UCRL-TR-216206

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

# Using Ancillary Information to Reduce Sample Size in Discovery Sampling and the Effects of Measurement Error

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.

# Using Ancillary Information to Reduce Sample Size in Discovery Sampling and the Effects of Measurement Error

Michael Axelrod

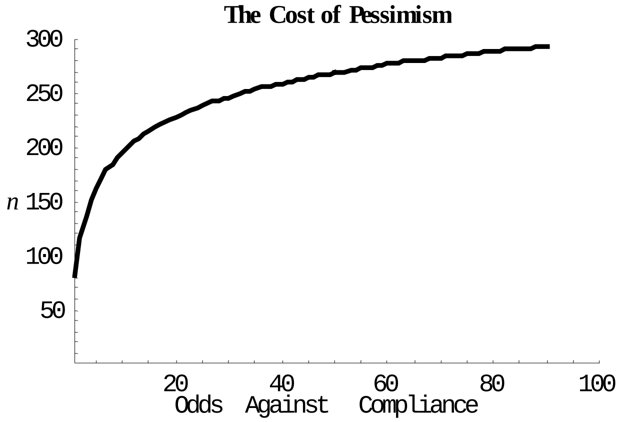
March 10, 2002

## Summary

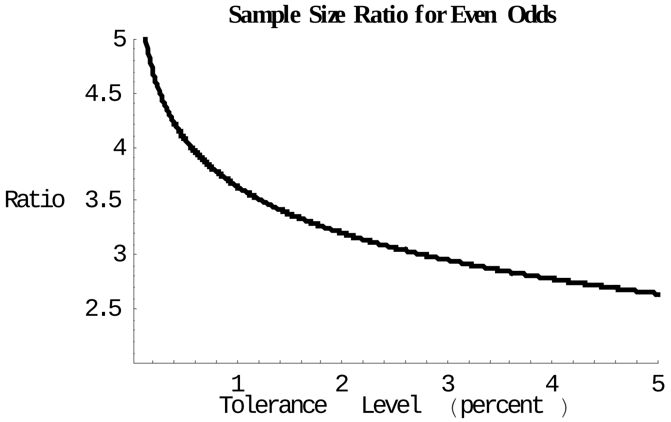
The classical statistical method of confidence intervals is unsuitable for discovery sampling because it requires unnecessarily large sample sizes. We show that sample sizes can be dramatically reduced if we are willing to relax an extremely pessimistic assumption that underlies the method. We also treat the effect of measurement error on sample size. This summary section states the main results, and can be read as a stand-alone. The body of this report together with the appendices provide further technical details.

Discovery sampling is a tool used in discovery *auditing*. In a discovery audit the auditors measure a random sample of items from an inventory to provide evidence that the whole inventory complies with a given set of criteria. Auditors expect the items in a sample to be free of compliance problems because they believe that the entire inventory has very few (if any) defects. As part of their work product, auditors are usually required to provide a *confidence statement* about the inventory such as: “We are 95% confident that less than 1% of the items in the inventory are defective.” The standard statistical tool for making this kind of statement uses the classical method of *confidence intervals*. A confidence interval brackets the estimated number of defects in the inventory based on the number of defectives in the sample (usually zero in a discovery audit). The size of the sample is chosen so the *tolerance level* (maximum number of defectives) appears at the upper end of the confidence interval. If the interval would span the actual number of defects in (say) 95% of future audits, then it’s called a “95% confidence interval.” Discovery audits usually specify tight tolerances such as 1% or even 0.1%. Tight tolerances come with a steep price: large sample sizes. For example an auditor would need to sample nearly a quarter of a large inventory to achieve 95% confidence for a 1% tolerance. The reason classical confidence intervals require large samples stems from an implicit assumption which becomes evident once a Bayesian framework is adopted. The assumption becomes very pessimistic for tight tolerance audits. For example if a 1,000-item inventory is to be vetted to a 1% tolerance, use of classical confidence intervals implicitly assumes that there is a 98.9% chance the inventory is *out* of compliance before the auditors draw and measure the sample. In other words, the sample information must shift the odds from 99:1 *against* compliance to 20:1 *for* compliance— a factor of almost 2,000 change in the odds. It’s no wonder that of confidence intervals require such large sample sizes, they must rebut an extremely pessimistic assumption. However by incorporating ancillary information about the inventory, expressed in terms of the prior

probability of compliance, we can dramatically reduce the required sample size. The figure below (based on a large inventory approximation) shows the cost of pessimism in terms of *n*, the sample size needed to achieve a 95% confidence for a tolerance of 1%.

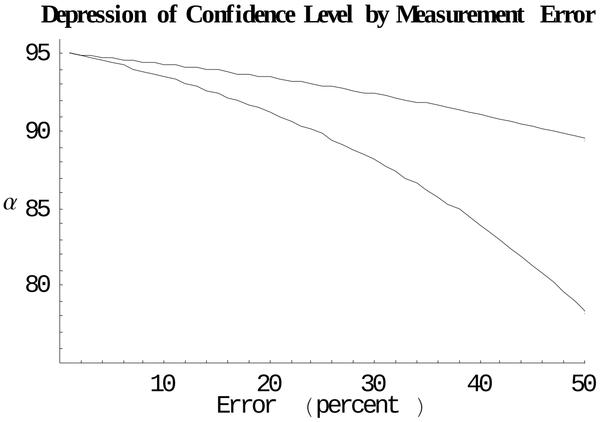


The method of confidence intervals operates at the extreme right of the curve, at the point corresponding to 99:1 prior odds against compliance. But we can operate anywhere along the curve by using the method of *Bayesian confidence*. Using Bayesian confidence, we can reduce the prior odds against compliance, causing sample size to drop dramatically. At even odds (extreme left of the curve) *n* plummets to 80 from nearly 300. If one goes further and accepts (say) 2:1 odds *for* compliance (not shown), the sample size shrinks to about 50. The figure below shows the cost in another way. In this figure (also based on a large inventory approximation), we assume even odds against compliance and calculate ratio of the sample sizes for Bayesian confidence to classical confidence intervals. The ratio is plotted against the tolerance levels usually encountered in a discovery audit. We see that the ratio grows explosively for extremely tight tolerances (below 1%). Even at a tolerance of 5% the ratio exceeds 2.5.

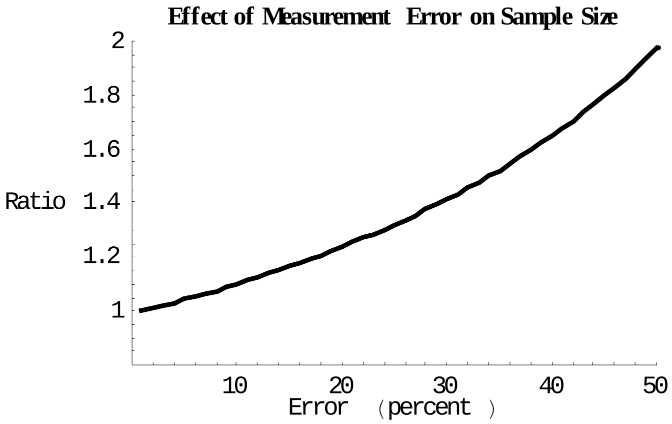


For finite inventories the curve above does not “blow up” at zero-tolerance. Indeed discovery sampling is both feasible and practical for zero-tolerance if one is willing to be somewhat optimistic. Given 2:1 odds *for* compliance, we can vet a 1,000-item inventory at zero-tolerance using a sample with only 400 items instead of the 950. Normally a 400-item sample would only vet at 1% tolerance. Note that “last one percent” is prohibitively costly using classical confidence intervals.

We also analyze the effects of measurement error. A sample could appear free of defectives because measurement error causes defective items to be misclassified as good items. Consequently, auditors will overstate the confidence. The figure below illustrates the depression on confidence level caused by measurement error. The lower curve corresponds to a classical 95% confidence interval for a tolerance of 1%. Note that a 25% error rate converts a 95% confidence interval into a 90% confidence interval. Above 30%, errors have a catastrophic effect on the confidence level, causing the audit to provide extremely misleading results. However using Bayesian confidence provides significant protection against measurement error as shown by the upper curve which corresponds to even odds against compliance. Note that the 95% confidence interval is now converted into a 93% confidence interval. Moreover, at least 90% confidence is preserved up to an error of 46%.



If the measurement error is known, then the sample size can be adjusted upward to restore the desired confidence level. At low error rates (below 12%) there is an approximately linear upward adjustment. A 10% error requires an approximate 10% increase in the sample size to maintain the confidence level. In the figure below we show the amount of upward adjustment versus error rate to preserve 95% confidence for 1% tolerance. The adjustment is expressed as the ratio of the sample size with error to the error-free sample size. Again, we see that for classical confidence intervals the increase in sample size grows explosively for errors that exceed 30%.



We have only considered errors where a defective item is misclassified as good. The converse, where a good item is misclassified as defective is generally not a problem in a discovery audit. Auditors will usually re-measure apparently defective items. Some auditors adopt a specific policy of multiple measurements for all items that fail a particular test. Consequently, errors that would to make the inventory look worse than it actually is tend to be corrected.

To use Bayesian confidence auditors need to choose a prior distribution for the number of defects in the inventory and some people consider this a problem as the choice of a prior can seem somewhat arbitrary. However, using classical confidence intervals does not circumvent this problem, it merely hides it by choosing the most conservative worst-case assumption: the uniform prior distribution. We advocate a workable one-parameter prior that we think makes sense in the special circumstances present in a discovery audit. However, one should bear in mind that the information from the sample adds to the prior, it does not validate it. The uniform distribution serves as a zero-information baseline prior, so the final statement of confidence is always conservative. Much too conservative! Our calculations show that even a very modest relaxation from the uniform prior yields significant savings in terms of sample size and protection from measurement error. There is of course some risk in using a non-uniform prior, but assuming a small risk for a large payoff (if samples are expensive in some sense) is a prudent practice. In some cases, it is the optimum strategy because resources can be diverted from taking samples to improving the measurement system, thus yielding a higher level of confidence. In discovery sampling extreme risk aversion can be counter productive.

We provide formulas to calculate the sample size for discovery sampling. To use the formulas auditors must make policy decisions in choosing the confidence level, and tolerance level, typically 95% and 1%. Next, they must use their knowledge about the inventory to choose the prior odds against compliance. They need only have a rough idea of these odds. We think even odds will cover many applications. Our large inventory approximation indicates a sample of size 82 for these specifications. Finally auditors need some knowledge of measurement errors. If the errors are less than 20%, a sample size of 100 will suffice. One hundred, the bottom-line!

## Introduction

Discovery sampling is a tool used in a discovery *auditing.* The purpose of such an audit is to provide evidence that some (usually large) inventory of items complies with a defined set of criteria by inspecting (or measuring) a representative sample drawn from the inventory. If any of the items in the sample fail compliance (defective items), then the audit has discovered an impropriety, which often triggers some action. However finding defective items in a sample is an unusual event— auditors expect the inventory to be in compliance because they come to the audit with an "innocent until proven guilty attitude." As part of their work product, the auditors must provide a confidence statement about compliance level of the inventory. Clearly the more items they inspect, the greater their confidence, but more inspection means more cost. Audit costs can be purely economic, but in some cases, the cost is political because more inspection means more intrusion, which communicates an attitude of distrust. Thus, auditors have every incentive to minimize the number of items in the sample. Indeed, in some cases the sample size can be specifically limited by a prior agreement or an ongoing policy. Statements of confidence about the results of a discovery sample generally use the method of *confidence intervals*. After finding no defectives in the sample, the auditors provide a range of values that bracket the number of defective items that could credibly be in the inventory. They also state a level of confidence for the interval, usually 90% or 95%. For example, the auditors might say: "We believe that this inventory of 1,000 items contains no more than 10 defectives with a confidence of 95%."

Frequently clients ask their auditors questions such as: How many items do you need to measure to be 95% confident that there are no more than 10 defectives in the entire inventory? Sometimes when the auditors answer with big numbers like "300," their clients balk. They balk because a big sample size might bust the budget, or the number seems intuitively excessive. To reduce the sample size, you can increase the tolerable number of defectives, the "10" in the preceding example, or back off on the confidence level, say from 95% to 90%. Auditors also frequently bump up the sample size as a safety factor. They know that something can go wrong. For example, they might find out that the measurements or inspections were subject to errors. Unless the auditors know exactly how measurement error affects sample size, they might be forced to give up the safety factor. Clients often choose to "live dangerously" (without a compelling argument to the contrary) to save money. Thus, sometimes the auditor finds that "you just can't get there from here," because the goals of the audit and the resources available are inherently in conflict. For discovery audits, there is a way out of this apparent conundrum. It turns out that the classical method of confidence intervals uses an implicit and very conservative assumption. We will see that this assumption is too pessimistic and too conservative in the context of a discovery audit. If we abandon this assumption and use ancillary information about the inventory, then we can significantly reduce the sample size required to achieve the desired confidence level. We will see exactly how the classical method ignores this ancillary information and misses the opportunity for an efficient audit.

In the following sections, we first review the standard approach using confidence intervals. Then we present a method that incorporates the ancillary information about the inventory to design a very efficient discovery audit. We also provide results on how measurement errors affect the audit, and how exactly how much the sample size must be modified to compensate for these errors. Finally, we state asymptotic formulas that provide useful approximations for large inventories. It is suggested that the reader review the glossary of symbols while reading the body and the appendices as there are numerous special symbols and notations used in the text.

## The Confidence Interval Approach[1]

Suppose *k* items out of an inventory of size *N* items are defective. We would like to say something about the unknown *k* given that all the items in a random sample of size *n* pass inspection. The best estimate of *k* is zero, but how good is this estimate? We measure goodness by the width of a confidence interval that would contain the true value of *k* in $\alpha \times 100\%$ repetitions of the sampling process. A narrow interval is a good interval, a wide interval means that the estimate of *k* is not well determined by the sample. So [0,1] is much better than [0,10] if they are both 95% confidence intervals. The lower limit of the confidence interval is always zero because *k* must be non-negative. To calculate the confidence interval when $n < N$ we use the hypergeometric distribution that describes random sampling from the inventory *without replacement*. This distribution gives the probability *p* that a random sample of size *n* will have *x* defective items. The formula for *p* is given by:

$$p(x,n,N,k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \quad \max(0,n+k-N) \le x \le \min(n,k) \tag{0.1}$$

We know both *N* and *n* beforehand, and we know *x* from the sample, but we don't know *k*. Somehow we have to "invert" (0.1) so we can say something about *k.* We do this as follows. In the typical discovery audit, the sample has no defectives so we put $x = 0$ in (0.1) and find the smallest integer $k^*$ that satisfies[2]:

$$p(0,n,N,k) = \frac{\binom{N-k}{n}}{\binom{N}{n}} \le 1-\alpha \tag{0.2}$$

Then the confidence interval is $\left[0,k^*-1\right]$. For example if $N = 1000$, $n = 238$, and $\alpha \times 100 = 95\%,$ the smallest value of $k^*$ that satisfies

[1] A complete list with explanations of all the mathematical symbols used here is given in the Appendix.

[2] *Technometrics*, Vol. 10, No. 1 (1968). For the basic ideas behind confidence intervals see *Mathematical Statistics*, Bickel and Doksum, Prentice Hall (1977). Some discussions of confidence limits for the hypergeometric distribution might give an upper confidence limit that differs by 1 from the formula presented here. This is because the maximum likelihood estimate of *k* is not unique when $x(N+1)/n$ is an integer.

$$p(0,238,1000,k^{*}) = 0.0494 \le .05 \qquad (0.3)$$

is $k^{*} = 11,$ which gives a confidence interval of [0,10] for the unknown parameter *k*. Note that upper value of the confidence interval is 1% of the total size of the inventory, the *tolerance level*. If we want the upper end of the confidence interval to be 1 instead of 10 then the sample size grows to $n = 776$, almost 80% of the inventory. For zero-tolerance we need $[0,0]$ which corresponds to $k^{*} = 1$ so we need $n = 950,$ or 95% of the inventory. This result applies in general. To have $\alpha \times 100\%$ confidence for zero-tolerance you have to sample $\alpha \times 100\%$ of the inventory. The message is clear: tight tolerances are very costly in terms of sample size. But discovery audit clients usually want tight tolerances, so we need the more efficient Bayesian confidence.

### Bayesian Approach to Gaining Confidence

Bayesian statistical methods use a subjective interpretation of probability. In this framework, “subjective probability” quantifies the degree of belief in the truth of some assertion. At 100%, we have absolute certainty in the truth of the assertion. At 50%, we have the same certainty that a coin flip will yield a “heads.” Since “confidence” is a degree of belief, Bayesian methods are a natural to discovery auditing where we want to gain confidence about an inventory. The Bayesian framework allows us to treat the parameter *k* (the total number of defectives) in (0.1) as a random variable *K*. We are confident about the inventory when we have a high probability that *K* is less than the tolerance level. Thus the fundamental calculation in discovery sampling is the following: given *x* (usually zero in a discovery audit), what is the probability $\alpha$ that $K \le k_{\text{tol}}$ where $k_{\text{tol}}$ is the tolerance level? In planning a discovery audit, we specify $k_{\text{tol}}$ and $\alpha$ as a policy decision and calculate the sample size *n* we need to achieve them. The customary value for $\alpha$ is 90% or 95%. The tolerance level is governed by the costs of having that many defectives present in the inventory. If the cost of even a single defective is catastrophic then zero tolerance $\left(k_{\text{tol}} = 0\right)$ makes sense, but then the sample size can be very large resulting in a very costly audit. In using the Bayesian approach, we need a prior distribution $w(k)$ for *K*. This is the probability that that the inventory has *exactly k* defective items before we observe the sample. Thus $w(k)$ represents the ancillary information we have about the inventory. If we believe the inventory has no defects, then $w(k) = 0$ except $w(0) = 1.$ In this case, we would not need to draw a sample at all because we would have absolute confidence that no defectives exist. The other extreme prior belief is one of maximum ignorance, where we have no ancillary information. In this case, the $w(k)$ is the discrete uniform distribution:

$$w(k) = \frac{1}{N+1} \quad k = 0,1,2\ldots N. \tag{0.4}$$

With a uniform prior the confidence for a zero-defect sample is:

$$\alpha = P\left[K \le k \middle| X = 0\right] = 1 - \frac{N-n-k}{N+1} \frac{\binom{N-k}{n}}{\binom{N}{n}} \tag{0.5}$$

(derived in Appendix I). Note that for large $N$ (0.5) is essentially the same as (0.2). Thus from a Bayesian viewpoint the classical confidence interval approach assumes a uniform prior distribution for the number of defectives. In Table 1 we present the sample sizes needed to achieve 95% confidence for a 1% tolerance. Note the nearly identical sizes both the classic and Bayesian approaches.

**Table 1 Sample Size for 95% Confidence**

| $N$ | $k$ | $n$<br>(Classical) | $n$<br>(Uniform Prior) |
|---|---|---|---|
| 10,000 | 100 | 291 | 290 |
| 1,000 | 10 | 238 | 237 |
| 500 | 5 | 196 | 195 |
| 200 | 2 | 126 | 125 |
| 100 | 1 | 78 | 77 |

The result holds generally for all tolerance levels. Thus if $k_{\text{tol}}$ is the upper end of a $100 \times \alpha$ percent confidence interval, then the probability that $K \le k_{\text{tol}}$ is also $100 \times \alpha$ percent assuming a uniform prior for $w(k)$.

In a discovery audit, we generally expect the number of defectives in the inventory to be extremely small even zero. From this viewpoint, we see that the assumption of a uniform distribution for $w(k)$ is extremely pessimistic and unduly conservative. For example in an inventory of 1,000 items a uniform prior assumes the auditors believe that there is a 98.9% chance that more than 10 items in the inventory are defective. A more credible distribution for $w(k)$ would one that decreases with increasing *k*, and where $w(0) \gg w(N)$. It makes sense to use a rapidly decreasing function for $w(k)$ to reflect the belief that it is much more likely that the inventory has few defectives instead of many defectives. If we actually believed that a thousand-item inventory had a 98.9% chance of being out of compliance, we would not be designing a discovery audit in the first place. A discovery audit is a vetting process; auditors want to confirm their expectation that the inventory is in compliance. Recall that the presence of a defective in a discovery sample aborts the normal audit process because an unexpected event has occurred and some action must be taken. If a defective item in a sample does not abort the process then the activity is not really a discovery audit. The activity is then more akin to a quality inspection where the goal is to estimate the number of defectives in the inventory. The use of ancillary information still applies, but the design of the sampling will differ from a discovery audit. It should be clear to the practitioners what the goal of their activity is before they design the sampling plan.

We choose the beta-binomial distribution (discrete form of the beta distribution) because it provides a flexible family of prior distributions:

$$w(k) = \frac{\binom{k+b-1}{b-1}\binom{N-k+a-1}{a-1}}{\binom{N+a+b-1}{a+b-1}} \qquad k = 0,1,2,\ldots N \tag{0.6}$$

In (0.6) the parameters *a* and *b* are non-negative and can be either integer or real. When both *a* and *b* are 1 then (0.6) reduces to the uniform distribution:

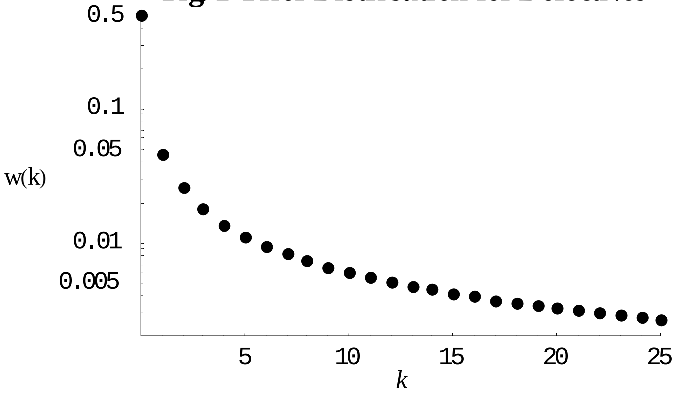
$$w(k) = \frac{1}{N+1} \quad k = 0,1,2,\ldots N \tag{0.7}$$

Setting $a = 1$ gives a useful family of priors for discovery sampling design, and (0.6) becomes:

$$w(k) = \frac{\binom{k+b-1}{b-1}}{\binom{N+b}{b}} \tag{0.8}$$

When *b* is non-integer (0.8) must be evaluated using the gamma function which generalizes the factorial function to non-integer arguments. In Fig. 1 we have a plot of a beta-binomial prior with $N = 1000,$ and *b* is set such that $P\left[K \le 10\right] = 65\%$; only the first 25 points are shown. Note that the vertical axis is a log scale so the probability decreases very rapidly with increasing *k* (with a linear scale and the entire *k*-axis out to 1000, the function would look like a "spike" $k = 0.$ This function is a proper probability mass function as it sums exactly to 1. The $w\left(k\right)$ in Fig. 1 expresses a 35% chance against compliance instead of 98.9%.

**Fig 1 Prior Distribution for Defectives**



After we observe a sample with no defectives the Bayesian confidence is:

$$\alpha = P\left[K \le k \middle| X = 0\right] = \frac{\sum_{j=0}^{k} \binom{N-j}{n} w\left(j\right)}{\sum_{i=0}^{N-n} \binom{N-i}{n} w\left(i\right)}, \tag{0.9}$$

(see Appendix I). Using the prior shown in Fig.1 in (0.9) with $k = 10,$ and $N = 1,000,$ the smallest value of *n* that gives at least 95% confidence is $n = 51.$ In other words, we attain the same level of confidence with 51 samples using an optimistic prior as we did using classical confidence intervals with $n = 237.$ If we use a less optimistic prior, one with even odds against compliance (50% probability) then sample sizes increases modestly to $n = 76.$

Table 2 provides a sample size comparison for 1% tolerance and 2:1 odds for compliance. We can see a significant drop in the required sample size to achieve the same level of confidence.

**Table 2 Sample Sizes for 95% Bayesian Confidence**

| *N* | *k* | *b* | *n*<br>(Classical) | *n*<br>(Bayesian) |
|---|---|---|---|---|
| 10,000 | 100 | 0.0936532 | 290 | 54 |
| 1,000 | 10 | 0.0946347 | 237 | 51 |
| 500 | 5 | 0.0957091 | 196 | 47 |
| 200 | 2 | 0.0988258 | 126 | 40 |
| 100 | 1 | 0.103674 | 77 | 31 |

Bayesian confidence can make zero-tolerance sampling practical. Given a 1,000-item inventory, a sample size of 390 is sufficient to establish zero-tolerance, far less than the 950 samples required in a classical confidence interval.

### Infinite Inventories

As the size $N$ of the inventory increases, the distribution of $X$ approaches the binomial from the hypergeometric:

$$\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \to \theta^{x}\left(1-\theta\right)^{n-x}\binom{n}{x} \qquad 0<\theta<1. \tag{0.10}$$

In the RHS of (0.10) $\theta$ replaces *k* on the LHS. The parameter $\theta$ is the *rate* of occurrence of defectives whereas *k* is *number* of defectives. Similarly, the discrete distribution used as a prior for *k* is replaced by its continuous version for $\theta$. The mathematical details are provided in Appendix II. Here are the results.

The sample size needed for the $\alpha\times 100\%$ classical confidence interval $\left[0,\theta\right]$ is the minimum value of *n* that satisfies:

$$\left(1-\theta\right)^{n} \le 1-\alpha \tag{0.11}$$

The Bayesian confidence $\alpha$ using a uniform distribution for $\Theta$ (the random version of $\theta$) is the minimum value of *n* that satisfies:

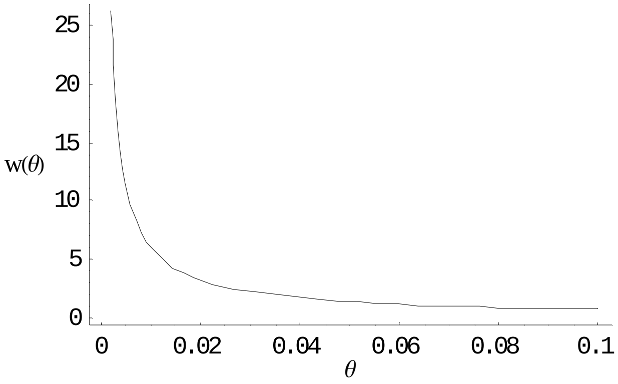
$$\left(1-\theta\right)^{n+1} \le 1-\alpha \tag{0.12}$$

For $\theta=.01$ and $\alpha=.95$ the required sample size is *n*= 298, close to the sample size of 291 required for *N*= 10,000. We see that (0.12) and (0.11) are essentially equivalent, so just as for finite inventories, the classical confidence interval uses the uniform prior. Using a beta distribution prior of the form $w\left(\theta\right)=b\theta^{b-1}$, the Bayesian confidence is:

$$\alpha = P\left[\Theta\le\theta \middle| X=0\right] = b\binom{n+b}{b}B_{\theta}\left(b,n+1\right) \tag{0.13}$$

where $B_{\theta}\left(\cdot,\cdot\right)$ is the incomplete beta function. The incomplete beta function is a three argument special *function*; it should be distinguished from the beta *distribution*. In Fig. 2 we show a plot of $w\left(\theta\right)$ versus $\theta$ corresponding a value of *b* that gives 2:1 odds in favor of compliance at a tolerance of 1% (uniform distribution corresponds to $b=1$). With these odds, *n* plummets from 291 to 54. Again the change for even odds is moderate, $n=82$. However, the binomial approximation cannot be used for zero-tolerance sampling.

**Fig. 2 Beta Density Function for W(.01) = 65%**

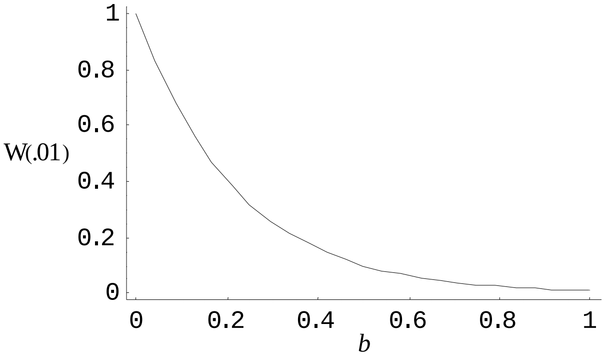


To choose a prior the auditors select the tolerance level $\theta_{\text{tol}}$ and the prior probability of compliance $W\left(\theta_{\text{tol}}\right)$ where:

$$W\left(\theta_{\text{tol}}\right) = P\left[\Theta \le \theta_{\text{tol}}\right] = \int_0^{\theta_{\text{tol}}} w\left(t\right) dt. \tag{0.14}$$

For example, if we choose $\theta = .01$ and $W\left(.01\right) = 0.65,$ then we see from Fig. 3 that the value of *b* is uniquely determined. Note that for classical confidence intervals $W\left(\theta_{\text{tol}}\right) = \theta_{\text{tol}} \Rightarrow b = 1$ and $w\left(\theta\right) \equiv 1.$

**Fig. 3 P[Θ≤.01] for Beta Prior as a Function of b Parameter**



Note that we are using a one-parameter family of beta priors. Using the two-parameter beta distribution would provide more control over the shape of $w\left(\theta\right),$ but in general that degree of detail is unnecessary, and for most cases the auditors don't have enough ancillary information to determine two parameters.

## Robustness Considerations

Suppose the auditors choose the wrong value of *b*, or equivalently the wrong prior odds against compliance for a specified tolerance level. What is the consequence for the confidence of the audit? If the value of *b* exceeds the true value of *b* then the stated confidence is less than the true confidence. The auditors will think they have (say) 95% confidence when they actually have something greater, such as a 98% confidence. The reverse is true if the auditors choose a value of *b* that is too low. However, this is unlikely to happen if *b* is chosen conservatively, which is easy to do for tight tolerances because tight tolerances always give excessively high odds against compliance for the uniform $w(\theta)$. Suppose $\Theta$ does not have a beta distribution, what are the consequences? In general, if the true distribution is monotonic decreasing and always less than the beta distribution corresponding to the chosen *b*, the calculated confidence will again be conservative. If we know nothing about the true distribution then all we can say is the classical confidence still has excessive sample size, but we don't know how excessive.

## Measurement Error

A *false positive* error occurs when a good item measures as being defective. This kind of error tends to be corrected because auditors will often re-measure apparently defective items suspecting that the problem lies with the measurement, not the item. On the other hand, truly defective items that measure as being good (*false negative* errors) tend not to get corrected because apparently good items don't usually receive further scrutiny in a discovery audit. False negatives make the inventory look better than it actually is, and the true level of confidence is smaller than the apparent level of confidence. Therefore, we need larger sample sizes in the presence of measurement error to achieve the desired level of confidence from the audit. We assume that false negatives occur randomly with a known probability, and that false positives never occur. These are approximate, but useful assumptions. We provide formulas to correct the sample size in Bayesian discovery sampling for both finite and infinite inventories. We also provide sample size corrections for classical confidence intervals. The results use some special functions: the hypergeometric function, and the incomplete beta function. Values for these functions are available in standard tables, or from readily available commercial packages that run on desktop computers. We also provide tables of corrected sample sizes for selected examples. The following sections provide the main results, mathematical details are covered in the appendices.

## Measurement Errors for Finite Inventories

With measurement error (0.2) becomes:

$$p\left(0,n,k,q\right)=\frac{\binom{N-k}{n}}{\binom{N}{n}}\,{}_2F_1\left(-k,-n,N-n-k+1,q\right)\le 1-\alpha \qquad (0.15)$$

where $q$ is the probability of a false negative error, and ${}_2F_1\left(\cdot,\cdot,\cdot,\cdot\right)$ is the hypergeometric function. To get classical $\alpha\times 100\%$ confidence interval for *k*, find the smallest value $k^*$ that satisfies the inequality (0.15). Then the interval is $\left\lfloor 0,k^*-1\right\rfloor$. The Bayesian confidence is given by a similar modification of (0.9)

$$\alpha = P\left[K \le k \middle| Y = 0\right] = \frac{\sum_{j=0}^{k} \binom{N-j}{n} {}_2F_1\left(-j,-n,N-n-j+1,q\right) w\left(j\right)}{\sum_{i=0}^{N-n} \binom{N-i}{n} {}_2F_1\left(-i,-n,N-n-i+1,q\right) w\left(i\right)} \tag{0.16}$$

In Table 3 we show the effects of measurement error on the sample size for classical and Bayesian confidence. Over the indicated range of measurement error, Bayesian with uniform prior matches classical. Table 3 also shows how measurement error affects the Bayesian confidence using the same beta-binomial prior as in Table 2. Note that Bayesian confidence attenuates the effect of measurement error.

**Table 3 Sample Size for 95% Confidence Interval With Indicated Measurement Error**

| *N* | *k* | *q* | *n*<br>Classical | *n*<br>Uniform | *n*<br>65% |
|---|---|---|---|---|---|
| 1000 | 10 | 0% | 237 | 237 | 51 |
| 1000 | 10 | 5% | 250 | 250 | 53 |
| 1000 | 10 | 10% | 264 | 264 | 56 |
| 1000 | 10 | 15% | 280 | 280 | 60 |
| 100 | 1 | 0% | 78 | 78 | 31 |
| 100 | 1 | 5% | 82 | 82 | 33 |
| 100 | 1 | 10% | 87 | 87 | 35 |
| 100 | 1 | 15% | 92 | 92 | 37 |

## Measurement Errors for Infinite Inventories

With measurement error, the defining equation for the classical confidence interval for infinite inventories corresponding to (0.11) becomes:

$$\left(1-\left(1-q\right)\theta\right)^{n} \leq 1-\alpha. \tag{0.17}$$

Similarly with measurement error the Bayes confidence for a uniform prior corresponding to (0.12) becomes:

$$\left(1-\left(1-q\right)\theta\right)^{n+1} \leq \left(1-q^{n+1}\right)\alpha. \tag{0.18}$$

Thus except for very small sample sizes, and large measurement errors, (0.17) and (0.18) are nearly identical. Thus just as for finite inventories, classical confidence intervals for infinite inventories contain the hidden assumption of a uniform prior when interpreted from a Bayesian viewpoint. Finally the Bayesian confidence using a beta prior for $\theta$ takes the form of the ratio of two incomplete beta functions:

$$\alpha = \frac{B_{\theta(1-q)}\left(b,n+1\right)}{B_{(1-q)}\left(b,n+1\right)}. \tag{0.19}$$

For $q = 10\%$ and $\theta = .01$ and $b = .09354$ (65% prior) the sample size increases from 54 to 60.

## Appendix I: Bayesian Confidence

To calculate the Bayesian confidence for discovery sampling, we randomize the parameter *k* in the hypergeometric distribution and find its distribution conditional on the number of defectives in a sample being zero.

$$p\left(X = x \middle| K = k\right) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \qquad \max\left(0, n+k-N\right) \le x \le \min\left(n,k\right) \tag{1.1}$$

Using Bayes’ theorem, we can reverse the conditioning in (1.1):

$$P\left[K = k \middle| X = x\right] = \frac{P\left[X = x \middle| K = k\right] w\left(k\right)}{P\left[X = x\right]}, \tag{1.2}$$

where $w\left(k\right)$ is the probability mass function for the integer-valued random variable *K*. Using the formula for total probability, the denominator in Eq. (1.2) can be written as:

$$P\left[X = x\right] = \sum_{i=0}^{N} P\left[X = x \middle| K = i\right] w\left(i\right) \tag{1.3}$$

Using (1.3) and (1.1) in (1.2) we get:

$$P\left[K = k \middle| X = x\right] = \frac{\binom{k}{x}\binom{N-k}{n-x} w\left(k\right)}{\sum_{i=0}^{N-n} \binom{i}{x}\binom{N-i}{n-x} w\left(i\right)}. \tag{1.4}$$

With $x = 0$ Eq. (1.4) becomes:

$$P\left[K = k \middle| X = 0\right] = \frac{\binom{N-k}{n} w\left(k\right)}{\sum_{i=0}^{N-n} \binom{N-i}{n} w\left(i\right)}. \tag{1.5}$$

Equation (1.5) is the basic formula for computing the Bayesian confidence for discovery sampling given the ancillary or prior information about the inventory as expressed by $w\left(k\right)$.

As an important special case, we assume that $w(k)$ is uniformly distributed. Then $w(k)$ is constant, and (1.5) becomes:

$$P\left[K = k \middle| X = 0\right] = \frac{n+1}{N+1}\frac{\binom{N-k}{n}}{\binom{N}{n}}. \tag{1.6}$$

If we sum the RHS of (1.6) from 0 to *k*, we get the probability that *K* is less or equal to *k*,

$$P\left[K \le k \middle| X = 0\right] = 1 - \frac{N-n-k}{N+1}\frac{\binom{N-k}{n}}{\binom{N}{n}}. \tag{1.7}$$

## Appendix II: Sample Size for Infinite Inventories

From Tables 1–4 we see that as the size of the inventory increases, the sample size required to achieve the sampling design levels off. It turns out that there is little difference in the sample size need to achieve the audit design between large inventories $(N > 10,000)$ and very large inventories $(N > 100,000)$. Mathematically this happens because as $N \to \infty$ with $\theta = k/N$ remaining constant, the hypergeometric distribution asymptotically approaches the binomial:

$$P\left[X = x\right] = \theta^x \left(1-\theta\right)^{n-x} \binom{n}{x} \quad 0 < \theta < 1. \tag{2.1}$$

The parameter $\theta$ is rate of occurrence of defectives in the inventory. The ordinary $\alpha \times 100\%$ confidence interval for $\theta$ is $\left\lfloor 0, \theta^* \right\rfloor$ where $\theta^*$ is the solution of $\left(1-\theta\right)^n = 1-\alpha$. The sample size $n$ needed to achieve meet specified values $\theta$ and $\alpha$ is:

$$n = \left\lceil \frac{\log\left(1-\alpha\right)}{\log\left(1-\theta^*\right)} \right\rceil \tag{2.2}$$

where $"\lceil \quad \rceil"$ denotes the “ceiling” function. For $\theta \times 100 = 1\%$, and $\alpha \times 100 = 95\%$, $n = 299.$ This is close to the sample size for a finite inventory with $N = 10,000$ and $k = 100.$

To get the infinite-inventory version of the Bayesian confidence corresponding to (1.5), let:

$$\begin{aligned} &\frac{K}{N} \to \Theta \qquad\qquad 0 < \Theta < 1 \\ &w\left(k\right) \to w\left(\theta\right) \;\; 0 < \theta < 1 \\ &P\left[K \le k\right] \to P\left[\Theta \le \theta\right] \\ &P\left[\Theta \le \theta\right] = W\left(\theta\right) = \int_0^\theta w\left(t\right) dt \end{aligned} \tag{2.3}$$

Note that $w\left(\theta\right)$ is a continuous probability *density* function whereas $w\left(k\right)$ is a discrete probability *mass* function. The Bayesian confidence is defined as:

$$\alpha = P\left[\Theta \le \theta \middle| X = 0\right] = \int_0^\theta f_{\Theta|X}\left(t \middle| X = 0\right) dt \tag{2.4}$$

which depends on the conditional density function $f_{\Theta|X}\left(\theta \middle| X=0\right)$ for $\Theta$. Using the probability density form of Bayes' law we get:

$$\begin{aligned} f_{\Theta|X}\left(\theta \middle| X=0\right) &= \frac{P\left[X=0 \middle| \Theta=\theta\right] w(\theta)}{\int_0^1 P\left[X=0 \middle| \Theta=t\right] w(t)dt} \\ &= \frac{(1-\theta)^n w(\theta)}{\int_0^1 (1-t)^n w(t)dt} \end{aligned} \tag{2.5}$$

Equation (2.5) is the continuous version of (0.9). If we choose the uniform density for $w(\theta)$, then $w(\theta)\equiv 1$, and the equation for Bayesian confidence becomes:

$$\begin{aligned} & f_{\Theta}\left(\theta \middle| X=0\right) = (n+1)(1-\theta)^n \quad 0<\theta<1 \\ & P\left[\Theta \le \theta \middle| X=0\right] = \int_0^{\theta} (n+1)(1-t)^n dt = 1-(1-\theta)^{n+1} \\ & (1-\theta)^{n+1} = 1-\alpha. \end{aligned} \tag{2.6}$$

The result in (2.6)matches (2.2), and we again see that from a Bayesian viewpoint the ordinary confidence interval contains a hidden assumption: the uniform prior— a very pessimistic assumption in the context of discovery sampling.

For the infinite-inventory version of the discrete beta-binomial pmf (0.6), we use the beta density function:

$$w(\theta) = \frac{(1-\theta)^{a-1}\theta^{b-1}}{B(a,b)} \qquad \operatorname{Re}(a)>0, \quad \operatorname{Re}(b)>0 \tag{2.7}$$

where $B(a,b)$ is the beta function. The beta function can be calculated from the gamma function by the formula:

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{2.8}$$

Using (2.7) with $a=1$ provides a family of tilted distributions useful as priors for discovery sampling: $w(\theta)=b\theta^{b-1}$. Using the beta distribution for $w(\theta)$ in (2.5) gives:

$$f_{\Theta|X}\left(\Theta \le \theta \middle| X=0\right) = b(1-\theta)^n \theta^{b-1}\binom{n+b}{b} \tag{2.9}$$

Finally integrating (2.9) gives the Bayesian confidence:

$$\alpha = P\left[\Theta \le \theta\right] = bB_\theta\left(b, n+1\right)\binom{n+b}{b} \qquad (2.10)$$

where $B_\theta$ is the incomplete beta function of order $\theta$. For a confidence of 95% with $\theta = .01$ and the density shown in Fig. 2, we get $n = 54$.

## Appendix III: Effects of Measurement Error for Finite Inventories

Recall that simple random sampling without replacement is described by the hypergeometric distribution:

$$P\left[X = x\right] = \frac{\binom{k}{x}\binom{N-k}{N-x}}{\binom{N}{n}} \qquad \max\left(0, n+k-N\right) \le x \le \min\left(n,k\right)$$

where *X* is the number of defectives measured in the sample of size *n*. When no defectives occur in the sample, we get:

$$P\left[X = 0\right] = \frac{\binom{N-k}{k}}{\binom{N}{n}} \tag{3.1}$$

When measurement errors are present, we can’t observe *X*. Instead we observe a possibly erroneous version of *X* which we call *Y*. Thus we can’t be sure that when we see *Y* = 0 that *X* = 0 because the sample could contain false negatives. To calculate $P\left[Y = 0\right]$, we have to modify (3.1) to for the effect of measurement errors. To do this we use the method of indicator random variables. An indicator random variable takes the value of 0 or 1. Thus, we can write the sample of size *n* as a set of indicator random variables.

$$\begin{aligned}&\left\{x_1, x_2, \ldots x_n\right\}\\&x_i = 1 \Rightarrow \text{ item } i \text{ is truly defective}\\&x_i = 0 \Rightarrow \text{ item } i \text{ is truly good}\end{aligned} \tag{3.2}$$

The measured outcome is the following:

$$\begin{aligned}&\left\{y_1, y_2, \ldots y_n\right\}\\&y_i = 1 \Rightarrow \text{ item } i \text{ measures as defective}\\&y_i = 0 \Rightarrow \text{ item } i \text{ measures as good}\end{aligned} \tag{3.3}$$

The true and measured counts of the number of defectives in a sample is:

$$\begin{aligned} x &= \sum_{i=1}^{n} x_i \\ y &= \sum_{i=1}^{n} y_i \end{aligned} \qquad (3.4)$$

A false-negative error will convert a $x_i = 1$ into a $y_i = 0.$ Since false positives cannot occur (by assumption), an $x_i = 0$ is never converted into a $y_i = 1.$ Here is an example for $n = 10$:

$$\begin{aligned} \{0,1,0,0,1,0,1,0,1,0\} &\Rightarrow x = 4 \\ \{0,1,0,0,0,0,0,0,1,0\} &\Rightarrow y = 2 \end{aligned} \qquad (3.5)$$

In (3.5) the true number of defectives is 4 while the number of measured defectives is 2 indicating that measurement error caused 2 conversions. Assuming the measurement errors occur independently of position in the sequence, the number of conversions *C*, can be modeled as a Bernoulli random process. The number of successes is the number of conversions, and the number of trials is the true number of defectives *x*. Thus in the example shown in (3.5) the number of trials is 4, and the number of successes is 2. Therefore, we can write the following conditional probabilities:

$$\begin{aligned} &P\left[Y_i = 0 \middle| X_i = 1\right] = q \quad \text{False Negative Probability} \\ &P\left[Y_i = 1 \middle| X_i = 0\right] = 0 \quad \text{False Positives Never Occur} \\ &P\left[C = c \middle| X = x\right] = \binom{x}{c} q^c (1-q)^{x-c} \qquad c = 0,1,2,\ldots x \\ &P\left[Y = y \middle| X = x\right] = P\left[C = x - y \middle| X = x\right] = \binom{x}{x-y} q^{x-y} (1-q)^y \\ &P\left[Y = 0 \middle| X = x\right] = P\left[C = x \middle| X = x\right] = q^x \qquad q > 0 \end{aligned} \qquad (3.6)$$

The last equation says that if the sample really has *x* defectives, then the only way to get a $Y = 0$ is for a total of *x* conversions to have occurred. The probability this happens is $q^x$ — assuming the measurement errors occur independently. Using the formula for total probability, we can write the unconditional probability that $Y = 0$:

$$
\begin{aligned}
P\left[Y=0\right] &= \sum_{x=0}^{\min(n,k)} P\left[Y=0 \middle| X=x\right] P\left[X=x\right] \\
P\left[Y=0\right] &= \sum_{x=0}^{\min(n,k)} q^x \frac{\binom{k}{x}\binom{N-k}{N-x}}{\binom{N}{n}} = P\left[X=0\right]\ {}_2F_1\left(-k,-n,N-n-k+1,q\right)
\end{aligned}
\tag{3.7}
$$

where ${}_2F_1\left(\cdot\right)$ is a special function called the *hypergeometric function*[3]. Equation (3.7) explicitly shows the modification of (3.1) needed to account for measurement errors. For small values of *q*, we have the following series approximation:

$$
{}_2F_1\left(-k,-n,N-n-k+1,q\right) \approx 1 + \frac{knq}{N-n-k+1} \tag{3.8}
$$

so we can see that we are more likely to get $Y = 0$ than $X = 0$. We can now calculate the sample sizes needed for a $100 \times \alpha\%$ confidence interval when measurement errors occur by solving for the minimum value of $k^*$ that satisfies:

$$
\frac{\binom{N-k^*}{n}}{\binom{N}{n}}\ {}_2F_1\left(-k^*,-n,N-n-k^*+1,q\right) \le 1-\alpha, \tag{3.9}
$$

then $\left[0, k^* - 1\right]$ is the confidence interval. A set of illustrative calculations appears in Table 3. The effect of increasing *q* is to increase the sample size necessary to achieve the desired level of confidence. Note that as *q* increases a point is reached where it is impossible to achieve the desired confidence. For example in a 1000-item inventory, if the upper range of the confidence is to be 1% of the size of the inventory then the probability of error cannot exceed 76%. If the size of the inventory is smaller say 100 then the maximum error drops, in the case of a 100-item inventory the error cannot exceed 20%.

Using (3.7) we can compute the Bayesian confidence by:

[3] The hypergeometric function is related to but more general than the hypergeometric distribution.

$$
\begin{aligned}
P\left[K=k\middle|Y=0\right] &= \frac{P\left[Y=0\middle|K=k\right]P\left[K=k\right]}{P\left[Y=0\right]} \\
&= \frac{P\left[X=0\middle|K=k\right]\ {}_2F_1\left(-k,-n,N-n-k+1,q\right)w\left(k\right)}{\sum\limits_{i=0}^{N-n}P\left[X=0\middle|K=i\right]\ {}_2F_1\left(-i,-n,N-n-i+1,q\right)w\left(i\right)} \\
P\left[X=0\middle|K=k\right] &= \frac{\begin{pmatrix} N-k \\ n \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} \\
P\left[K\le k\middle|Y=0\right] &= \frac{\sum\limits_{j=0}^{k}\begin{pmatrix} N-j \\ n \end{pmatrix}\ {}_2F_1\left(-j,-n,N-n-j+1,q\right)w\left(j\right)}{\sum\limits_{i=0}^{N-n}\begin{pmatrix} N-i \\ n \end{pmatrix}\ {}_2F_1\left(-i,-n,N-n-i+1,q\right)w\left(i\right)}
\end{aligned}
\tag{3.10}
$$

The last equation in (3.10) give the Bayesian confidence in terms of the prior distribution and the measurement error.

## Appendix IV: Effects of Measurement Error for Infinite Inventories

Let *X* be the true number of defectives and *Y* be the number of measured defectives in a sample of size *n*. Again, let *q* be the probability of a false negative measurement. Using the same methods as Appendices II and III, we get the following equations:

$$\begin{aligned} &P\left[X=x\right]=\theta^{x}\left(1-\theta\right)^{n-x}\binom{n}{x} \quad 0<\theta<1 \\ &P\left[Y=0\middle|X=0\right]=q^{x} \qquad (4.1) \\ &P\left[Y=0\right]=\sum_{x=0}^{n}q^{x}\theta^{x}\left(1-\theta\right)^{n-x}\binom{n}{x}=\left(1-\left(1-q\right)\theta\right)^{n}. \end{aligned}$$

The sample size required for a classical $\alpha\times100\%$ confidence interval is the smallest integer *n* that satisfies the following inequality:

$$\left(1-\left(1-q\right)\theta\right)^{n}\leq1-\alpha \qquad (4.2)$$

Equation (4.2) is the infinite-inventory version of (3.9).

To get the Bayesian confidence for infinite inventories, let $\theta\rightarrow\Theta$ where $\Theta$ is a random variable with probability density $w(\theta)$. Then (4.1) becomes:

$$P\left[Y=0\middle|\Theta=\theta\right]=\sum_{x=0}^{n}q^{x}\theta^{x}\left(1-\theta\right)^{n-x}\binom{n}{x}=\left(1-\left(1-q\right)\theta\right)^{n}. \qquad (4.3)$$

Applying the density form of Bayes' law as in (2.5), we get:

$$f_{\Theta|Y}\left(\theta\middle|Y=0\right)=\frac{\left(1-\left(1-q\right)\theta\right)^{n}w\left(\theta\right)}{\int_{0}^{1}\left(1-\left(1-q\right)t\right)^{n}w\left(t\right)dt} \qquad (4.4)$$

If $w(\theta)$ is uniform then $w(\theta)\equiv1$, and:

$$f_{\Theta|Y}\left(\theta\middle|Y=0\right)=\frac{\left(n+1\right)\left(1-q\right)\left(1-\left(1-q\right)\theta\right)^{n}}{1-q^{n+1}} \qquad (4.5)$$

Integrating (4.5), we get the infinite-inventory Bayesian confidence for a uniform prior with measurement error:

$$\begin{aligned}\alpha = P\left[\Theta \le \theta \middle| Y = 0\right] &= \int_0^\theta f_{\Theta|Y}\left(t \middle| Y = 0\right)dt \\ &= \int_0^\theta \frac{(n+1)(1-q)\left(1-(1-q)t\right)^n}{1-q^{n+1}}dt \\ &= \frac{1-\left(1-(1-q)\theta\right)^{n+1}}{1-q^{n+1}}\end{aligned} \tag{4.6}$$

Finally using the beta prior (2.7) with $a = 1$ we get the conditional density for $\Theta$:

$$\begin{aligned}f_{\Theta|Y}\left(\theta \middle| Y = 0\right) &= \frac{\left(1-(1-q)\theta\right)^{n+1} b\theta^{b-1}}{\int_0^1 \left(1-(1-q)t\right)^{n+1} bt^{b-1}dt} \\ &= \frac{b\left(1-(1-q)\theta\right)^{n+1}\theta^{b-1}}{{}_2F_1\left(b,-n-1,1+b,1-q\right)}.\end{aligned} \tag{4.7}$$

Note that (4.7) reduces to (2.9) for $q \to 0$.

Integrating (4.7) gives the infinite-inventory Bayesian confidence for a beta prior with measurement error:

$$\alpha = \frac{B_{\theta(1-q)}\left(b,n+1\right)}{B_{(1-q)}\left(b,n+1\right)} \tag{4.8}$$

where $B_z\left(a,b\right)$ is the *incomplete beta function*. Note that for $q \to 0$ the denominator in (4.8) becomes the beta function and the ratio reduces to (2.10).

## Appendix V: Glossary of Variables

The following is a recap of the variables used in the discussion. In general upper case Latin letters denote random variables. Lower case Latin letters can denote either known constants or the realization of a random variable. For example "*X*" denotes a random variable while "*x*" denotes the value that the random variable realizes. So we can talk about the mean of *X* or the probability that $X = 3$ or the probability that $X = x$. This is standard usage in probability and statistics.

- *X*: the number of items that fail inspection in the sample *before* the sample is observed. Thus, *X* is a random variable. However, *X* cannot be observed when measurement error is present.

- *Y*: the number of items that fail inspection in the sample *before* the sample is observed when measurement error is present.

- *x*: the number of items that fail inspection in the sample *after* the sample is observed. In a discovery audit we expect $x = 0$.

- *y*: the number of items that fail inspection in the sample *after* the sample is observed when measurement error is present. If $y = 0$ it is still possible that $x \neq 0$.

- *n*: the number of items in the sample. In general *n* is specified by the design of the discovery audit. We usually want small *n* because the cost of the audit is proportional to *n*.

- *N*: the total number of items in the inventory being audited. Auditors usually know the value of *N* beforehand.

- $\alpha$ : The level of confidence. The customary confidence levels used in statistics are 90% and 95%. The choice is a policy decision. The confidence is the probability that the number of defects in the inventory is less than the tolerance level?

- *k*: an unknown constant that is the total number of defective items in the inventory. The true value of *k* cannot be directly observed without measuring all the items. Since *k* is not a random variable, we cannot make direct probability statements about the value of *k*, but we can make statements about our knowledge of *k* by using confidence intervals.

- *K*:  the total number defective items in inventory expressed as a random variable. Using the random variable framework allows the auditors to assign a prior probability to various values of *K*. We can therefore make probability statements about *K* after we measure the sample. We can also assign confidence probabilities that use the information in the prior distribution for *K*.

- $w(k)$: prior probability mass function for the integer-valued random variable *K*. This is the probability that the number of defective items is exactly equal to *k* for $k = 0,1,\ldots N$. We cannot observe $w(k)$ directly; instead we must rely on ancillary information such as expert opinion. If the auditors feel that all values of *K* are equally likely, then $w(k)$ would be the discrete uniform distribution. In the context of discovery auditing we would expect that $w(k)$ decreases with increasing *k*. If the *n* is small compared to *N*, then $w(k)$ will dominate our knowledge about *K*. Conversely as we increase *n*, the effect of $w(k)$ diminishes. If $n = N$ we sample everything and the choice of $w(k)$ is irrelevant.

- $\theta$: the rate of occurrence of defectives in large inventories. A parameter used in the binomial approximation to the hypergeometric.

- $\Theta$: the rate of occurrence of defectives in large inventories expressed as a bounded continuous random variable.

- $w(\theta)$: probability density function for the random variable $\Theta$.

- $k_{\text{tol}}$: the tolerance level for the number of defects. The greater the consequences of defectives, the smaller the assigned $k_{\text{tol}}$. If $k_{\text{tol}} = 0$ then extremely risk averse "zero-tolerance" policy has been adopted. Zero tolerance leads to absurd sample sizes unless Bayesian confidence methods are used.

- $\theta_{\text{tol}}$: the tolerance level for the rate of defects. Assigned as a policy decision. In discovery audits $\theta_{\text{tol}}$ is typically less than 5%. The greater the consequences of defectives the smaller the tolerance.

- *q*: probability of a "false negative" measurement error. A positive is the measurement of a defect. A negative is the measurement of a non-defective item. A false negative is a measurement that indicates a defective item is good. Terminology is sometimes reversed.

- $P[\cdot]$ the probability of the event contained with the bracket. For example $P[X=0]$ is the probability the number of defective items in the sample is exactly zero.

- pmf: probability mass function. See below.

- $p(\cdot)$: probability mass function. A deterministic function of the variables contained within the parenthesis. For example $p(x,n,N,k)$ can be the hypergeometric distribution. The function $p(\cdot)$ gives a number for every allowed combination of its arguments *x*, *n*, *N*, and *k*.

- $B(a,b)$: beta function. Special two-argument function than can be expressed in terms of the gamma function.

- $B_\theta(a,b)$: incomplete beta function of order $\theta$. Special three-argument function. Must use tables or approximations to compute value. Not expressible in terms of elementary functions.

- ${}_2F_1(a,b,c,q)$: hypergeometric function with 2 upper parameters and 1 lower parameter. The argument is the variable *q*. Special function that requires tables or numerical approximations to evaluate. Not to be confused with the hypergeometric distribution.

- $\Gamma(z)$: gamma function for real argument *z*. When *z* is integer valued, $\Gamma(z+1)=z!$
- $\begin{pmatrix} N \\ n \end{pmatrix}$: binomial coefficient, also called “*N* choose *n*,” the number of ways *n*

distinct objects can be selected from *N* distinct objects without regard to order. The values of *N* and *n* need not be integers as the binomial coefficient can be defined in terms of gamma functions, however the combinatorial interpretation is then lost.