc. and has held positions within the company as Laboratory Manager and Regional Operations Manager prior to accepting his current position in 2003. He holds a B.S. in Engineering Technology and a M.S. in Industrial Engineering from the University of Central Florida.

Howard has collected a wealth of knowledge in many disciplines during the span of 23 years in Metrology. Starting with his fundamental and advanced PMEL training in the United States Air Force, his Metrology experience includes:

- Metrology support for production manufacturing with Martin Marietta (Lockheed Martin) for the Patriot and Copperhead Missile programs and components of the Apache Helicopter (Hellfire Missiles and the TADS/PNVS FLIR system);
- A decade with NASA-contractors working in the Standards and Calibration Laboratories at the Kennedy Space Center in support of the US Space Program and the Merritt Island National Wildlife Refuge located on the center;
- Design and development of a fully automated RF/Microwave and Fiber Optics calibration laboratory with Philips Electronics

Howard has authored white papers that are being presented at the NCSLI and MSC, as well as seminars throughout the United States in 2006:

- "Metrology Concepts: Understanding Test Uncertainty Ratio (TUR)"
- "k=3.9? . . . Why Not???"

With these credentials, it is no surprise that Mr. Zion is sought for his invaluable experience. He currently serves on two NCSLI committees:

- Working Group on the Electronic Format of Instrument Specifications
- Working Group on Standardization of Position Descriptions in Metrology

Additionally, Mr. Zion is on the advisory board at the University of North Carolina at Charlotte (UNCC) for the Engineering Technology programs. Through Transcat, he sponsors and mentors Engineering Capstone projects for undergraduate students at the Rochester Institute of Technology (RIT). Additionally, he has been named a Principal Investigator for the Measurement and Calibration Onsite – Inline Project, through Sinclair Community College in Dayton, OH, who is in pursuit of a National Science Foundation – ATE grant.

You may contact Howard at hzion@transcat.com

Transcat sells and markets test and measurement instrumentation as well as calibration and repair services to a variety of industries including life sciences, pharmaceutical, petroleum refining, chemical manufacturing, and public utility. Transcat celebrates its 42nd year in business in 2006. For more information, go to www.transcat.com