UCRL-TR-216207

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Confidence Probability versus Detection Probability

M. Axelrod

October 17, 2005

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Confidence Probability versus Detection Probability

Michael Axelrod

January 27, 2003

Introduction

In a discovery sampling activity the auditor seeks to vet an inventory by measuring (or inspecting) a random sample of items from the inventory. When the auditor finds every sample item in compliance, he must then make a confidence statement about the whole inventory. For example, the auditor might say: "We believe that this inventory of 100 items contains no more than 5 defectives with 95% confidence. Note this is a retrospective statement in that it asserts something about the inventory after the sample was selected and measured. Contrast this to the prospective statement: "We will detect the existence of more than 5 defective items in this inventory with 95% probability." The former uses *confidence probability* while the latter uses *detection probability*. For a given sample size, the two probabilities need not be equal, indeed they could differ significantly. Both these probabilities critically depend on the auditor's prior belief about the number of defectives in the inventory and how he defines non-compliance. In other words, the answer strongly depends on how the question is framed.

Confidence Probability

We define *confidence probability*, as the conditional probability that the number of defective items in the target inventory is no greater than a specified *tolerance limit*.

$$\alpha = P\left[K \le t \middle| X = x\right] \tag{0.1}$$

In (0.1) α is the confidence probability, K is the unknown number of defectives in the whole inventory, and x is the number of defectives measured in the sample and t is the tolerance limit. Note that both K and X are random variables, while x and t are integer-valued numbers. In the absence of data (before we draw the sample), our prior confidence probability is:

$$P\left[K \le t\right] = \sum_{k=0}^{t} w(k) \tag{0.2}$$

In (0.2) $w(k)$ is the prior probability (mass) function that *exactly* k items are defective. Thus $w(k)$ reflects our prospective knowledge about the inventory, while α reflects our retrospective knowledge given the results of measuring the sample. Using the hypergeometric distribution and Bayes' formula, we can write the confidence probability as:

$$\alpha = \frac{\sum_{k=0}^{t} \binom{k}{x} \binom{N-k}{n-x} w(k)}{\sum_{i=0}^{N-n} \binom{i}{x} \binom{N-i}{n-x} w(i)}. \tag{0.3}$$

where n is the size of the sample and N is the size of the inventory. When the sample has no defectives $x = 0$, and (0.3) becomes:

$$\alpha = \frac{\sum_{k=0}^{t} \binom{N-k}{n} w(k)}{\sum_{i=0}^{N-n} \binom{N-i}{n} w(i)}. \tag{0.4}$$

An important special case occurs when the prior probability $w(k)$ becomes the discrete uniform distribution, then (0.4) becomes:

$$\alpha = 1 - \frac{N-n-t}{N+1} \frac{\binom{N-t}{n}}{\binom{N}{n}}. \tag{0.5}$$

Detection Probability

We define *detection probability* as the conditional probability that we will detect non-compliance by seeing one or more defective items in the sample. In contrast to confidence probability, detection probability is prospective. The conditioning event refers to the true state of the inventory instead the observed state of the sample.

$$\beta = P\left[X > 0 \middle| K > t \right] \tag{0.6}$$

Using the elementary formulas for conditional probability, we can write (0.6) as:

$$\begin{aligned} P\left[X > 0 \middle| K > t \right] &= 1 - P\left[X = 0 \middle| K > t \right] \\ &= 1 - \frac{\sum_{k=t+1}^{N} P\left[X = 0 \middle| K = k \right] P\left[K = k \right]}{\sum_{k=t+1}^{N} P\left[K = k \right]} \end{aligned} \tag{0.7}$$

We have seen all the probabilities in (0.7) before. The first formula below comes from the hypergeometric distribution, while the second formula is the prior distribution discussed above.

$$P\left[X=0\middle|K=k\right]=\frac{\binom{N-k}{n}}{\binom{N}{n}} \quad 0\le k\le N-n$$
$$P\left[K=k\right]=w\left(k\right). \tag{0.8}$$

Therefore, the detection probability is:

$$\beta=1-\frac{\sum_{k=t+1}^{N-n}\binom{N-k}{n}w\left(k\right)}{\binom{N}{n}\sum_{i=t+1}^{N}w\left(i\right)} \tag{0.9}$$

When the prior distribution is uniform, we get

$$\beta=1-\frac{N-n-t}{\left(n+1\right)\left(N-t\right)}\frac{\binom{N-t}{n}}{\binom{N}{n}} \tag{0.10}$$

as the companion equation to (0.5). With these two equations, we can get a relation between confidence probability and detection probability:

$$\frac{1-\alpha}{1-\beta}=\frac{\left(n+1\right)\left(N-t\right)}{N+1}. \tag{0.11}$$

From (0.11)we see that detection probability always exceeds confidence probability. For example, if the inventory has $N=100$ items and we use a tolerance limit of $t=5$ (5%), then we need $n=10$ samples for a detection probability of 95%. But, our confidence probability is then only about 50% even if all 10 items in the sample conform. Thus, a good design for detection is generally not a good design for confidence. On the other hand, if we use $n=37$, our confidence probability climbs to 95%, while the detection probability hits 99.85%. However, this second design is almost four times as expensive in terms of sample size. Thus, if a uniform prior is the operating assumption, the sample size critically depends on whether one is interested in confidence probability or detection

probability. You could end up with either an overly expensive design or one that does not meet the need of a discovery audit where confidence probability is the key result.