Scale Counting

Stefan H. STEINER and R. Jock MACKAY

Department of Statistics and Actuarial Sciences
University of Waterloo
Waterloo, N2L 3G1 Canada
(*shsteine@uwaterloo.ca*, *rjmackay@uwaterloo.ca*)

In many applications, small parts are counted using a scale. The number of parts is estimated by dividing the total weight by the (estimated) average weight of an individual part. This procedure avoids counting individual parts and can save time and money and improve the accuracy of counts. In this article we explore and quantify the effect of the estimation procedure used to determine the average weight, measurement error, and discretization on the accuracy of the scale count. We present guidelines for the successful implementation of scale counting and suggest a change in the standard procedure.

KEY WORDS: Bulk and dribble counting; Multiple scale counting; Ratio estimation.

1. INTRODUCTION

In many applications, small parts are counted using a weighing method called scale counting (see Fig. 1 for an illustration; and Anonymous 1981 for more background). We count the parts by weighing them and dividing the total weight by the (estimated) average weight of each individual part. This procedure avoids counting individual parts and can save time and money and improve the accuracy of large counts. By accuracy we refer to both bias and variability in the counts.

In this article we show how the accuracy of scale counting depends on

- The variability of the weight of the individual parts,
- The number of parts being counted,
- Calibration, the procedure used to estimate the average weight of individual parts,
- Measurement error, the resolution, bias, and variability of the scale.

As far as we know, the nature of these dependencies has never been extensively studied. The purpose of this article is to provide some guidelines and recommendations for scale counting.

This article was motivated by several visits to an organization that counts and packages automotive parts for shipment overseas. The warehouse from which the shipments originate contains more than 3,000 different components, ranging from engines and transmissions to small fasteners. When we first visited this operation, a shipment consisted of all parts required to build 96 vehicles. At the assembly plant, the plan was to unpack a shipment and build 96 vehicles. In theory, there was no extra inventory or borrowing from future shipments. To execute this plan, it was essential to have the correct number of parts of each type in each shipment. If there were too few of any component, then the planned number of vehicles could not be assembled, whereas if there were too many, then there was waste, confusion, and substitutions in future shipments. In our later visits, the plan had changed. Now parts are shipped in stated volumes, and there is a small inventory of parts at the assembly plant. With this plan, the consequences of small counting errors to the assembly plant is reduced.

In either plan, the shipping organization is responsible for rectifying and explaining all counting errors detected at the assembly plant, often at a cost far greater than the value of the parts in error. In an attempt to reduce these costs, we were asked to audit the scale-counting procedures used throughout the warehouse. We developed the proposed guidelines and recommendations as a result of this audit.

With the current shipping plan, the number of pieces for each part varies from 100 to more than 3,500 (e.g., some common bolts). Low-volume parts have specific target numbers; high-volume items are shipped in bulk, so that the count is important, but there is only a rough target value. We illustrate the implementation of scale counting using two parts. Part *A* is plastic hose, [Fig. 2(a)], for which the goal of the counting is to produce bags of exactly 100 units. Part *B* is a small clamp [Fig. 2(b)] that must be packaged in groups of about 2,000 units.

Scale counting involves three steps. First, the tray (as shown in Fig. 1) is placed on the scale, and the scale is zeroed; this is called taring. Next, in the calibration step, a sample of parts is hand-counted and weighed to determine an estimate of the average part weight. Finally, the operator determines the total weight of a group of parts and determines the count using the estimated average part weight from the calibration sample. The software in the scale calculates and displays the count. There are two versions of the third step. With "bulk" counting, all parts are placed on the tray at one time, and the count is determined. With "dribble counting," parts are added and removed from the scale until the estimated number of parts equaled the desired target value. In the application, bulk counting is used for the clamps and dribble counting is used for the hoses. Although these two procedures have different goals, we show in the Appendix that the statistical properties of bulk and dribble counting are virtually identical. This near equivalence has also been verified using simulation. The remainder of the article addresses bulk counting only, although the results are valid for dribble counting as well.

At the shipping firm, the calibration step is repeated whenever an operator needs to count a particular part due to possible variability in the average part weight between batches, and to avoid record keeping. Note that the calibration sample has size 25 for all parts, which yields a simple calibration procedure.

TECHNOMETRICS, AUGUST 2004, VOL. 46, NO. 3
DOI 10.1198/004017004000000374

Figure 1. A Counting Scale.

The article is organized as follows. In Section 2 we analyze the impact of variability in part weight, the number of pieces being counted, and the calibration procedure on the accuracy of the scale count. We derive these initial results assuming a scale with no measurement error and ignoring the discretization needed to yield an integer count. Next, we quantify the effect of relaxing these assumptions and determine when measurement error and discretization can substantially effect the statistical properties of the count. In Section 3 we propose a new procedure, called multiple scale counting, that yields more accurate counts than the standard procedure in certain circumstances. In Section 4 we turn to issues of importance in the implementation of scale counting. We make recommendations for changes to the standard practice and provide guidelines for dealing with the impact of measurement error and discretization. Finally, we summarize our results in Section 5.

(a)

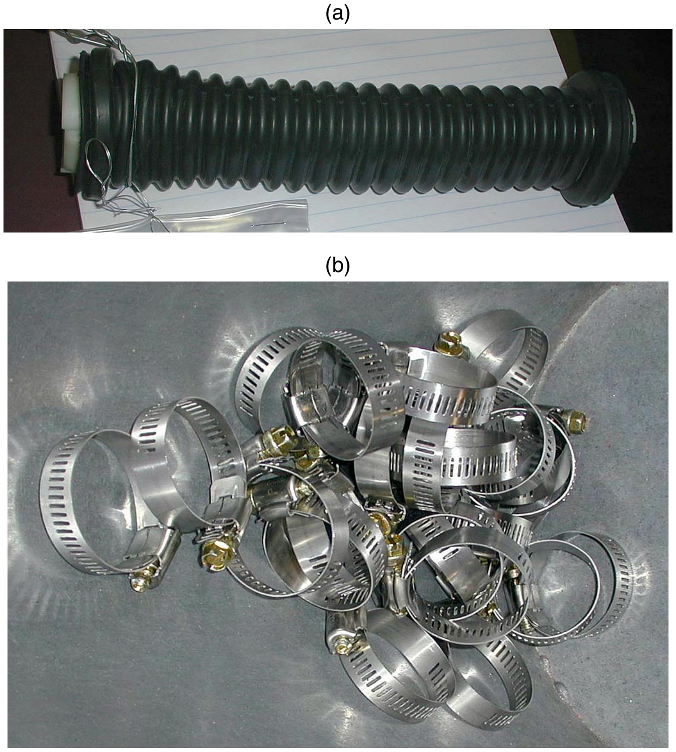

Figure 2. Sample Parts to be Scale Counted (a) Hose, (b) Clamps.

2. STATISTICAL PROPERTIES OF SCALE COUNTING

In the calibration step, we count p parts (usually a grab sample from a bin or basket) and determine their total weight. We then estimate the average part weight μ by

$$\hat{\mu} = \frac{\hat{t}_p}{p}, \tag{1}$$

where $\hat{t}_p$ is the total weight of the p parts subject to sampling error. For the moment, we assume that here are no measurement errors; see Section 2.3. With this calibration procedure, we cannot estimate the standard deviation, σ, of the individual part weights.

In the counting step, we measure the total weight of a group of parts, denoted by $\hat{t}_n$, (again collected as a grab sample) of unknown size n. Then an estimate of the total number of parts is

$$\hat{n} = \frac{\hat{t}_n}{\hat{\mu}}. \tag{2}$$

In practice, the result from (2) is rounded to the nearest integer. In Section 2.2 we show that, except when the variability in the count is very small, the effect of rounding is not important.

We denote estimators with a tilde, so, for example, $\tilde{n}$ is the estimator corresponding to the estimate $\hat{n}$ given by (2). If n and p are large (>100 and 5) and the coefficient of variation $\gamma = \sigma/\mu$ is small ($<.05$), then we can accurately estimate the mean and standard deviation of $\tilde{n}$ using statistical differentials (i.e., Taylor series expansion) retaining terms up to second order (Kotz and Johnson 1982). We obtain

$$E(\tilde{n}) \cong \frac{E(\tilde{t}_n)}{E(\tilde{\mu})}\left[1 + \frac{\mathrm{var}(\tilde{\mu})}{E(\tilde{\mu})^2}\right]$$
$$= n\left(1 + \frac{\gamma^2}{p}\right) = n + \beta \tag{3}$$

and

$$\mathrm{var}(\tilde{n}) \cong \left[\frac{E(\tilde{t}_n)}{E(\tilde{\mu})}\right]^2 \left[\frac{\mathrm{var}(\tilde{t}_n)}{E(\tilde{t}_n)^2} + \frac{\mathrm{var}(\tilde{\mu})}{E(\tilde{\mu})^2}\right]$$
$$= n\gamma^2\left(1 + \frac{n}{p}\right) = \alpha^2. \tag{4}$$

In what follows, we ignore any error in these approximations. Note that the bias β is positive and the ratio $\mathrm{var}(\tilde{n})/\beta = p + n$. If we further assume that part weights are independent and normally distributed for parts selected from the bin, then we can show, using the method proposed by Cabuk and Springer (1990), that the estimator $\tilde{n}$ is approximately normally distributed under conditions where the approximations (3) and (4) are reasonable. The assumption of normality of the part weights is not critical to the results.

As pointed out by some reviewers, bulk counting can be considered in the context of simple random sampling and the estimation of a ratio (Cochran 1977; Raj 1968). For instance, Mendenhall, Ott, and Scheaffer (1971) discussed an example to determine the number of oranges in a shipment using simple random sampling. The results in (3) and (4) differ slightly from the their calculations, because in our case $\tilde{t}_n$ corresponds to a sample from all of the available parts and we also assume that the part weights are independent. Equations (3) and (4) and the normality can also be derived as approximations from the properties of simple random sampling.

It is apparent from (3) and (4) that the effectiveness of scale counting depends on the variability in the individual part weights (i.e., γ), the calibration sample size p, and the unknown group size n. In our application, where $n < 3{,}000$, $\gamma \le .05$, and $p = 25$, the bias is small ($< .3$) and the standard deviation is as large as 30. If we increase p to 50, then the standard deviation is < 21.3. Note that we cannot find the standard error of the estimate without an estimate of γ that is not available from the scale counting procedure.

2.1 Effect of Discretization

In practice, the scale count is reported as round($\hat{n}$), the integer closest to $\hat{n}$. Here we examine the effects of the discretization on the estimator.

We have, approximately, $\text{round}(\tilde{n}) - n = \text{round}(\tilde{n} - n) \sim \text{round}(\text{N}(\beta, \alpha^2))$ so we can numerically determine the bias and standard deviation α_{round} of the discretized estimator for any values of β and α. For our application, the rounding has minimal effect on the bias because $\beta = \alpha^2/(p+n)$ and $p+n > 100$. Because the bias is small relative to the variance, using the symmetry of the normal distribution, we can see that rounding has little effect on the bias. If α is small, then α_{round} can be substantially less than α. Figure 3 plots α_{round} against α for small values of α, setting the bias to 0.

If the standard deviation $\alpha \ge .5$, we can approximate the effect of the discretization by modeling round($\tilde{n}$) as $\tilde{n} + U$, where $\tilde{n}$ and U are independent and U is a continuous uniform random variable that ranges between $-1/2$ and $1/2$, with $E(U) = 0$ and $\text{var}(U) = 1/12$. In this case, $\alpha_{\text{round}} \cong \sqrt{\alpha^2 + 1/12}$ is slightly increased.

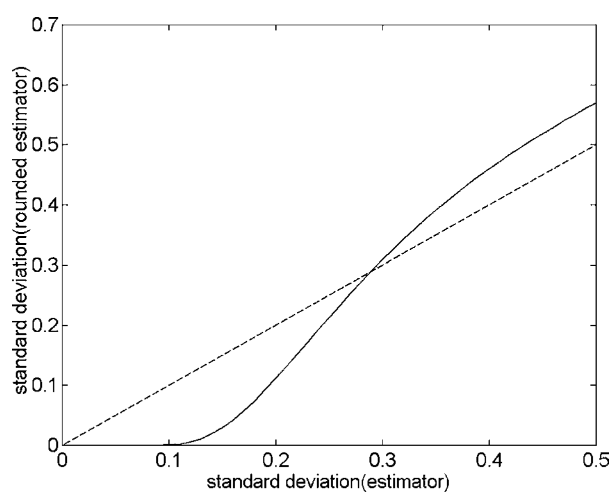

Figure 3. The Effect of Rounding on the Standard Deviation of the Count.

In summary, because the bias is typically small and the standard deviation exceeds .75, the discretization in the scale counting procedure has little effect on the properties of the estimator. However, there are dramatic beneficial effects of discretization when $\alpha < .2$. This property is exploited in Section 3, where we introduce the idea of multiple scale counting.

2.2 Effect of Measurement Error

To derive the expressions (3) and (4), we assumed no measurement error. We now relax this assumption and show that the negative effect of measurement error is usually small.

Suppose that the measurement bias and standard deviation are β_m and σ_m, constant for any repeated weighing of the same group of parts. Applying statistical differentials, we obtain the approximations

$$E(\tilde{n}) \cong \frac{n\mu + \beta_m}{\mu + \beta_m/p}\left(1 + \frac{p\sigma^2 + \sigma_m^2}{(p\mu + \beta_m)^2}\right) \tag{5}$$

and

$$\text{var}(\tilde{n}) \cong \left(\frac{n\mu + \beta_m}{\mu + \beta_m/p}\right)^2 \left(\frac{n\sigma^2 + \sigma_m^2}{(n\mu + \beta_m)^2} + \frac{p\sigma^2 + \sigma_m^2}{(p\mu + \beta_m)^2}\right). \tag{6}$$

We suppose that the scale has been chosen so that β_m/μ and σ_m/σ are relatively small and hence these quantities divided by n or p are negligible. If we ignore negligible terms in (5) and (6), we have that the bias is $E(\tilde{n}) - n \cong \beta + \beta_m/\mu$ and var($\tilde{n}$) is unchanged from (4). The effect of measurement bias can be substantial.

Note that, as in the current application, we can eliminate measurement bias using the taring procedure where the weight of a group of parts is determined as the difference in weight of the tray with the parts and the tray on its own. Using differencing eliminates the measurement bias, and increases σ_m by a factor of $\sqrt{2}$. As shown earlier, this increase has negligible effect on the bias and standard deviation of the estimator. These results depend on the assumption that the properties of the measurement system do not depend on the actual weight of the item or items on the scale.

2.3 Effect of Measurement Resolution

All of the previous analysis has ignored the possible effect of poor measurement resolution. Measurement resolution is defined by the smallest unit of measurement. For example, we may measure weights to the nearest 2 g. Then any group of parts that weighed between 99 and 101 g yields a measured value of 100 g. We quantify the resolution of a measurement device as the inverse of the scale's minimum discrimination weight. Then, ignoring measurement error, a scale with resolution r that weighs a part of weight w yields the result round(rw)/r. For example, if a scale has a capacity of 20 kg with 10,000 divisions, then the scale's minimum discrimination is 2 g, and $r = 1/2$ (when parts are weighed in grams). Similarly, $r = 100$ means that the scale is capable of providing around 2 decimal points.

As before, we can approximate the effect of the rounding due to the measurement resolution by modeling round(X) as $X + U$, where U is a continuous uniform random variable that ranges between $-1/2$ and $1/2$. Assuming no measurement error, we can write

$$\tilde{n} = \frac{\text{round}(\text{N}(rn\mu, nr^2\sigma^2))/r}{\text{round}(\text{N}(rp\mu, pr^2\sigma^2))/pr} \doteq \frac{p\text{N}(n\gamma, n) + pU/r\sigma}{\text{N}(p\gamma, p) + U/r\sigma}. \quad (7)$$

Applying the method of statistical differentials, as in (2) and (3) to the ratio (7), we get

$$E(\tilde{n}) - n \cong \beta + \frac{n}{12p^2r^2\mu^2} = n\left(\frac{\gamma^2}{p} + \frac{1}{12p^2r^2\mu^2}\right) \quad (8)$$

and

$$\begin{aligned} \text{var}(\tilde{n}) &\cong \alpha^2 + \frac{1}{12r^2\mu^2}\left(1 + \frac{n^2}{p^2}\right) \\ &= n\left(\gamma^2 + \frac{1}{12nr^2\mu^2} + \frac{n}{p}\gamma^2 + \frac{n}{12p^2r^2\mu^2}\right). \end{aligned} \quad (9)$$

If r is small, then the bias and the standard deviation can be adversely affected. Because generally $n \gg p$, we know that $\frac{n}{12p^2r^2\mu^2} \gg \frac{1}{12nr^2\mu^2}$ and $\frac{n}{p}\gamma^2 > \gamma^2$. Thus, to bound the effect of the measurement resolution on both the bias and variance of $\tilde{n}$, we compare $\frac{n}{p}\gamma^2$ and $\frac{n}{12p^2r^2\mu^2}$. We can limit the extra bias and variation added by the resolution to $< 50\%$ approximately by ensuring that $\frac{n}{12p^2r^2\mu^2} \le \frac{1}{2}\frac{n}{p}\gamma^2$ or $r \ge \frac{1}{\sigma\sqrt{6p}}$.

In the motivating example, the minimum discrimination of the scales used is 2 g, so the measurement resolution $r = .5$. The calibration step uses p equal to 25 parts. Assuming a coefficient of variation equal to .025, the rule of thumb implies that the average weight of individual parts should be greater than 6.5 g. Otherwise, the effect of the resolution of the scale on the properties of the estimator is (relatively) substantial.

3. MULTIPLE SCALE COUNTING

As shown in Section 2.1, rounding generally increases $\text{var}(\tilde{n})$ by approximately $1/12$. However, when the variance is $<.2$, the discretization can reduce the variance (and bias) dramatically. This large potential reduction in variability suggests an alternative scale counting strategy using multiple scale counts. The idea is simple. To scale count n parts, we divide the n parts into k subgroups of around m parts each, where $km = n$. Then we scale count each subgroup of parts separately and add the results at the end. By choosing m sufficiently small, we can obtain a count that is extremely accurate. In the extreme, we can consider subgroups of size 1. In this case, unless there is large variability in the individual part weights or large measurement error, the multiple scale count will result in the correct count for the total number of items; that is, the bias and standard deviation of the corresponding estimate are 0.

When using multiple scale counts, the estimate of the group size is

$$\hat{n}_{\text{mult}} = \sum_{j=1}^{k} \text{round}\left(\frac{\hat{t}_{m(j)}}{\hat{\mu}}\right), \quad (10)$$

where $\hat{t}_{m(j)}$ is the observed weight of the m items in the jth subgroup, $\hat{\mu}$ is the estimated average part weight obtained from the calibration step, and $mk = n$. Note that $\hat{\mu}$ is the same for all subgroups, because the calibration step is performed only once.

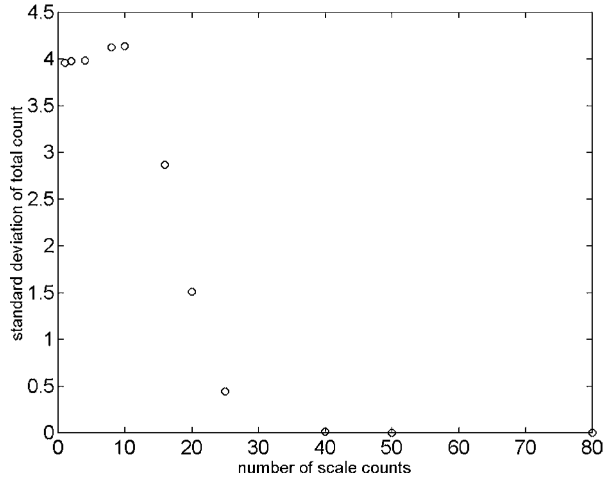

Figure 4. Standard Deviation of $\tilde{n}_{mult}$ for $n = 2{,}000$, $\sigma/\mu = .01$, $p = 25$.

We cannot find a convenient algebraic expression for the mean and variance of the estimator $\tilde{n}_{\text{mult}}$ when m is small enough to produce the dramatic reduction in variance. However, we can numerically calculate the mean and standard deviation by conditioning on $\tilde{\mu}$.

As an example, consider a situation where we wish to scale count 2,000 items, the coefficient of variation of the individual part weights is .01, and we use a calibration sample of size $p = 25$. Figure 4 plots the standard deviation of $\tilde{n}_{\text{mult}}$ as a function of the number of subgroups k. We obtain a substantial reduction in the variation by using as few as 20 scale counts of about 100 items each.

Although the distribution of $\tilde{n}_{\text{mult}}$ is complex, we can derive a simple rule of thumb to determine when multiple scale counting is beneficial. In Figure 3 we see that the discretization substantially reduces the variability in the count whenever $\text{var}(\tilde{n}) < .2^2 = .04$. Now, using the approximation (4) with subgroup size m, we want $m\gamma^2(1 + m/p) < .04$ or, equivalently, $m < \sqrt{\frac{p}{\gamma^2 25} + \frac{p^2}{4}} - \frac{p}{2}$. For example, for $p = 25$ and $\gamma = .01$, the rule of thumb suggests that the subgroup size should be less than around 88 units. This matches closely the results shown in Figure 4. Note that the maximum subgroup size suggested by the foregoing rule of thumb increases roughly linearly in $1/\gamma$. Multiple scale counting is feasible only if the coefficient of variation of the parts is not too small.

One potential problem with multiple scale counting is that the number of subgroups may be miscounted. To avoid this problem, we recommend finishing the multiple scale count with a scale count of all subgroups together. If the combined result is close to the desired value, then we have verified that the correct number of subgroups was used.

4. IMPROVING SCALE COUNTING

The motivation for this paper was an audit of a shipping organization that used scale counting on a routine basis for large number of parts. The goal of the audit was to identify opportunities for improvement to reduce counting errors as determined by the customer. In the earlier discussion, we considered the statistical properties of scale count estimators and their sensitivity to rounding, measurement error, and the resolution of the scale. Here we consider some practical issues for the improvement of scale counting. We use the two parts, A and B, for illustration.

4.1 Estimating the Variability of Individual Part Weights

As shown in Section 2, the statistical properties of the scale count are dependent on the coefficient of variation γ of the individual part weights. Without an estimate for γ, we cannot deal with planning issues, such as the choice of scale and the size of the calibration sample, or assess the precision of a particular count. We get no information about the current value of γ from a single application of the standard scale counting procedure.

We recommend that a small study be conducted periodically on every part to assess γ. We can conduct this study by changing the calibration step of a planned scale count so that each of the p parts is weighed individually. At the same time, the variability of the scale can be assessed by measuring each of the parts twice. There is no need to reestimate γ as often as μ, because the coefficient of variation is not needed to determine a scale counting result.

In our application, investigations, all with 25 parts, to estimate γ for a range of parts (but not all, due to time and cost constraints) were conducted. For parts A and B, we have $\hat{\gamma}_A = .015$ and $\hat{\gamma}_B = .039$. The same scale is used for each part, and the measurement error was negligible. Using (3) and (4), we estimate the bias and standard deviation of $\tilde{n}$ for a count of 100 units of part A as .0009 and .342, and for a count of 2,000 units of part B as .119 and 15.5. For round($\tilde{n}$), using Figure 3, the standard deviation is .380 for part A. The other characteristics are unchanged.

For part A, assuming that γ does not change substantially, we estimate that about 85% of the counts to 100 will be accurate, with a deviation of more than 1 highly unlikely. With the current procedure, for part B we can use the standard error to get an approximate 95% confidence interval (± 30) for a future count of close to 2,000 units. We can reduce this range by half if, using (4), we increase the calibration sample size to $p = 107$.

4.2 Checking the Calibration Sample Size

An error in the hand counting of the calibration sample can lead to a substantial bias in the scale counting procedure. For example, if the calibration sample size was suppose to be 25 but is actually only 24, then using (3), the extra bias in the count is around 4%. A larger calibration sample reduces the standard deviation of the estimate but increases the likelihood of miscounting. In our review, we found that several operators verified the hand count of the calibration sample by adding a specified number of additional parts to the scale and checking that the scale count matched the expected total. We use the foregoing derivations to determine a reasonable number of parts to add.

We denote the number of additional parts by q and the actual number of parts hand counted in the calibration step by p^*. Then, after some algebra, we find that the estimated count has mean $p + \frac{pq}{p^*}(1 + \frac{\gamma^2}{p^*})$ and variance $\frac{qp^2\gamma^2}{(p^*)^2}(1 + \frac{q}{p^*})$ approximately. We want to choose q large enough so that if p^* does not equal p, the resulting scale count will likely not yield $p + q$. Figure 5 shows contours of minimum value of q needed so that 95% of the time the checking procedure would identify a calibration sample whose size differs from the desired number p.

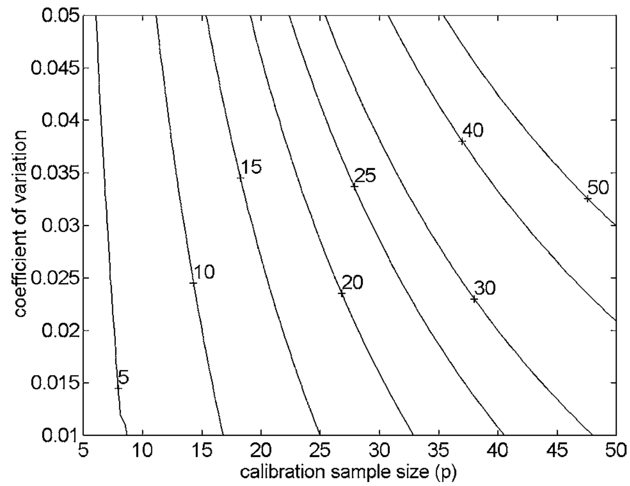

Figure 5. Minimum Additional Size Sample Needed to Check the Hand Count.

In our application, the planned calibration sample size is $p = 25$. To keep the calibration step uniform for all parts and to conserve time, we recommended that operators add 20 parts after weighing the calibration sample and verify that the scale count reads 45. Based on Figure 5, checking the calibration count by adding 20 units is adequate for part A but not quite adequate for part B.

4.3 Asymmetric Loss

The estimated number of parts obtained by scale counting may be less or greater than the actual number of parts on the scale. In some applications, the consequences of making a mistake may depend on whether the actual number of parts is overestimated or underestimated. This is the case in our example of dribble counting, where sending too few parts is a more severe problem than sending too many parts. Similarly, in bulk counting, incorrect estimates of the number of parts in a group may result in errors in the inventory count with overestimates and underestimates have different consequences.

Unequal consequences for over estimation and underestimation suggest an asymmetric loss function. This suggests that when dribble counting, we aim for more parts than are actually needed to make sure at least the desired number of parts is obtained. Similarly, with a bulk count, we may wish to underreport the observed scale count.

The amount by which the target should exceed the desired number of parts depends on the variability and bias in the dribble count, and the level of protection against too few parts desired. As shown in Section 2, the bias and variability depend mostly on the coefficient of variation of the individual part weights, and the sample size used in the calibration step. Suppose, for example, that $p = 25$ and that we aim for 2% over the actual target. Figure 6 explores the effect of the coefficient of variation, and changing either the aimed-for percent over target or the calibration sample size on the chance the dribble counting will yield less than n parts. The probability of too few

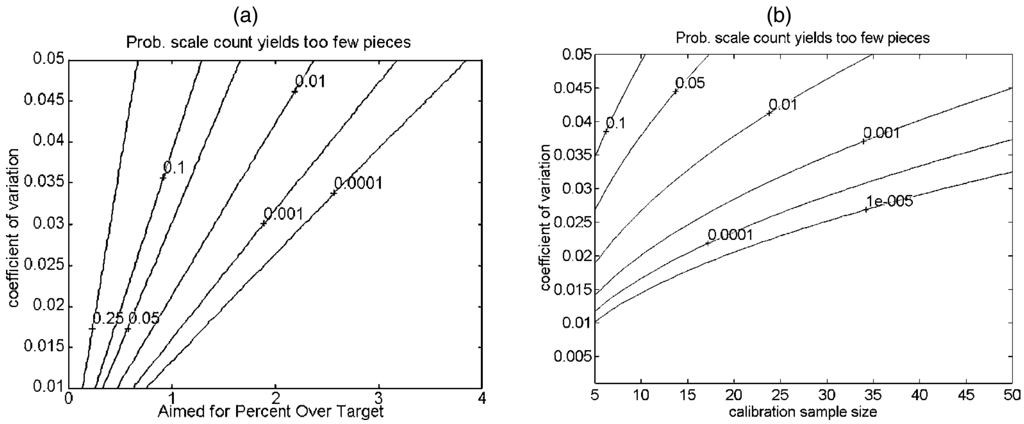

Figure 6. Contours of the Estimated Probability That Scale Counting Yields Too Few Parts (a) p = 25, (b) Aim for 2% Over Target.

parts is determined using (3), (4), and a normal approximation. Figure 6 is based on the assumption that $n = 500$ parts are desired, although the results are not very dependent on the value for n.

In our example, part B is relatively inexpensive, and sending too few parts is undesirable. Based on Figure 6, if we want to have only a 1% risk of sending fewer than 2,000 clamps in a bag, we should aim for roughly an extra 2% of units; that is, we should aim for 2,040 clamps rather than 2,000. For part B, adding extra parts was chosen over other ways of improving the scale count, such as increasing the calibration sample size, to balance cost and complexity.

4.4 Multiple Scale Counting

Using the rule of thumb, multiple scale counting is feasible for part A so long as each group consists of less than roughly 50 units. For part A, combining the counts of two groups of 50 units each and taking into account the effect of discretization reduces the standard deviation of the overall count to .12, resulting in the correct count around 98.5% of the time (up from 85% of the time with one group of 100 units). However, for part B with its large coefficient of variation, multiple scale counting is not practical, because the largest group size for multiple scale counting to be effective is about 20 units, and we wish to count a total of 2,000 units.

In our application, even for parts such as A, we were unable to sell the idea of multiple scale counting because of the increased cost and complexity.

5. DISCUSSION AND RECOMMENDATIONS

The results derived in this article suggest that scale counting can be an effective method of counting parts. However, the accuracy of the count depends critically on a number of factors, including the sample size used to estimate the average part weight, the number of parts that we wish to count, and the coefficient of variation of the individual part weights. The effects of these factors on the estimator were examined in Section 2. Poor measurement resolution and measurement variability can have a substantial negative effect on scale counting. We show, however, that for reasonably good measurement devices, the negative effect of measurement variability and resolution is small. Measurement bias, on the other hand, can have a substantial effect but can be avoided using differencing. Finally, the discretization used to obtain integer estimates of the count usually has little effect but can be beneficial in eliminating the bias and variability in the estimate in certain circumstances.

Choosing an appropriate scale counting procedure is a management decision in which we must balance the required precision of the count with the cost and complexity of the counting procedure. The results of this article provide information to help make an informed choice.

To summarize, in applications where scale counting is economical, we recommend the following steps in planning scale counting:

- Check that the measurement system variability is sufficiently small.
- Estimate the coefficient of variation for the individual part weights.
- Ensure that the measurement resolution is sufficiently large.
- Choose the calibration sample size to ensure the desired precision.
- Aim for more parts than needed if undercounts are more serious than overcounts.
- Consider using multiple scale counting.

When implementing scale counting we recommend the following steps:

- Check the calibration sample hand count (Sec. 3.2).
- Use taring to eliminate the effect of possible measurement bias.

ACKNOWLEDGMENTS

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada. The authors thank Maria Cruz and Jeffrey Chevers for allowing and encouraging studies of scale counting at TDS Automotive. They also thank the editor for insightful comments that improved the article.

APPENDIX: CORRESPONDENCE BETWEEN BULK AND DRIBBLE COUNTING

In this appendix we show that the statistical properties of bulk and dribble counting are virtually identical. Suppose that the target for dribble counting is n and the number of parts actually on the scale is Y. For a bulk count of n units, let X be the corresponding scale count. Consider any sequence of cumulative sums $S_1, S_2, \ldots, S_i$, where $S_i = \tilde{t}_i/\tilde{\mu}$ and each term in the sequence arises by adding one more item to the scale. Because γ is small, we make negligible error assuming that this sequence is increasing. The value of Y is the index i for which $n - 1/2 \le S_i \le n + 1/2$ and thus the scale count reads n. Rarely, there are two possible values of i depending on how dribble counting is implemented—namely, whether the last part is added or subtracted. Suppose, without loss of generality, that $Y = n - k \le n$. Now in the sequence S_n corresponds to the addition of k more units to the scale, and because γ is small, the scale count is likely to increase to $n + k$. This is true in all cases except when S_i is close to a boundary, which generally does not happen in a consistent manner. That is, $X = n + k$ with high probability when $Y = n - k$. Because this holds for all such sequences, it follows that $\Pr(Y = n - k) \doteq \Pr(X = n + k)$, and so the two counting methods have essentially the same properties.

[Received May 2001. Revised January 2004.]

REFERENCES

Anonymous (1981), "Industrial Weighing—Part 1: Electronic Counting Scales," *Modern Materials Handling*, 36, 50–59.

Cabuk, S., and Springer, M. (1990), "Distribution of the Quotient of Noncentral Normal Random Variables," *Communications in Statistics, Part A—Theory and Methods*, 19, 1157–1168.

Cochran, W. G. (1977), *Sampling Techniques* (3rd ed.), New York: Wiley.

Kotz, S., and Johnson, N. L. (eds.) (1982), *Encyclopedia of Statistical Sciences*, New York: Wiley.

Mendenhall, W., Ott, L., and Scheaffer, R. (1971), *Elementary Survey Sampling*, Belmont, CA: Wadsworth.

Raj, D. (1968), *Sampling Theory*, New York: McGraw-Hill.

Wolynetz, M. S. (1979), "Maximum Likelihood Estimation From Confined and Censored Normal Data," *Applied Statistics*, 28, 185–195.