Revisiting activity sampling:
a fresh look at binomial proportion confidence intervals

Stephen M. Disney
Logistics Systems Dynamics Group, Cardiff Business School, Cardiff University.
Email: DisneySM@cardiff.ac.uk. Telephone: +44(0)2920 876310.

Abstract
The Wald interval is typically used to assign confidence to the accuracy of activity sampling studies. It is known the performance of the Wald interval is poor, especially when the observed probability is near zero or one. The suitability of the Wald interval for activity sampling is not often discussed in the operations management literature; if it is, this is usually followed by inappropriate and incorrect advice. Herein a range of alternative binominal confidence intervals for activity sampling is reviewed. A number of selection criteria are considered including achievement of the target nominal coverage probability, size of the interval, and ease of use and presentation. It is recommended that the Clopper-Pearson interval is used for activity sampling. A table of confidence intervals and sample sizes that is specifically designed to be used within a new activity sampling procedure based on the Clopper-Pearson interval is developed. Finally, pedagogical issues are considered.

Key words
Activity sampling, work sampling, binomial proportion confidence intervals, coverage probability.

1. Introduction and motivation

Activity sampling is an empirical data collection technique attributed to Tippet (1935). It can be used to determine the proportion of time: an activity is being conducted by an operator; an operator is doing productive, value adding work; a machine or operator being delayed, or whether an entity (customer, supplier, product) possesses a particular (quality) characteristic or not.

Activity sampling has been defined by British Standard 3138 (1992) as “A technique in which a large number of observations are made over a period of time of one group of machines, processes or workers. Each observation records what is happening at that instant and the percentage of observations recorded for a particular activity or delay is a measure of the percentage of time during which that activity or delay occurs”. Activity sampling has also been known as: work sampling in the U.S.; snap reading, Tippets’ original name for the technique; the ratio delay technique, as it can be used to ascertain the proportion of time a machine or operator is delayed or idle; and in the service sector, random moment studies.

Activity sampling is a relatively inexpensive technique that can be used to determine the proportion of time spent on a particular activity. It can be applied to long, varied and intermittent work. It does not use a stop watch, so it is more acceptable to the subjects of the study and is not as intrusive as other methods. However, activity sampling is not efficient when activities are of a short duration, regular and predictable when time and motion studies might be more suitable.

Almost all Operations Management (OM) textbooks that the recommended activity sampling procedure use the Wald interval to assign confidence to the results obtained
from the study. It is known in the statistical literature that this interval is a rather poor approximation to the true interval. Herein we test this claim and find that, 80% of the time, the Wald interval does not achieve the desired confidence. This is a serious issue for anyone using activity sampling for safety assurance. To address this we review a range of alternative confidence intervals and select one for inclusion in an updated activity sampling procedure. A new procedure is recommended that actually achieves the desired confidence and accuracy levels.

1.1. The standard activity sampling procedure

For simplicity we assume that we are monitoring the types of activities that an operator is undertaking. The standard activity sampling procedure usually involves the following steps. First, a pilot study (an initial look at the situation being studied) is undertaken to ascertain the range of activities an operator undertakes. The second step is to design the sampling tour and data collection forms. The sampling tour specifies when instantaneous observations of the situation being studied are to be taken. It is assumed that the underlying probability of an activity occurring does not change over time (at least within the period of study). It is important to ensure that there is no systematic effect present in the data collected by taking observations after random intervals of time. The study period should be long enough to capture the complete range of activities the operator undertakes but the study can be interrupted if the necessary. It may be also shortened or lengthened as required. The data collection forms are simply a table with rows that note the time a observation is taken and columns that document the activity observed.

In the next step, the main data collection is undertaken. The situation being studied is observed at the prescribed moments of time and the activity being conducted by the operator is recorded in the data collection form as a tick mark. After several dozen or so observations the tick marks can be tallied to obtain an initial estimate of the probability of an activity occurring, \( \hat{p} \). Alternatively we could use judgement or experience to make an initial estimate of \( \hat{p} \). We could even believe that we are being conservative and (incorrectly) use \( \hat{p} = 0.5 \). Note there will be a different \( \hat{p} \) for each of the activities. We will not index them there – it is fairly obvious which activity is being considered as there will be a unique estimate for each of the columns in the data collection form.

Based on this initial estimate, \( \hat{p} \), we can then calculate how many instantaneous observations \( n \) we need to take in order to be 95% confident that the true (real, constant, but unknown) probability of the activity occurring \( p \) is within a certain boundary with

\[
 n = 4 \hat{p} (1 - \hat{p})/L^2 ,
\]

where \( L \) is a tolerance of the form \( \hat{p} \pm L \), within which a desired level of accuracy has to be achieved. Note \( L \) is an absolute distance and not a relative percentage of \( \hat{p} \). \( L \) can be thought of as the confidence interval half-width.

We continue to take the required number of observations, checking periodically with (1) for an updated value of \( n \) (which may have changed due to the new values of \( \hat{p} \), the estimate of the true probability \( p \)). Note that \( n \) will be different for each of the observed activities (as \( \hat{p} \) is likely to be different for each activity). In order to gain a complete picture of the situation, the largest \( n \) will determine the required number of samples to be taken. Observe that the required level of accuracy \( (L) \) is a strategic decision that
influences the trade-off between the efficiency of the procedure and the accuracy obtained.

1.2. Motivation

At first sight the activity sampling procedure seems fine. It has such a long history and is relatively straightforward, so what is the problem? Take a look at (1). What happens when \( L > \hat{p} \) (or \( L > (1 - \hat{p}) \))? A negative probability of \( p \) (or one that is greater than one) is not possible, so there is an issue here. What happens when \( \hat{p} = 0 \) or \( \hat{p} = 1 \)? Equation (1) incorrectly suggests that no samples should be taken. But rarely do operations management (OM) textbooks discuss these issues.

Equation (1) is actually a simplified and rearranged version of the so-called Wald interval for the binomial proportion, Wald and Wolfowitz (1939). A literature review (see, for example, Pires and Amado (2008) for a modern and particularly comprehensive treatment of 20 different binomial proportion confidence intervals) reveals that statistical scholars have serious concerns about the adequacy of the Wald interval. However, this is the interval used in most the OM textbooks.

Table 1 summarises a review of OM books for the terms activity sampling, work sampling, snap reading, ratio delay, random moment studies and confidence intervals. It highlights that in general it can be said that the Wald interval is almost exclusively recommended and guidance on the interval coverage probability is limited to either a “large sample size” or “\( n\hat{p} > 5 \) and \( n(1 - \hat{p}) > 5 \)”, if it is given at all. If there is indeed a problem with the Wald interval, then it means that the confidence that is assigned to the confidence interval cannot be trusted. The purpose of this paper is to investigate this issue.

1.3. Organisation of this paper

Section 2 presents a short review of the literature that exploits activity sampling to demonstrate the relevance and possible use of activity sampling. Section 3 reviews some background theory and defines notation. Section 4 studies the Wald interval. Key performance measures for confidence intervals, suitable for use in an activity sampling procedure, are defined and justified in Section 5. Section 6 studies some alternative confidence intervals from an OM activity sampling viewpoint. Section 7 reflects upon the considered confidence intervals and makes recommendations. Section 8 details a new updated activity sampling procedure that exploits the recommended confidence interval. Section 9 provides pedagogical reflections. Section 10 concludes. A blank example data collection form and the necessary tables required for the updated activity sampling procedure are presented in the appendices.

2. Recent activity sampling studies

It is interesting to quickly review studies that have used activity sampling in order to gain an understanding of the range of problems and issues that the methodology could be applied too. Farrell et al. (2009) determined the unit labour cost of activities in a bank across multiple branches. Tsai (1996) incorporated activity sampling into an Activity Based Costing methodology. Thomas (1991) investigated labour productivity in the nuclear industry. Liou and Borcherding (1986) studied power plant productivity.
<table>
<thead>
<tr>
<th>Recommended sample size</th>
<th>Guidance given</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>None</td>
<td>Adam and Ebert (1992), Barnes (2008), Finch (2008), Krajewski, Ritzman and Malhotra (2013), Slack, Brandon Jones and Johnston (2013), Waters (2002)</td>
</tr>
<tr>
<td>‘100 samples’</td>
<td>None</td>
<td>Naylor (1996), Schonberger and Knod (1988)</td>
</tr>
<tr>
<td>‘perhaps 250 samples’</td>
<td>None</td>
<td>Bicheno and Holweg (2009)</td>
</tr>
<tr>
<td>‘200 samples’</td>
<td>None</td>
<td>Bicheno (2008)</td>
</tr>
<tr>
<td></td>
<td>n = ( \frac{4 \hat{p}(1 - \hat{p})}{L^2} )</td>
<td>Meredith (1992), Rosenkrantz (2009)</td>
</tr>
<tr>
<td></td>
<td>( \left( \frac{z}{\alpha} \right)^2 \hat{p}(1 - \hat{p}) ), where ( z_{\alpha} ) is the standard normal variant that refers to the confidence level required.</td>
<td>Brisley (2001), Chase and Aquilano (1992), Davis and Heineke (2005), Evans et al. (1984), Greasley (2009), Heizer and Render (2014), Jacobs, Chase and Aquilano (2009), Khanna (2015), Lee and Schniederjans (1994), Noori and Radford (1995), Reid and Sanders (2002), Russell and Taylor (2009), Stevenson (2012), Whitmore (1987)</td>
</tr>
<tr>
<td></td>
<td>‘n\hat{p} &gt; 5 and n(1 - \hat{p}) &gt; 5’</td>
<td>Anderson et al. (2007), Silver (1997)</td>
</tr>
<tr>
<td></td>
<td>‘&gt; 30 samples’</td>
<td>Curwin and Slater (1990)</td>
</tr>
<tr>
<td></td>
<td>‘n\hat{p} &gt; 5 and n(1 - \hat{p}) &gt; 5’</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Treatment of activity sampling confidence intervals in OM textbooks

Buchholz et al. (1996) characterised ergonomic hazards in the American highway construction industry. Chen, Peacock and Schlegel (1989) conducted an ergonomic study to assess physical work stress. Construction site productivity was measured by Heinze (1984). Kaming et al. (1997) found that craftsmen in the Indonesian construction industry spent 75% of their time productively and identified five different root causes of productivity problems.

73% of working time was productive in Gunesoglu and Meric’s (2006) study of the garment industry. Rutter (1994) identified the activities undertaken by operators in a pharmaceutical plant. The results were also used to justify the purchase of additional equipment. Kelly (1964) studied executive behaviour in a factory.

Williams, Harris and Turner-Stokes (2009) identified the proportion of time on patient-related care issues (as supposed to other nursing activities) within a UK neuro-rehabilitation setting. Pelletier and Duffield (2003) also considered hospital scenarios. Finkler et al. (1993) compared activity sampling with time-and-motion studies and reflected upon the policy implications of sampling accuracy in the health services industry. Foley (1999) investigated the impact of restraints in nursing homes.
3. Background theory and notation

Statisticians have developed several formulae to determine, with a particular level of confidence, a range of estimated probabilities that will contain the true probability. These formulae are known as binomial proportion confidence intervals or just confidence intervals for brevity. These confidence intervals are based on the binomial distribution as we are concerned with the observing $x$ number of successes (the activity occurring or not) in $n$ observations. Clopper and Pearson (1934) developed an exact solution and there is also some approximate confidence intervals available that possess various properties.

Let $\hat{p}$ be the observed value of the probability of a particular activity occurring. This observed value is not the true probability which can only be obtained by taking an infinite number of observations. However, it is an appropriate, approximate value for the true probability $p$. As the number of observations increases, the more confident we are that the observed probability is representative of the true probability. However, while $\hat{p} \to p$ as $n \to \infty$, $\hat{p}$ does not approach $p$ asymptotically (Brown, Cai and DasGupta, 2001).

The observed probability $\hat{p} = x/n$, where $x$ is the number of successes in the $n$ samples (the number of times an activity occurs in the $n$ observations). The confidence interval equation will give us an upper, $\hat{p}_U$ and lower, $\hat{p}_L$ bound for the unknown $p$, for a desired level of confidence, $1 - \alpha$. In other words, the confidence interval is a range of values, which we can be sure that, $(1 - \alpha) \times 100\%$ of the time, will include the true value of $p$. Thus if $\alpha = 0.05$, we can be 95% confident that $\hat{p}_L \leq p \leq \hat{p}_U$. The coverage probability, $\Lambda$, is the actual confidence achieved by interval, and $\Lambda_i$ is the actual length of the interval.

In summary,

- $p$ is the (unknown) true value of the probability of a particular activity occurring from the entire population, $0 \leq p \in \mathbb{R} \leq 1$
- $n$ is the number of observations that have been taken
- $x$ is the number of times a particular activity was observed in the $n$ observations
- $\hat{p} = x/n$ is an estimate of the true probability $p$
- $\hat{p}_U$ is the upper limit of the confidence interval
- $\hat{p}_L$ is the lower limit of the confidence interval
- $1 - \alpha$ is the desired confidence level. $\alpha$ is the ‘confidence coefficient’
- $L$ is the interval half-width. It is the difference between the estimated value of $\hat{p}$ and the unconstrained (upper and lower) confidence interval, $\hat{p}_U - \hat{p} = \hat{p} - \hat{p}_L = L$. Note that when $\hat{p}$ is near 0 or 1, the limits of the confidence interval must be truncated to ensure $0 \leq \{\hat{p}_U, \hat{p}_L\} \leq 1$. In which case $L = \max[\hat{p}_U - \hat{p}, \hat{p} - \hat{p}_L]$.
- $\Lambda$ is the coverage probability. The coverage probability is the level of confidence actually achieved (not the desired confidence level).
- $\Lambda_i$ is the actual length of the confidence interval. $\Lambda_i = \hat{p}_U - \hat{p}_L$. Note, that $\Lambda_i < 2L$ near $p = \{0,1\}$.
- $z_\alpha$ is the standard normal variant for a given $\alpha$. 

---
4. Wald interval & measures of performance for activity sampling

The simplest equation for the confidence interval is a simplified and rearranged version of the Ward interval, see Table 1. It is based on a normal approximation to the binomial distribution. The upper and lower limits of the Wald interval are defined as

\[
p_U = \min \left( \hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, 1 \right), \quad p_L = \max \left( 0, \hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)
\]

(2)

where, \( \hat{p} = x/n \) and \( z_\alpha = \Phi^{-1}(1-\frac{\alpha}{2}) \) is the \( 1-\frac{\alpha}{2} \) quartile of the cumulative density function of the standard normal distribution. \( z_\alpha \) for popular confidence levels are;

- 90% confidence, \( \alpha = 0.1 \) and \( z_\alpha = \Phi^{-1}(1-\frac{\alpha}{2}) = \Phi^{-1}(0.95) = 1.64 \),
- 95% confidence, \( \alpha = 0.05 \) and \( z_\alpha = \Phi^{-1}(1-\frac{\alpha}{2}) = \Phi^{-1}(0.975) = 1.96 \),
- 99% confidence, \( \alpha = 0.01 \) and \( z_\alpha = \Phi^{-1}(1-\frac{\alpha}{2}) = \Phi^{-1}(0.995) = 2.58 \).

Notice some boundary conditions have been introduced into (2) to ensure that \( 0 \leq \hat{p} \leq 1 \).

For 95% confidence, rounding \( z_\alpha = 1.96 \) to 2, ignoring the boundary conditions and recognising that \( \hat{p}_U = \hat{p} + L \) and \( \hat{p}_L = \hat{p} - L \), it is easy to see that one of the standard approaches for determining the confidence interval in OM texts is based on the Wald interval,

\[
L = 2\sqrt{\hat{p}(1-\hat{p})/n} \Rightarrow n = 4\hat{p}(1-\hat{p})/L^2 .
\]

(3)

If we consider the same procedure without rounding \( z \) then the other popular OM textbook sample size requirement formulae results from the Wald interval as follows,

\[
L = z_\alpha \sqrt{\hat{p}(1-\hat{p})/n} \Rightarrow n = z_\alpha^2 \hat{p}(1-\hat{p})/L^2 .
\]

(4)

It is often stated that the Wald interval is a conservative estimate but it performs badly when \( n \) is small or when \( p \) is near to zero or one, Blyth and Still (1983). A simple rule of thumb frequently relied upon (see for example Johnson, Kemp and Kotz (2005) and Table 1) is that the Wald interval should only be used when \( np > 5 \) and \( n(1-p) > 5 \). However, analysis suggests even this guidance is questionable. Brown, Cai and DasGupta (2001) also point out that the Wald interval can perform badly for all \( p \) and all \( n \).

4.1. Coverage probability for the Wald interval

The confidence level actually achieved by a certain interval is called the coverage probability. As the true coverage probability is determined by the discrete binomial distribution, the coverage probability can never exactly equal the desired confidence level for all values of \( p \). However, if an interval performs properly, then the coverage probability is always greater than the confidence level desired. If this is the case then we say the interval is exact. In order to obtain the coverage probability, \( \Lambda \) the following formula can be used,
\[ \Lambda = 1 - \sum_{x=0}^{n} \left( \text{If } \hat{p}_L < p, 0, \binom{n}{x} p^x (1-p)^{n-x} \right) + \text{If } \hat{p}_U > p, 0, \binom{n}{x} p^x (1-p)^{n-x} \right), \]  

where \( \text{If } [c, t, f] \) is the conditional statement, “if \( c \) holds then \( t \), otherwise \( f \)”. This expression can be found by deduction (see infra, the Clopper-Pearson interval) and verified via a Monte Carlo simulation. An alternative to this equation using an indicator function can be found in Pires and Amado (2008). Making (5) specific for the case for \( \{n = 50, \alpha = 0.05 \} \) when the Wald interval (2) is used to generate the upper and lower limits of the confidence interval produces Figure 1.

\[ \text{Figure 1. Coverage probability of the Wald interval, } (n = 50, \alpha = 0.05) \]

In Figure 1 there are 1001 points (at 0, 0.001, 0.002, … 0.999, 1) of the unknown, but constant value of the binomial parameter, \( p \). 907 of these 1001 tests fail to meet the 95% confidence limit (at the extremities, (7) is used in these results). Note that these results are based on using \( z_{\alpha} = 1.96 \). If \( z_{\alpha} = 2 \) is used, then the chance of meeting the desired confidence level will be slightly higher, with 799 of the 1001 tests failing to meet the desired 95% confidence interval (note \( z_{\alpha} = 2 \) implies 95.47% confidence). If we follow the advice of excluding \( np > 5 \) and \( n(1-\hat{p}) > 5 \) then 707 of the 1001 tests fail with \( z = 1.96 \) (if \( z_{\alpha} = 2 \) is used, this reduces to 601 failures). For 90% desired confidence, there are 801 failures in the 1001 tests. For 99% confidence there are 999 failures in the 1001 tests.

4.2. Length of the Wald interval

It is also interesting to investigate the length of the confidence interval. Figure 2 illustrates the upper and lower bound of the 95% Wald interval. It also highlights the length of the interval given by

\[ \Lambda_t = \hat{p}_U - \hat{p}_L. \]
Note that for the Wald interval when \( L < \hat{p} < (1 - L) \) then \( \Lambda_i = 2L \). However, when \( \hat{p} > (1 - L) \) or \( \hat{p} < L \), then \( \Lambda_i < 2L \). This is a result of the boundary conditions in (2). It can be seen in Figure 2 that the interval narrows as more samples are taken. Here we have illustrated the case of \( n = 50, 100 \) and \( 250 \). Furthermore, the confidence interval is largest at \( \hat{p} = 0.5 \) and has zero length at \( \hat{p} = 0 \) and \( \hat{p} = 1 \). This is incorrect as it is known

\[
\hat{p}_L = \begin{cases} 
0 & \text{if } x = 0 \\
(\alpha / 2)^{1/n} & \text{if } x = n \\
1 & \text{if } x = n 
\end{cases} \quad \text{and} \quad \hat{p}_U = \begin{cases} 
1 - (\alpha / 2)^{1/n} & \text{if } x = 0 \\
1 & \text{if } x = n 
\end{cases}
\]

(7)

should be used at the extremities, Pires and Amano (2008). These are true the Clopper-Pearson limits.

Figure 2 also contains a contour plot of the number of samples required to ensure 95% confidence (according to (4)) as a function of the underlying binomial probability \( p \), and the interval half-width, \( L \). It can be seen that the Wald interval (incorrectly) advises, for a given interval half width, that the maximum number of samples required occurs when \( p = 0.5 \). The Wald interval also assumes (again incorrectly) that all the contours originate from \( L = 0, p = \{0, 1\} \).

5. Evaluating the performance of a confidence interval for activity sampling

The review of the Wald interval has allowed us to introduce the terminology and the issues involved in binomial proportion confidence intervals. There are many such intervals in the literature (see Pires and Amado, 2008). In order to select a confidence interval for professional OM activity sampling we need some criteria to judge the field.

First, and most importantly, the interval should actually achieve the desired confidence interval. That is, the coverage probability should be greater than the desired confidence level. Second, the length of the interval should be as small as possible, which probably means that excessive confidence should be avoided. Third, the confidence interval should work for different confidence levels. Forth, it should be easy to present and understand in a classroom setting.
We should also not need to make any *a priori* assumptions of the binomial probability $p$. Whilst there are good performing conservative intervals (see Sterne (1954) for example) that use *a priori* information, they are not suitable for activity sampling as the required information is rarely available in a usable form. The procedure to determine the confidence interval is also rather complicated as the solution has no explicit form. Similar arguments were made by Clopper and Pearson (1934), although Buck and Tanchoco (1974) and Buck, Askin and Tanchoco (1983) have developed an activity sampling procedure that does use *a priori* information.

There are also procedures available to modify confidence interval guidance. For example Wang (2007) and Wang (2009) identifies the minimum coverage probability for a given sample and size and interval specification. He also provides a mechanism to determine the average coverage probability. Having determined the exact coverage probability (or the exact average coverage probability) one can then adjust (in an iterative approach) the safety factor to achieve target confidence levels, see Agresti and Caffo (2000). We have not pursued this approach either as this is a rather complex task. Rather we prefer to have a one-step calculation to the confidence interval calculation.

6. Alternative confidence intervals

This section reviews four other binomial confidence intervals. Three of these intervals are approximate, one is exact. These four intervals where selected from Pires and Amado (2008) as being described to possess (or nearly possess) the characteristics outlined in Section 5.

6.1. Agresti and Coull’s ‘Adjusted Wald’ interval

The first alternative confidence interval we will consider is based on a modification to the Wald interval that was introduced by Agresti and Coull (1998), sometimes called the Adjusted Wald interval. The expression for the boundaries of the interval with arbitrary confidence levels is given by,

$$\hat{p}_u = \min \left[\zeta + z_\alpha \sqrt{\frac{\zeta(1-\zeta)}{n+z_\alpha^2}}, 1\right] \text{ and } \hat{p}_l = \max \left[\zeta - z_\alpha \sqrt{\frac{\zeta(1-\zeta)}{n+z_\alpha^2}}, 0\right],$$

where $\zeta = \left(x + z_\alpha^2/2\right) \left(n + z_\alpha^2\right)^{-1}$. Using (8) in (5) allows us to determine its coverage probability, see Figure 3.
Figure 3. Coverage probability of the Agresti and Coull interval, \((n = 50, \alpha = 0.05)\)

For the record, 232 of 1000 tests fail to meet the 99% confidence target, 217 fail to meet the 95% target (above) and 305 fail the 90% target. While this is an improvement to the coverage compared to the Wald interval, 20% to 30% of time, the desired level of confidence is not actually achieved.

Figure 4 highlights the length of the Agresti and Coull interval for 95% confidence. Of note here is the fact that at the extremities of the probability, \(\hat{p} = \{0,1\}\), the interval has a finite length. Also, the maximum and minimum operators in (8) are activated near the extremities. The sample size contours also exhibit more natural behaviour as they do not all originate from the same point in the \((L, p)\) plane. Note that in the contour plot, the untruncated behaviour (for \(L\)) was considered, as it was for Figure 2.

Figure 4. Agresti and Coull interval, its length and sample size requirements for 95% confidence
6.1.1. Agresti and Coull’s ‘Add 4’ interval

If you round the normal variant for 95% confidence from \( z_\alpha = 1.96 \) to \( z_\alpha = 2 \) and add two successes and two failures to the sample population the Wald interval becomes Agresti and Coull’s Add 4 interval. Starting with

\[
\hat{p}_{4} = \frac{x + 2}{n + 4},
\]

the upper and lower intervals can be calculated with

\[
\hat{p}_U = \min \left[ \hat{p}_{4} + 2 \sqrt{\frac{\hat{p}_{4}(1 - \hat{p}_{4})}{n + 4}}, 1 \right] \quad \text{and} \quad \hat{p}_L = \max \left[ \hat{p}_{4} - 2 \sqrt{\frac{\hat{p}_{4}(1 - \hat{p}_{4})}{n + 4}}, 0 \right].
\]

Now only 76 of the 1000 tests fail to reach 95% coverage. Equation (10) is interesting as it shows that if you do not collect any samples i.e. \( n = x = 0 \), then you can be confident that \( 0 \leq p \leq 1 \), which is at least a logical result. It is possible to manipulate the un-truncated intervals in (10) to find a concise expression for the number of observations required to ensure \( p \) is within \( \hat{p} \pm L \),

\[
n = 4 \left( \hat{p}_{4} \left( 1 - \hat{p}_{4} \right) - L^2 \right) / L^2.
\]

\[\text{Figure 5. Coverage probability of the Add 4 interval, } (n = 50, \alpha = 0.05)\]

Equation (12) adapts the Add 4 interval for arbitrary confidence levels. It produces 345 failures in the 1001 tests at the 90% confidence level, 211 failures at 95% and 351 failures at 99%.

\[
\hat{p}_U = \hat{p}_{4} + z_\alpha \sqrt{\hat{p}_{4}(1 - \hat{p}_{4})(n + 4)^{-1}} \quad \text{and} \quad \hat{p}_L = \hat{p}_{4} - z_\alpha \sqrt{\hat{p}_{4}(1 - \hat{p}_{4})(n + 4)^{-1}}
\]

In summary, the Add 4 interval has a relatively simple closed form that is only slightly more complex than the Wald interval. It can be manipulated for \( n \), the number of samples needed to be collected in order to reach a defined tolerance band. Although the coverage achieved is vastly better than the Wald interval, especially for 95% when \( z = 2 \) is used, it is still not exact. Agresti and Caffo (2000) report that the Add 4 interval is received well
in a classroom setting, especially when students realise the implications of the Wald interval when \( p = \{0, 1\} \).

### 6.2. Wilson Score interval

The Wilson Score interval is often quoted to be the statistician’s preferred choice for an approximation to the exact interval. The continuity corrected version of the Wilson score interval is recommended by Pires and Amado (2008) and is given by,

\[
\hat{p}_U = \begin{cases} 
1 & \text{if } x = n \\
\frac{2x + z_{\alpha}^2 + 1 + z_{\alpha} \sqrt{z_{\alpha}^2 + 2 - \frac{1}{n} + 4x(1 - \hat{p} - \frac{1}{n})}}{2(n + z_{\alpha}^2)} & \text{if } 0 < x < n \\
\frac{2x + z_{\alpha}^2 - 1 - z_{\alpha} \sqrt{z_{\alpha}^2 - 2 + \frac{1}{n} + 4x(1 - \hat{p} + \frac{1}{n})}}{2(n + z_{\alpha}^2)} & \text{if } x = 0, n
\end{cases}
\] (13)

and

\[
\hat{p}_L = \begin{cases} 
0 & \text{if } x = 0 \\
\frac{2x + z_{\alpha}^2 + 1 + z_{\alpha} \sqrt{z_{\alpha}^2 + 2 - \frac{1}{n} + 4x(1 - \hat{p} - \frac{1}{n})}}{2(n + z_{\alpha}^2)} & \text{if } 0 < x < n \\
\frac{2x + z_{\alpha}^2 - 1 - z_{\alpha} \sqrt{z_{\alpha}^2 - 2 + \frac{1}{n} + 4x(1 - \hat{p} + \frac{1}{n})}}{2(n + z_{\alpha}^2)} & \text{if } x = 0, n
\end{cases}
\] (14)

As we can see in Figure 6, the coverage probability for the 90% and 95% confidence levels is achieved. However, for the 99% confidence level the coverage probability is not met in 36 instances of the 1001 tests.

![Figure 6. Coverage probability for the Wilson “Score” interval (n=50)](image)

Figure 6 highlights the length of the Wilson Score interval for different sample sizes. It also shows that it deals with the extremities appropriately. It is possible to subtract (13) from (14), set it equal to twice the interval half length \((2L)\) and solve for \(n\). It results in a rather unwieldy solution to a fourth order equation. However, it is easily plotted, see Figure 7. By closely comparing the sample size requirements the Wilson interval with the Agresti and Coull interval sample size requirements, we can see that the Wilson Score interval always requires more samples to be taken for a given probability and a given interval half width. This is perhaps the reason why the Wilson Score interval has such a high coverage probability.
6.3. The Arc Sin interval
The final approximate binomial confidence interval considered is the continuity corrected Arc Sin interval discussed in Pires and Amado (2008). This is an interval based on the approximate normal distribution interval after a variance stabilizing transformation. The Arc Sin interval is given by

\[
\hat{p}_U = \begin{cases} 
1 & \text{if } x = n \\
\sin^2 \left( \arcsin \frac{x + \frac{x}{2}}{n + \frac{1}{2}} + \frac{z_{\alpha/2}}{2\sqrt{n + \frac{1}{2}}} \right) & \text{else}
\end{cases}
\]

and

\[
\hat{p}_L = \begin{cases} 
0 & \text{if } x = 0 \\
\sin^2 \left( \arcsin \frac{x - \frac{1}{2}}{n + \frac{1}{2}} - \frac{z_{\alpha/2}}{2\sqrt{n + \frac{1}{2}}} \right) & \text{else}
\end{cases}
\]

The coverage probability of the Arc Sin interval when \( n = 50 \) is highlighted in Figure 8.
The Arc Sin interval actually achieves 90% and 95% coverage, but fails to meet 6 of the 1001 tests at 99% coverage probability. This is an improvement upon the Wilson Score interval, especially as 4 of the 6 failures at 99% coverage occur very close to the extreme values of the binomial probability. However, the interval is very large, see Figure 9.

Figure 9. The 95% Arc Sin confidence interval, its length and sample size requirements

6.4. The Clopper-Pearson interval

An exact solution of the confidence interval is one that always reaches the desired coverage probability. The original approach to this solution is based on finding the real root within the valid probability range of a polynomial of an order equal to \( n+1 \), Clopper and Pearson (1934). Specifically, the upper boundary of the Clopper-Pearson confidence interval is given by the real solution in the range \([0..1]\) to

\[
\sum_{k=0}^{\alpha/2} \binom{n}{k} \hat{p}_U^k (1 - \hat{p}_U)^{n-k} = \frac{\alpha}{2}. \tag{17}
\]

Similarly the lower boundary is given by the real solution, in the range \([0..1]\), to

\[
\sum_{k=x}^{n} \binom{n}{k} \hat{p}_L^k (1 - \hat{p}_L)^{n-k} = \frac{\alpha}{2}. \tag{18}
\]

This form of the Clopper-Pearson interval is rather hard to deal with when \( n \) becomes large. However, the Clopper-Pearson interval can also be expressed in a more manageable form that uses the Beta distribution (see Newcombe (1998) and Pires and Amado (2008)) as follows,

\[
\hat{p}_U = \begin{cases} 
1 - \left(\frac{\xi}{2}\right)^{1/n} & \text{if } x = 0 \\
1 & \text{if } x = n \\
B^{-1}\left[1 - \frac{\xi}{2}, x+1, n-x\right] & \text{if } 0 < x < n 
\end{cases}
\]  
and  
\[
\hat{p}_L = \begin{cases} 
0 & \text{if } x = 0 \\
\left(\frac{\xi}{2}\right)^{1/n} & \text{if } x = n \\
B^{-1}\left[\frac{\xi}{2}, x, n-x+1\right] & \text{if } 0 < x < n 
\end{cases}
\]  

where \( B^{-1}\left[\gamma, \theta_1, \theta_2\right] \) is the \( \gamma \) percentile of the Beta[\( \theta_1, \theta_2 \)] distribution. This form of the Clopper-Pearson interval is easy to handle with modern statistical and mathematical software. They are also computable in Microsoft Excel with the following expressions;
\[ \hat{p}_U = \text{BETAINV}(1 - \alpha/2, x + 1, n - x) \]  
and

\[ \hat{p}_L = \text{BETAINV}(\alpha/2, x, n - x + 1). \]

Johnston, Kemp and Kotz (2005) provide a comprehensive list of references to tables of confidence intervals. They also note the link between the Clopper-Pearson interval and the F distribution and provide the following guidance for the upper limit of the confidence interval,

\[ \hat{p}_U = \frac{v_1 F_{v_1, v_2, 1 - \alpha/2}}{v_2 + v_1 F_{v_1, v_2, 1 - \alpha/2}}, \]

where \( v_1 = 2(x + 1) \) and \( v_2 = 2(n - x) \). The lower limit of the interval is

\[ \hat{p}_L = \frac{v_1 F_{v_1, v_2, \alpha/2}}{v_2 + v_1 F_{v_1, v_2, \alpha/2}}, \]

with \( v_1 = 2x \) and \( v_2 = 2(n - x + 1) \). Using (19) in (5) allows us to inspect the coverage probability, see Figure 10.

\[ \text{Figure 10. Coverage probability of the Clopper-Pearson interval, } (n = 50, \ \alpha = 0.05) \]

Figure 10 confirms that for the Clopper-Pearson interval when \( n = 50 \) and \( \alpha = 0.05 \) the coverage probability is always greater than the desired confidence level. However, because of the discrete nature of the binomial distribution it can often be rather conservative, especially in the extremities of the binomial probability. The confidence interval and its length is portrayed in Figure 11. Here we can see that the interval is largest near \( \hat{p} = 0.5 \), is symmetrical about \( \hat{p} = 0.5 \), and has a finite length at \( \hat{p} = \{0, 1\} \).
The inverse problem where we solve the Clopper-Pearson equations for \( n \), was studied by Johnston, Kemp and Kotz (2005). They highlight that (24) and (25) can be solved to yield an upper and lower limit on \( n \), \( n_U \) and \( n_L \), such that \( p \) is within \( \hat{p} \pm L \) with \((1 - \alpha)\times100\%\) confidence;

\[
\sum_{k=0}^{n_U} \binom{n_U}{k} (\hat{p} + L)^k (1 - \hat{p} - L)^{n_U - k} = \frac{\alpha}{2},
\]  

(24)

\[
\sum_{k=n_L}^{n} \binom{n_L}{k} (\hat{p} - L)^k (1 - \hat{p} + L)^{n_L - k} = \frac{\alpha}{2}.
\]  

Solving (24) and (25) plotting \( n_U \) and \( n_L \) for different values of \( p \) when the interval half length, \( L = 0.1 \) and \( \alpha = 0.05 \) yields Figure 12. Here we can see the two curves for \( n_U \) and \( n_L \). Obviously we will need to pick the largest \( n \) for a particular observed probability, \( \hat{p} \). Hence, the parts of the curves that are plotted in grey become redundant. We can also see that solutions \( n_L < L \) and \( n_U > 1 - L \) do not exist as \( \hat{p} \) cannot be less than zero or greater than unity. Interestingly, the maximum number of observations required does not occur when \( \hat{p} = 0.5 \) (which was advocated by all of the previously considered intervals - see the solid line in Figure 12 for the Wald interval). Rather the maximum observations occur near \( \hat{p} \approx \frac{1}{2} \pm L \) (but not precisely at \( \hat{p} = \frac{1}{2} \pm L \), due to the discrete nature of the binomial distribution). The nature of the two solutions to (24) explains why Figure 11 does not have contours that are maximal in \( L \) at \( p = 0.5 \). Figure 12 also highlights that the Wald interval never takes enough observations to ensure that both sides of the interval have less than \( \alpha/2 \) error probability (error probability = 1 - \( L \)).
Defining \( L^+ \) as the maximum \( L \) in the level sets of the contour plot in Figure 11 allows us to obtain an upper bound of the number of samples required to achieve a desired interval half-width. A visualisation of the relationship between \( L^+ \) and \( n \) is given in Figure 13. Here we can see that the maximum interval half-width reduces as more samples are taken, but there is a law of diminishing returns with large samples sizes.

7. Recommendation of a confidence interval for activity sampling

Table 2 summaries the test results from Sections 4 and 6. If you require the most simple interval possible for 95% confidence, then you could (continue to) use the simplified Wald interval. This does not guarantee the coverage required (about 80% of the time) and assumes the user lacks the capability to exploit a more sophisticated approach. A literature review demonstrated that this approach is almost exclusively advocated by the OM field.
A more meaningful interval for 95% confidence is the Add 4 interval. This greatly improves the coverage probability of the Wald interval and also has a simple closed form solution for the number of samples required to be within a desired tolerance.

To be sure that the coverage probability was achieved (at least for 90% and 95%, and almost sure for 99% confidence) in situations where a closed form expression is required, the continuity corrected Wilson Score interval is recommended. Whilst it is possible to manipulate the interval for n, given a target interval length, it is rather complex for classroom use.

If one is willing to drop the requirement for a closed form, the Arc Sin interval ensures the coverage probability is assured for both 90% and 95% confidence. However, for 99% confidence the interval does not achieve the required coverage at the extremities (and for a small number of points in the middle). It is possible to find the sin and arcsin functions on most good handheld calculators, hence it may be relevant for practising industrial engineers.

Finally, if you must ensure the coverage probability is met and have access to the Clopper-Pearson tables given in Appendix B (or access to the inverse beta distribution in Microsoft Excel) then the Clopper-Pearson interval is recommended. This is the only interval that guarantees the desired coverage probability for all values of $\alpha$ and $p$.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Confidence level desired</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90%</td>
<td>95%</td>
</tr>
<tr>
<td>Wald</td>
<td>801 fails</td>
<td>997 fails (799 fails if $z=2$)</td>
</tr>
<tr>
<td>Agresti &amp; Coull’s ‘Adjusted Wald’</td>
<td>305 fails</td>
<td>217 fails</td>
</tr>
<tr>
<td>Agresti &amp; Coull’s ‘Add 4’</td>
<td>345 fails</td>
<td>211 fails (76 fails if $z=2$)</td>
</tr>
<tr>
<td>Wilson Score</td>
<td>0 fails</td>
<td>0 fails</td>
</tr>
<tr>
<td>Arc Sin</td>
<td>0 fails</td>
<td>0 fails</td>
</tr>
<tr>
<td>Clopper-Pearson</td>
<td>0 fails</td>
<td>0 fails</td>
</tr>
</tbody>
</table>

Table 2. Summary results from 1001 confidence interval trials for activity sampling

Notwithstanding the arguments above, let’s now investigate the contour plots in Figures 2, 4, 7, 9 and 11 again. Overlaying them all into one single plot as in Figure 14, it is possible to see when one interval dominates the other in terms of reducing the interval to a specific half-width with a particular sample size. Note we are ignoring whether the coverage probability actually meets the desired confidence level here. It can be seen that the Agresti and Coull interval generally dominates all of the other intervals. That is, it requires fewer samples to meet its specific half-width. The Wilson interval generally dominates all other intervals, except the Agresti and Coull interval. This can be regarded as good performance as the Wilson Score interval is almost exact. The Arc Sin interval generally dominates the Clopper-Pearson interval. Despite this, it is recommended that the standard operations management procedure is updated to include the Clopper-Pearson interval.
8. Activity sampling procedure using the Clopper-Pearson interval

A procedure to exploit the Clopper-Pearson interval for activity sampling will now be described. Appendix A provides a blank data collection form. The Clopper-Pearson confidence interval requires either; the solution of pair of high order equations (equations (17) and (18)); access to Microsoft Excel to enumerate (20) or specialist statistical software to realise (22) and (23)); or access to Clopper-Pearson confidence interval tables. Appendix B provides Clopper-Pearson confidence intervals that fit the data collection forms for the case of \( n = 50, 100 \) and 150. Appendix C gives the maximum of the solution to (24) and (25), rounded up to the nearest integer, for the required number of samples \( n \) to be within a desired interval half-width for 90%, 95% and 99% confidence for different values of the estimated binomial parameter, \( \hat{p} \).

The following step-by-step guide highlights how to use these tools.

**Step 1.** Determine how much confidence you need to assign to your results. From this, determine the value of \( \alpha \).

**Step 2.** Determine an acceptable confidence interval half length (\( L \)) for the purposes of your study. Note. Steps 1 and 2 should be done in consultation with the “customer” of the activity sampling exercise.

**Step 3.** Explain the purpose of the activity sampling to the subjects of the exercise, (the workers, customers, operators etc). Obtain an initial impression of range of activities being undertaken by the workers, their duration and their frequency. Make a note of the working time, breaks and any data collection issues present.

**Step 4.** Modify the activity sampling data collection form in Appendix A for your situation. Name each of the activities the subject (worker, customer, machine etc) undertakes in the columns A to J. If there are not enough columns then you need to develop your own forms. Specify when the observations are to be taken (ensuring random intervals of time between observations) using the uniformly
distributed random numbers appropriately scaled to fit your scenario. Make sure the study is long enough to capture the range of activities that are undertaken.

**Step 5.** Collect the activity sample. First collect 50 samples (a page worth). Total up the observations for each activity and calculate the upper and lower limits of confidence interval, using the table of Clopper-Pearson confidence intervals in Appendix B. Decide if the confidence interval meets the needs of the study. If it is not then use Appendix C to determine how many samples should be taken in order to reduce the interval half-width ($L$) to an acceptable level.

**Step 6.** If the confidence interval half length ($L$) is too large for your purpose, collect another 50 samples using the data collection form and re-calculate the length of the interval.

**Step 7.** Repeat steps 5 and 6 until you are satisfied with the accuracy of the confidence interval for each of the activities.

### 9. Teaching activity sampling with the new Clopper-Pearson interval

This new activity sampling procedure has been taught to over 2000 students over the last 5 years in Undergraduate, MSc, MBA and Executive courses in the U. S. and the U. K. Only the Clopper-Pearson interval is discussed. No detail on other alternative intervals needs to be given. The data collection forms and tables of intervals and sample sizes are no more difficult to use than a standard normal table. Students often struggle with the difference between and selection of appropriate accuracy ($L$) and confidence ($\alpha$), so it is worthwhile spending some time on this matter. However, on balance, the new Clopper-Pearson approach is actually easier to teach as students often find the circular reference to $p$ in the Wald interval formula difficult. It is also easy to break-out of a presentation to illustrate the Inverse Beta function in Microsoft Excel.

A worked example (activity sampling my 2 year old daughter), a class exercise of activity sampling a secretary’s duties and case study at Sam’s Tailor in Hong Kong is used. These teaching materials are available upon request from the author. Sohal and Oakland (1990) also offer some advice on an engaging method to teach activity sampling that could also be adopted for use within this Clopper-Pearson activity sampling procedure.

### 10. Concluding remarks

The standard operations management textbook treatment of activity sampling has been evaluated. It was found that the Wald interval is almost exclusively used to assign confidence to the results. A review of modern statistical knowledge on binomial proportion confidence intervals revealed that the statisticians have serious concerns with the adequacy of this interval and have developed a range of more sophisticated approximations. There is even an exact solution to the problem.

These modern confidence intervals have been reviewed and analysed for the purpose of activity sampling. The Clopper-Pearson interval was found to be the only interval that actually achieves the desired confidence interval for any desired coverage probability that does not use a-priori information. The Clopper-Pearson interval has an explicit solution, but sadly has no closed form. However, given that Microsoft Excel has an Inverse Beta Distribution function built-in, the Clopper-Pearson interval can be easily determined.
Two look-up tables for the interval and the sample size requirements have been developed for classroom use.

A new activity sampling procedure was developed that can be used without advanced statistical knowledge to gather information about many operations management scenarios. The interval obtained now economically achieves the desired coverage. The existing activity sampling procedure in many OM textbooks cannot guarantee this.

References


Appendix A: Activity sampling data collection form

<table>
<thead>
<tr>
<th>Observation Number</th>
<th>Uniformly distributed random numbers</th>
<th>Intervals driven by random numbers? Y/N</th>
<th>Activity</th>
<th>Time of observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>98837 67136</td>
<td></td>
<td>A</td>
<td>198837</td>
</tr>
<tr>
<td>2</td>
<td>81781 03347</td>
<td></td>
<td>B</td>
<td>81781</td>
</tr>
<tr>
<td>3</td>
<td>35311 82655</td>
<td></td>
<td>C</td>
<td>35311</td>
</tr>
<tr>
<td>4</td>
<td>24689 72155</td>
<td></td>
<td>D</td>
<td>24689</td>
</tr>
<tr>
<td>5</td>
<td>93492 29368</td>
<td></td>
<td>E</td>
<td>93492</td>
</tr>
<tr>
<td>6</td>
<td>10281 25221</td>
<td></td>
<td>F</td>
<td>10281</td>
</tr>
<tr>
<td>7</td>
<td>95991 99716</td>
<td></td>
<td>G</td>
<td>95991</td>
</tr>
<tr>
<td>8</td>
<td>57579 12203</td>
<td></td>
<td>H</td>
<td>57579</td>
</tr>
<tr>
<td>9</td>
<td>12468 46011</td>
<td></td>
<td>I</td>
<td>12468</td>
</tr>
<tr>
<td>10</td>
<td>67468 45271</td>
<td></td>
<td>J</td>
<td>67468</td>
</tr>
<tr>
<td>11</td>
<td>86051 48309</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>31430 42324</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>96616 99230</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>92753 37184</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>55855 79446</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>12736 75741</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>36111 99785</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>58755 87409</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>67051 07571</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>34731 54073</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>94730 49294</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>45804 30298</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>50371 35805</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>53271 59910</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>25086 38884</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>72736 75741</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>67051 07571</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>68051 48309</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>31430 42324</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>96616 99230</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>92753 37184</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>55855 79446</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>86051 48309</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>36111 99785</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>58755 87409</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>45804 30298</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>50371 35805</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>53271 59910</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>25086 38884</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>72736 75741</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>67051 07571</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>34731 54073</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>94730 49294</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>45804 30298</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>50371 35805</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>53271 59910</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>25086 38884</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>72736 75741</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>67051 07571</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>34731 54073</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Running count from any previous samples

Total sum of observations, \( x \)

Observed probability, \( \hat{p} = \frac{x}{n} \)

Lower confidence interval, \( \hat{p}_L \) (App. B)

Upper confidence interval, \( \hat{p}_U \) (App. B)

Sample size required based on \( L \) (App. C)
### Appendix B: Table of Clopper-Pearson confidence intervals

<table>
<thead>
<tr>
<th>n = 50</th>
<th>$\alpha$ = 0.10</th>
<th>$\alpha$ = 0.05</th>
<th>$\alpha$ = 0.01</th>
<th>$\alpha$ = 0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n \times \hat{p}$</td>
<td>$n \times \hat{p}$</td>
<td>$n \times \hat{p}$</td>
<td>$n \times \hat{p}$</td>
</tr>
<tr>
<td>1 (5)</td>
<td>0.03390244</td>
<td>0.03390244</td>
<td>0.03390244</td>
<td>0.03390244</td>
</tr>
<tr>
<td>2 (10)</td>
<td>0.06780488</td>
<td>0.06780488</td>
<td>0.06780488</td>
<td>0.06780488</td>
</tr>
<tr>
<td>3 (15)</td>
<td>0.10170732</td>
<td>0.10170732</td>
<td>0.10170732</td>
<td>0.10170732</td>
</tr>
<tr>
<td>4 (20)</td>
<td>0.13561016</td>
<td>0.13561016</td>
<td>0.13561016</td>
<td>0.13561016</td>
</tr>
<tr>
<td>5 (25)</td>
<td>0.16951292</td>
<td>0.16951292</td>
<td>0.16951292</td>
<td>0.16951292</td>
</tr>
<tr>
<td>6 (30)</td>
<td>0.20341568</td>
<td>0.20341568</td>
<td>0.20341568</td>
<td>0.20341568</td>
</tr>
<tr>
<td>7 (35)</td>
<td>0.23731844</td>
<td>0.23731844</td>
<td>0.23731844</td>
<td>0.23731844</td>
</tr>
<tr>
<td>8 (40)</td>
<td>0.27122120</td>
<td>0.27122120</td>
<td>0.27122120</td>
<td>0.27122120</td>
</tr>
<tr>
<td>9 (45)</td>
<td>0.30512396</td>
<td>0.30512396</td>
<td>0.30512396</td>
<td>0.30512396</td>
</tr>
<tr>
<td>10 (50)</td>
<td>0.33902472</td>
<td>0.33902472</td>
<td>0.33902472</td>
<td>0.33902472</td>
</tr>
</tbody>
</table>

### Examples:

- For different sample sizes $n$ or for different confidence level then the following formulas can be used in Microsoft Excel: $\hat{p}_L = \text{BETAINV}(1 - \frac{\alpha}{2}, x + 1, n - x)$ and $\hat{p}_U = \text{BETAINV}(\frac{\alpha}{2}, x, n - x + 1)$. 

- Note: For Table 1, keep $n$ to 10 different confidence level then the following formulas can be used in Microsoft Excel:
  - $\hat{p}_L = \text{BETAINV}(1 - \frac{\alpha}{2}, x + 1, n - x)$
  - $\hat{p}_U = \text{BETAINV}(\frac{\alpha}{2}, x, n - x + 1)$
### Appendix C: Activity sample size requirements

<table>
<thead>
<tr>
<th>p or 0.1 (1-p)</th>
<th>90% Confidence</th>
<th>95% Confidence</th>
<th>99% Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>----------------</td>
<td>---------------</td>
<td>----------------</td>
<td>---------------</td>
</tr>
<tr>
<td>p or (1-p)</td>
<td>L</td>
<td>p or (1-p)</td>
<td>L</td>
</tr>
<tr>
<td>0.01</td>
<td>73</td>
<td>26</td>
<td>0.04</td>
</tr>
<tr>
<td>0.02</td>
<td>58</td>
<td>46</td>
<td>0.03</td>
</tr>
<tr>
<td>0.03</td>
<td>43</td>
<td>36</td>
<td>0.02</td>
</tr>
<tr>
<td>0.04</td>
<td>33</td>
<td>28</td>
<td>0.01</td>
</tr>
<tr>
<td>0.05</td>
<td>29</td>
<td>24</td>
<td>0.01</td>
</tr>
<tr>
<td>0.06</td>
<td>26</td>
<td>22</td>
<td>0.00</td>
</tr>
<tr>
<td>0.07</td>
<td>23</td>
<td>20</td>
<td>0.00</td>
</tr>
<tr>
<td>0.08</td>
<td>21</td>
<td>18</td>
<td>0.00</td>
</tr>
<tr>
<td>0.09</td>
<td>19</td>
<td>16</td>
<td>0.00</td>
</tr>
<tr>
<td>0.10</td>
<td>17</td>
<td>14</td>
<td>0.00</td>
</tr>
<tr>
<td>0.11</td>
<td>16</td>
<td>13</td>
<td>0.00</td>
</tr>
<tr>
<td>0.12</td>
<td>15</td>
<td>12</td>
<td>0.00</td>
</tr>
<tr>
<td>0.13</td>
<td>14</td>
<td>11</td>
<td>0.00</td>
</tr>
<tr>
<td>0.14</td>
<td>13</td>
<td>10</td>
<td>0.00</td>
</tr>
<tr>
<td>0.15</td>
<td>12</td>
<td>9</td>
<td>0.00</td>
</tr>
<tr>
<td>0.16</td>
<td>11</td>
<td>8</td>
<td>0.00</td>
</tr>
<tr>
<td>0.17</td>
<td>10</td>
<td>7</td>
<td>0.00</td>
</tr>
<tr>
<td>0.18</td>
<td>9</td>
<td>6</td>
<td>0.00</td>
</tr>
<tr>
<td>0.19</td>
<td>8</td>
<td>5</td>
<td>0.00</td>
</tr>
<tr>
<td>0.20</td>
<td>7</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>0.21</td>
<td>6</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>0.22</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>0.23</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0.24</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.26</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example:** Assume that 95% confidence and an interval half width of \( L = 0.05 \) is desired. After the first 50 samples, the estimate of \( p, \hat{p} \) is 0.73. The sample size requires are symmetrical about \( p \), thus in order to save space the table only considers 0 \leq p \leq 0.5. So we need to make the calculation \( p = 1 - 0.73 = 0.27 \). The required sample size is then found in the table (as this refers to the 95% confidence interval) and we pick the number associated with \( L = 0.05 \) and the row associated with \( 0.27 \). Thus the sample size is 174.