A Systematic Approach to Common and Expected Uncertainty

by Eli Schragenheim

**ABSTRACT**

*This paper suggests a systematic approach for dealing with routine decisions that contain high levels of uncertainty, which is hard to measure with known statistical means because of lack of formal data. The focus of the paper is to assist in management decision making through the estimation of the confidence interval, rather than by the estimation of the expected value and the standard deviation. It Is claimed that most management decisions should be based on the extreme values of the confidence interval rather than on the expected value. It is further argued that when the uncertainty estimation is based on intuition, it is best to directly estimate a reasonable confidence interval.*

*The systematic approach to intuitive uncertainty estimation allows for an appropriate feedback on the estimation, which might lead to an improvement of future intuitive estimations.*

**INTRODUCTION**

No decision is made with absolute certainty. This means that the actual results from decisions can vary. Does management have any guiding rules for making uncertain decisions?

Does statistics provide such a practical methodology? Even though many managers get a fairly good education on both probability and statistics, they do not use it very often. The thinking paradigms created by probability theory and statistics do not seem very useful in management reality. Too often there is not enough data to use any statistical tool. Collecting huge amounts of data to be used for statistical models is deemed a nuisance and is therefore not done. Most information systems are not designed to collect such vast amounts of data, just for the purpose of estimating uncertainty.

How should uncertain decisions be made when all that is available is intuition? While the estimation itself has to be intuitive, the decision making can still be structured. The intuitive estimation itself can be more structured and collected in order to allow for feedback and future improvement.

**UNCERTAINTY CLASSIFICATION**

Uncertainty can be classified into three categories:

1. Unexpected events/results. As it is unexpected from the manager's point of view, it is impossible to estimate the likelihood of its appearance, and no measures can be taken to control it.

2. Infrequent events, usually of catastrophic nature, greatly impact the organization and put it in danger. This is where insurance comes in. The vast majority of risk management practices are geared towards this type of uncertainty.

3. Common and expected uncertainty that impacts the daily life of every manager. All the frequent and regular decisions that are uncertain because of the inherent randomness of certain variables. This uncertainty is handled by forecasting and by mere intuition, only seldom rooted in any procedure or quantitative algorithm. The focus of this paper concerns this type.

There are two ways to deal with uncertainty. The first is to reduce the amount of uncertainty by impacting the source of the uncertainty. This is advocated in TQM and can be effective when the source of the uncertainty lies within the organization.

The other way is to maintain a protection mechanism in order to reduce the damage that is caused by the uncertainty. In a way, this can be viewed also as reducing the uncertainty by considering the damage to the organization as the main random variable.

Protection mechanisms exist in various forms. However, when the amount of uncertainty is not measured and not even addressed, the actual protection is often either too high or too low. Some protection mechanisms protect, directly or indirectly, the sales and/or the performance of the internal constraint, thus reducing the throughput of the system and/or creating more expenses and higher inventories than needed.

**FORECASTING FOR A PEAK DEMAND PERIOD**

The following example is used in order to highlight the problem. Make-to-forecast is a notorious policy. It is widely used as a compromise to gap between the need to produce according to the market demand and the inability of the shop floor to respond in a very short time. Many times make-to-forecast is done in order to achieve high efficiencies. In other cases it is used to anticipate peaks of demand. This is the case with consumer goods that have a very significant peak of sales at Christmas. Usually the manufacturing companies do not have enough capacity to produce all the market demand within the peak period itself. Thus, it has to produce based on a forecast.

Suppose the forecast for product XYZ during Christmas is 1,544,670 units. What is the probability that during Christmas 1,544,670 units exactly will be actually sold? Practically zero! In reality either we have unsold units of product XYZ after Christmas or all the units were sold out, meaning some potential customers couldn't find it and probably bought the competitive product.

The decision the production manager faces, given the forecast that 1,544,670 units of XYZ will be sold, is how many units should be produced? Is it reasonable to produce the expected quantity to be sold? No, it isn't.

Another consideration needs to be taken: What is the perceived damage caused by not being able to supply all the demand, and what is the perceived damage from having surplus finished goods in inventory? There is no reason to believe that the two types of damage are equal.

Theoretically it is possible to find an “optimal” solution, all you need is to know the distribution function of the sales and the two damage functions. Suppose that the damage from a lost sale is much greater than the damage from excessive finished goods inventory. It is reasonable to produce more than the expected value of the number of units to be sold. How many more? Reasonably so, say an amount that ensures with perhaps 90% confidence that no missed sales will occur. As the actual distribution function of the actual sales is unknown, specifying a 90% confidence is a mere intuitive guide.

My only claim is that it is better than an intuitive decision that is made without any guidelines whatsoever. It is evident that when the damage of a lost sale is greater than the damage of excess finished goods, there is no need to assess how many units will be sold "on average" but the amount that roughly represents 90% confidence that there will be no shortage of that product.

When the damage from surplus inventory is greater than that of a lost sale, we'd like to produce the quantity that provides about 90% confidence that the actual sales will be no less than that.

In an uncertain environment, it is reasonable to expect that many decisions made do not need the expected value ("the average") of the main random variable at all. What it needs is a reasonably extreme value of that variable in order to protect the system from greater damage. Only in cases where the two damages are about the same the common sense decision should look for an assessment of the average v value.

Statistics does provide a valid mechanism for such a reasonably extreme value: The expected value plus-or-minus two or three standard deviations (SD). The actual number of the SDs depends upon your definition of "reasonably extreme value," the characteristics of the particular distribution function, and the data you have on hand (assuming the expected value and standard deviations are estimated rather than known). Only when the distribution function is known can one match the exact number of SD to a valid confidence interval.

Here is where the basic assumptions of statistics fail in reality. In order to determine the confidence interval, we need fairly good estimations of both the expected value and the standard deviation. In most cases there is not enough data to calculate a good estimation of the expected value, let alone of the standard deviation. When one cannot rely on a mathematical/statistical forecast, what is left is the intuition of the manager. Many times this is all we have.

When one has to decide how many XYZs are going to be sold during Christmas, very often last year's numbers have little to offer and prior sales are of no relevance. In such a case it is the marketing manager who supposedly has the best intuition. However, when that person is asked to make a judgment regarding the most likely sales volume, it turns out to be an thankless job. Forecasting a specific number means that you are always wrong, because no one has any idea about the value of the standard error of such a forecast. The standard deviation is not intuitive; hence, it cannot be forecasted by intuition. Even the expected value is not easy to forecast intuitively.

It is probably much more intuitive to forecast the extreme values. Why? Intuition is based on past experience. As a manager you are inclined to notice extreme values because they are problematic and not common. You're less inclined to notice actual results that fall safely within the reasonable interval you have in your mind.

Can you determine the average and standard deviation for the time that it takes you to go to work in the morning? In thinking about the average time, one is tempted to consider the interval, calculate the middle point and round it to a "nicer-looking" number. Isn't it easier to estimate the shortest time it might reasonably take and how long it might sometimes take when the all the traffic is against you?

And, it is the reasonable interval rather than the average that is needed for the decision itself. In the example above, assuming the forecast is generated by intuition, it is suggested that the marketing person closest to the market of the specified product should forecast a reasonable interval for the peak sales so that there would be about 90% confidence that the peak sales would fall in this interval, based on intuition.

That forecasted interval would be passed to the production manager, who will have to consider overall demand intervals. Produce for a peak of sales is an acknowledgment that you have a capacity constraint within the shop floor. Otherwise, you should wait to the peak time itself and produce to the actual demand! In other words, we have a temporary capacity constraint, which emerges close to the peak time and vanishes afterwards.

The production manager should devise a plan to exploit the constraint(s) based on the various market demand min-max intervals forecasted by marketing. The minimum amount per interval represents the “must have” from every product.  As long as the constraint's capacity is enough for producing all the minimums, those amounts should be produced, no matter what the “throughput per constraint minute” indicates.  This is because these minimums are “fairly certain sales” relative to higher volumes, which are “maybe sales.”

Please remember that it is assumed that we don't know the distribution function, so we are looking for a “good enough” kind of solution.  Producing for the minimum forecasted demand for all products ensures that the vast majority of what the constraint produces will be turned into real throughput.  When the peak demand approaches, it might be beneficial to get another forecast from the marketing person, hopefully a better one, that uses a smaller and more precise interval to specify 90% confidence.

Assuming we cannot produce all the maximum forecasted demand, it is necessary to decide to produce less than the maximum for certain products.  This should be based on the assessment of the damage caused by a lost sale of a certain product per constraint minute.  The most obvious damage is the loss of the potential throughput.  However, higher damage may be caused because customers may switch to a competitor, which means loss of future throughput.

**A DETAILED EXAMPLE**

Let's look at the following example:

| Product | Throughput | Constraint Minutes | Minimum Demand | Maximum Demand | Damage from a Lost Sale | Damage per Minute |
|---|---|---|---|---|---|---|
| ABC | 1 | 0.01 | 4,000,000 | 8,000,000 | 20 | 2,000 |
| DEF | 1 | 0.005 | 2,000,000 | 5,000,000 | 3 | 600 |
| XYZ | 1 | 0.002 | 1,000,000 | 4,000,000 | 1 | 500 |

The reason the damage of a lost sale of ABC is valued much higher than the actual throughput loss is because the company’s name is associated with product ABC, there is stiff competition for that product, and the ABC product is crucial for the customers.  So ABC customers are going to buy the competitive product if they don't find it on the shelves.  The other products are not so central to the company reputation, so the actual damage is closer to the loss of direct throughput.

Suppose the constraint's capacity, which may include several units, is 87,000 minutes, and we're now two months before the peak (approximately 8.5 weeks).  The basic decision should be to produce the minimum forecasted demand.  As this takes 52,000 minutes out of 87,000, we are now about 3 weeks before the peak.  Later, we may get a better forecast, with smaller intervals between the min and the max, but for now suppose that we are still following the original forecast.

We still need to specify the amounts to produce in the coming three weeks.  Adding 3,000,000 pieces of ABC (75% of the interval) will add 30,000 minutes of load on the constraint, meaning the total load will be 82,000 minutes out of 87,000.

The problem we may face now is the following.  The probability of a lost sale for the DEF product, when we decide to produce only the minimum amount, is greater than the probability of a lost sale for the ABC product, where we decide to produce 3/4 of the interval between the minimum and maximum.  Does this higher probability offset the large difference in the damage from lost sales?  It is difficult to know.

In the above situation any decision to produce 7 to 8 million ABCs, 2 to 3.5 million of DEF, and 1 to 1.5 million of XYZ will load the constraint to its maximum capacity and can be considered to be a GOOD ENOUGH decision.

This example is targeted to demonstrate our search for a valid good enough decision rather than "optimal" one. Human intuition has been used in:

a. Evaluating the reasonable range for the min-max demand.

b. Assessment of the "damage of a lost sale."

c. Fixing the concrete quantities to be produced, generally preferring the products that cause the larger damages within the specified interval.

Compare these decision-making criteria to the intuitive decision based on:

| Product | Throughput | Constraint Minutes | Average Demand |
|---|---|---|---|
| ABC | 1 | 0.01 | 6,500,000 |
| DEF | 1 | 0.005 | 3,500,000 |
| XYZ | 1 | 0.002 | 2,000,000 |

The "common sense" decision would be to produce according to the average. We won't feel the need to assess the damages of lost sales or surplus inventory. We won't push for a better forecast, because when the forecast is not based on statistical models, no quality measure is attached to it. There is more than .5 probability that the actual financial results will be worse than in the previous method, based on the same intuition. And finally, we cannot really hope that the intuition of the forecaster will be better next year, because only a small part of that intuition has been recorded and is thus available for comparison with actual results.

**A NEGATIVE BRANCH**

A reservation about the idea of basing decisions on the estimated interval is that people might tend to exaggerate the limits of the "reasonable interval," in order to be "right" every time. This enlarges protection more than necessary, causing additional expenses to the organization, and resulting in an inferior bottom line.

A possible way to relieve this negative branch is to establish two conflicting measures as a control mechanism for forecast quality. The first is the percentage of times the forecast was right, where an "ideal" measure should be 90%. Another measurement should be the ratio between the size of the interval and the midpoint [2x(BA)/(A+B)], where B and A are the limits of the interval, B is greater than A, and A and B are greater than zero. This measure, taken per forecast, influences the forecaster to shorten the gap between the minimum and the maximum, thus creating a "better" forecast for the decision maker. The down-side is that it will produce more instances where the actual result will be outside of the forecasted interval. Such a conflict forces the people in charge of forecasting to state their sincere judgment without adding personal "safety measures" to it.

These measurements should be used as a comparison between the same set of forecasted results, done in subsequent years. The idea is to show that either we have succeeded to lower the uncertainty, or we have gained a better understanding of the market demand or suppliers' lead-time we have to assess. When a person monitors a group of forecasts he or she has done in a year, the results may show whether, on the whole, the forecasts are too conservative or a too liberal.

A more detailed analysis may include a result that is very close to the boundaries of the interval as an "almost wrong" forecast. This insight comes from Buffer Management methodology, where "almost late" is treated the same as "already late." The idea is to expand the sample in order to get a better picture of the quality of the forecasts.

## THE SUGGESTED PROCEDURE

The ideas above boil down to certain basic systematic process that improves the current state of facing the common and expected uncertainty.

**Step 1.** Identify the most important random variables impacting the performance of the system and its key decisions. In classic TOC terminology it means that those variables affect the exploitation of the system constraints. Such variables include future sales, capacity, lead times, availability of certain materials, budget usage (future OE), and certain departments capabilities (flexibility to face changes in the environment).

**Step 2.** Determine the key decisions that involve the above variables. Verbalize the decision-making procedure and how an estimation of the reasonable intervals should impact the decisions.

**Step 3.** Create a systematic assessment of the uncertainty by estimating a reasonable interval of the random variables chosen. This means also designating the person in charge of the reasonable estimation of the possible values of every variable chosen in Step 1. Please note that it is not implied that the person who makes the forecast is the same person who makes the decision!

**Step 4.** Maintain a control system on the quality of the forecasts. Every reasonable interval estimation should be recorded. When the actual result is known, a classification of the reasonable interval should be made to distinguish among "within the interval," "too close to the boundaries," and "outside the interval." Every interval estimation should yield the ratio of the interval size to the mid-point.

These measures should be grouped together periodically to produce a global and departmental analysis of the forecast's quality, to be compared with the next period. The average ratios represent the magnitude of the uncertainty as viewed by the forecasters. The classification analysis indicates how reliable those forecasts are. These measures should drive the organization to lower the level of uncertainty and/or to assess it better in the future.

## THE "BUFFER" CONCEPT AS A SPECIAL CASE

The concept of a *time buffer*, is central to the DBR methodology. The shipping buffer is a protection mechanism for the reliable due-date performance. The constraint buffer provides protection that ensures availability of material at the capacity-constrained area. (The DBR methodology also maintains an assembly buffer which is, to my mind, redundant.)

The time buffer represents a "fairly long lead time." The actual lead time—the time it takes to move an order from the raw materials stock room to the protected area—is the random variable we would like to assess. This assessment dictates the release time relative to the need time reflected in the date/time a particular instruction (i.e., shipping, constraint working on a particular job) is scheduled.

The current concept of the buffer makes it clear that one should use the high end of the lead-time gross estimation as the value of the buffer. This is a "common sense" kind of idea, because in the vast majority of the cases, the damage from a late order is significantly larger than the damage from a somewhat earlier release of materials. Fixing that high-end lead time estimation reduces the overall level of work in process (WIP). Using uncertainty estimation, the need for excess and redundant protection is removed, while still providing adequate due-date protection. How good is the buffer estimation? Once the question is posed in this sense, making it clear that the buffer is an extreme assessment of the lead-time, it lends itself to being a control mechanism.

Buffer Management is a methodology that monitors the actual arrival of material to the protected area. One of its objectives is to monitor the effectiveness of the buffer as a protection mechanism that answers the question, "Is the current size of the buffer about right or significantly wrong?" This is the same general approach as suggested by Step 4 of the procedure above.

DBR ignores the minimum side of the "reasonable interval." However, those familiar with the specifications of *The Goal System*, DBR software (Goldratt, 1990), may recall the prescribed use of "half of the buffer size" as a way to describe "short enough but still reasonable lead time." The concept of "half the size of the buffer" is used in two different applications. The first is in the subordination step, where the feasibility of the constraint's schedule is checked. As long as half of the original buffer is still available, it is assumed that the schedule is realistic. When this assumption is found to be invalid, it points to the possibility that another capacity constraint may be emerging.

The second application deals with the schedule of constraint's operation that one feeds the other. The minimum displacement in time is determined to be "half of the buffer," assuming that with the right type of priorities this amount of time is still feasible.

Suppose that the production manager should assess a minimum and maximum lead times between the release and the critical areas as defined by the DBR methodology. It is clear that the high end of that estimation should be used to schedule the material release. However, when there is a need to estimate whether it is still feasible to complete an order on time, assuming expediting efforts are taken, the value of the minimum time is of great value. This concept of "fairly short lead-time that is still feasible" seems to be more accurate than the "half of the buffer" concept. This would also clarify the role of the buffer as a protection mechanism. The current definition makes it difficult because there is no distinction between the minimum time that an order has to be processed and protection against "Murphy." By establishing an interval for the lead time that represents the difference between the net processing time (the main contributor for any minimum lead time estimation) and the maximum estimation (which is the concept of the time buffer), one can estimate the amount of buffer allowed for "Murphy."

## SUMMARY AND CONCLUSIONS

The primary contention of this paper is that the use of an intuitive estimate of a confidence interval can lead to superior decision making in cases where no hard data exist to support routine statistical procedures.

It is suggested that most decisions rely on the limits of such an interval rather than on its central measures. This is a more natural way to intuitively estimate the uncertainty, quantify it and use it in the decision making process.

Better understanding the nature of uncertainty may lead to better controlled protection needed for appropriate behavior in every organization. This will eventually lead to much better utilization of the constraint and much less Operating Expense and Inventory. Further research may show benefits of using this same idea in different functions, such as budget management and strategic planning.

A future study might explore cases when complex situations require the combination of several intuitive estimations. Current procedures assume that the individual cases are independent, thus the global average and standard deviation can be calculated. In reality, in most cases there is some dependency between the individual elements of complex systems. A more realistic way is needed to intuitively assess the combination of individual uncertain variables. One open question is whether intuition is better in estimating the parts or in estimating the total. Or maybe both are needed. In any event, more focused research on the quality of intuitive estimation of uncertainty is badly needed.

**REFERENCE**

Goldratt, E. M. (1990). The Haystack Syndrome. Croton-on-Hudson, NY: North River Press.