The 10% Rule for the formula for a Confidence Interval for a proportion

We now ask when does the boxed formula in Section 8.3 of the text based on the Normal Approximation to the Binomial distribution perform satisfactorily? Many of the standard rules like $n\widehat{p}(1-\widehat{p}) \geq 10$ can perform very badly in either tail of the Binomial distribution. Since confidence intervals focus on tail probabilities (e.g. 0.025 at each end for a 95% interval), this rule should not be used for confidence intervals. Instead we have chosen to make use of a table given by Samuels and Lu [1992], which we have produced in slightly modified form in Appendix A3 of the text. This table was constructed by first of all computing an exact confidence interval and then seeing how the boxed formula matched up with it. A close match will occur when the distance between the lower endpoints of the two intervals plus the distance between the upper endpoints (total d, say) is small relative to the width, w say, of the exact interval. On this criterion we want d/w (expressed as as percentage) to be small, say nor more than 10%. For example suppose that our exact interval is [0.12,0.35] and our approximate interval is [0.11,0.37] then $d = 0.01 + 0.02 = 0.03$, $w = 0.35 - 0.12 = 0.23$ and $d/w = .03/0.23$ (or 13%). The table in Appendix A3 gives the smallest value of n needed for a given $\widehat{p}$ for different percentage accuracies, namely 15%, 10% and 5%, respectively.[1] However, we shall use 10% in most cases, which will give us good accuracy. For future reference we call this rule the **10% rule**. In situations where accuracy is not so crucial we could make do with a 15% accuracy.

What do we do when our value of n is smaller than that prescribed by our 10% rule, or even by a 15% rule? We can no longer use our quick formula for finding the confidence interval and more accurate methods are needed. For example, an "exact" interval, sometimes known as the Clopper-Pearson interval, can be calculated by computer. Alternatively we can use one of the more complex approximate methods which can be used for smaller n such as those described by Bohning [1994] and Newcombe [1998a].

[1]For a 15% accuracy, Samuels and Lu [1992] give the following "Rule of 5 and 25": If there are at least 25 observations in each category (i.e. x and $n - x$ are both at least 25), use the boxed interval; if there are fewer than 5 observations in either category, don't use the interval; in intermediate cases use the table to decide.