Chapter 3

Methodology

In the practical part, the objective is to compare three new types of confidence intervals: using direct and inverse binomial sampling method, using only direct binomial sampling method, and using only inverse binomial sampling method. Moreover, we will compare new confidence intervals with the previously known.

The criteria for comparing between methods are median, mean lengths of confidence intervals and the coverage probabilities, with shorter median lengths as the leading factors. The 95% confidence level is considered, the values for probabilities of success p are .1,.3,.5,.7,.9 (we could choose any particular values), and three levels of sample sizes $n, m = 30,50,100$ are chosen (small, moderate and large sample sizes). For each particular case, the simulations are repeated 10,000 times by using program R version 2.6.1 for calculating the coverage probability , mean ,median , and standard seviation of the interval lengths. The following are the steps for our study.

3.1 Comparing five new confidence intervals

3.1.1 Confidence intervals using direct and inverse binomial sampling methods

3.1.1.1 when the asympotic variance is used.

Step 1: For the first variable we simulate binomial numbers X ~ Binomial(n,p$_1$) with fixed sample sizes $n = 30,50,100$ and probabilities of success $p_1 = .1,.3,.5,.7,.9$. For the second variable Y we apply inverse binomial sampling. That is, let $n = 30,50,100$ and probabilities of success $p_2 = .1,.3,.5,.7,.9$. That is, the stopping time ν is defined by the number of the observation that results in achieving $m = n * p_2$ successes.

Step 2: Calculate the point estimate of $\hat{\theta}_{n,m} = \overline{X}_n \overline{Y}_m$

where $T = \sum_1^n X_k$, $\overline{X}_n = T/n$, $\overline{Y}_m = \upsilon/m$

Step3 : Compute the confidence interval by the formula

$$\hat{\theta}_{n,m} \pm Z_{\alpha/2}\sqrt{\overline{X}_n\overline{Y}_m\left(\frac{\overline{Y}_m(1-\overline{X}_n)}{n}+\frac{\overline{X}_n(\overline{Y}_m-1)}{m}\right)}$$

Step 4: Report the interval length for a ratio of binomial proportions and observe whether the interval covers the true value of the ratio or not.

Step 5: Steps 1-4 are repeated 10,000 times.

Step 6: Compute the mean coverage probabilities, median, mean , and standard deviation of confidence interval lengths.

3.1.1.2 when the true value of the variance is used.

Step 1: For the first variable we simulate binomial numbers X ~ Binomial(n,p_1) with fixed sample sizes $n = 30{,}50{,}100$ and probabilities of success $p_1 = .1,.3,.5,.7,.9$. For the second variable Y we apply inverse binomial sampling. That is, let $n = 30{,}50{,}100$ and probabilities of success $p_2 = .1,.3,.5,.7,.9$. That is, the stopping time υ is defined by the number of the observation that results in achieving $m = n * p_2$ successes.

Step 2: Calculate the point estimate of $\hat{\theta}_{n,m} = \overline{X}_n\overline{Y}_m$

where $T = \sum_1^n X_k$, $\overline{X}_n = T/n$, $\overline{Y}_m = \upsilon/m$

Step3 : Compute the confidence interval by the formula

$$\hat{\theta}_{n,m} \pm Z_{\alpha/2}\hat{\theta}_{n,m}\sqrt{\overline{X}_n\overline{Y}_m\left(\frac{(\overline{Y}_m-1)(1-\overline{X}_n)}{nm}+\frac{\overline{Y}_m(1-\overline{X}_n)}{n}+\frac{\overline{X}_n(\overline{Y}_m-1)}{m}\right)}$$

Step 4: Report the interval length for a ratio of binomial proportions and observe whether the interval covers the true value of the ratio or not.

Step 5: Steps 1-4 are repeated 10,000 times.

Step 6: Compute the mean coverage probabilities, median, mean , and standard deviation of confidence interval lengths.

3.1.1.3 when the number of successes in the second sample is fixed as in the first experiment.

Step 1: For the first variable we simulate binomial numbers X ~ Binomial(n,p_1) with fixed sample sizes $n = 30,50,70,100$ and probabilities of success $p_1 = .1,.3,.5,.7,.9$. For the second variable Y we apply inverse binomial sampling. That is, let $n = 30,50,70,100$ and probabilities of success $p_2 = .1,.3,.5,.7,.9$. In this case, the stopping time υ is defined by the number of the observation that results from the number of successes as in the first experiment.

Step 2: Calculate the point estimate of $\hat{\theta}_{n,m} = \dfrac{\upsilon}{n}$

where $T = \sum_1^n X_k$, $\overline{Y}_T = \upsilon / T$

Step 3: Compute the confidence interval by the formula

$$\hat{\theta}_{n,m} \pm Z_{\alpha/2} \sqrt{\frac{\hat{\theta}_n}{n}(2\overline{Y}_T - \hat{\theta}_n - 1)}$$

Step 4: Report the interval length for a ratio of binomial proportions and observe whether the interval covers the true value of the ratio or not.

Step 5: Steps 1-4 are repeated 10,000 times.

Step 6: Compute the mean coverage probabilities, median, mean , and standard deviation of confidence interval lengths.

3.1.2 Confidence intervals using only direct binomial sampling method

Step 1: For the case of both direct sampling we simulate binomial numbers

X ~ Binomial(n,p_1)

Y ~ Binomial(m,p_2)

with fixed sample sizes $n, m = 30,50,100$ and probabilities of success $p = .1,.3,.5,.7,.9$. In order to achieve such samples for the fixed value of p we apply the following procedure n times. Let u be a uniform $[0,1]$ random number. If $u < p$, then put the sample value equal to 1, otherwise take it equal to 0.

Step 2: Calculate the point estimate of $\hat{\theta} = \hat{\theta}_{n,m} = \dfrac{\overline{X}_n(m+1)}{m\overline{Y}_m + 1}$,

where $n\overline{X}_n = \sum_1^n X_i$, $m\overline{Y}_m = \sum_1^m Y_i$

Step 3: Compute the confidence interval by the formula

$$\hat{\theta}_{n,m} \pm Z_{\alpha/2}\sqrt{\hat{\theta}_{n,m}\left(\frac{1-\overline{X}_n}{n\overline{Y}_m} + \hat{\theta}_{n,m}\frac{(1-\overline{Y}_m)}{m\overline{Y}_m}\right)},$$

Step 4: Report the interval length for a ratio of binomial proportions and determine whether the interval covers the true value of the ratio or not.

Step 5: Steps 1-4 are repeated 10,000 times.

Step 6: Compute the mean coverage probabilities, median, mean , and standard deviation of confidence interval lengths.

3.1.3 Confidence intervals using inverse binomial sampling method

Step 1: For both inverse sampling we apply inverse binomial sampling

X ~ Negative binomial(m_1,p_1)

Y ~ Negative binomial(m_2,p_2)

Let $n = 30{,}50{,}100$ probabilities of success $p = .1,.3,.5,.7,.9$ the stopping time $\upsilon_i; i = 1,2$ is defined by the number of the observation that results in achieving $m_i = n_i * p_i; i = 1,2$ successes.

Step 2: Calculate the point estimate of $\hat{\theta}_{n,m} = \dfrac{\upsilon_2(m_1 - 1)}{(\upsilon_1 - 1)m_2}$.

Step 3: Compute the confidence interval by the formula

$$\hat{\theta}_{n,m} \pm Z_{\alpha/2}\sqrt{\hat{\theta}_{n,m}\left(\frac{\hat{p}_1(1-\hat{p}_2)}{m_2} + \hat{\theta}_{n,m}\frac{1-\hat{p}_1}{m_1}\right)},$$

where $\hat{p}_i = (m_i - 1)/(\upsilon_i - 1)$, $i = 1,2$

Step 4: Report the interval length for a ratio of binomial proportions and observe whether the interval covers the true value of the ratio or not.

Step 5: Steps 1-4 should be repeated 10,000 times.

Step 6: Compute the mean coverage probabilities, median, mean , and standard deviation of confidence interval lengths.

3.2 Comparing new confidence interval with the previously known.

We compare our results with the results presented in the paper by Koopman (1984) and Bailey (1987). We would like to mention that we are using the results included in tables in these papers. For example, we cannot write a program for confidence intervals presented in Koopman (1984). In the paper Koopman (1984) no actual formula for the confidence intervals is presented, because it is completely unclear how to invert some convex function to solve for the confidence interval endpoints. The Koopman's confidence intervals are interesting only from a theoretical point of view, up to our knowledge, they do not have any practical applications due to the computational difficulty mentioned above.

We expect to get some gain in coverage probability and an improvement in the median and mean length of our confidence interval in comparison with all previously known.

We would also like to mention once again that, up to our knowledge, nobody considered the case of inverse binomial sampling before. Because of that we cannot compare our confidence intervals for inverse sampling with the previously known.

Figure 3.1
Procedure of CI using direct and inverse binomial sampling method when the asympotic variance is used.

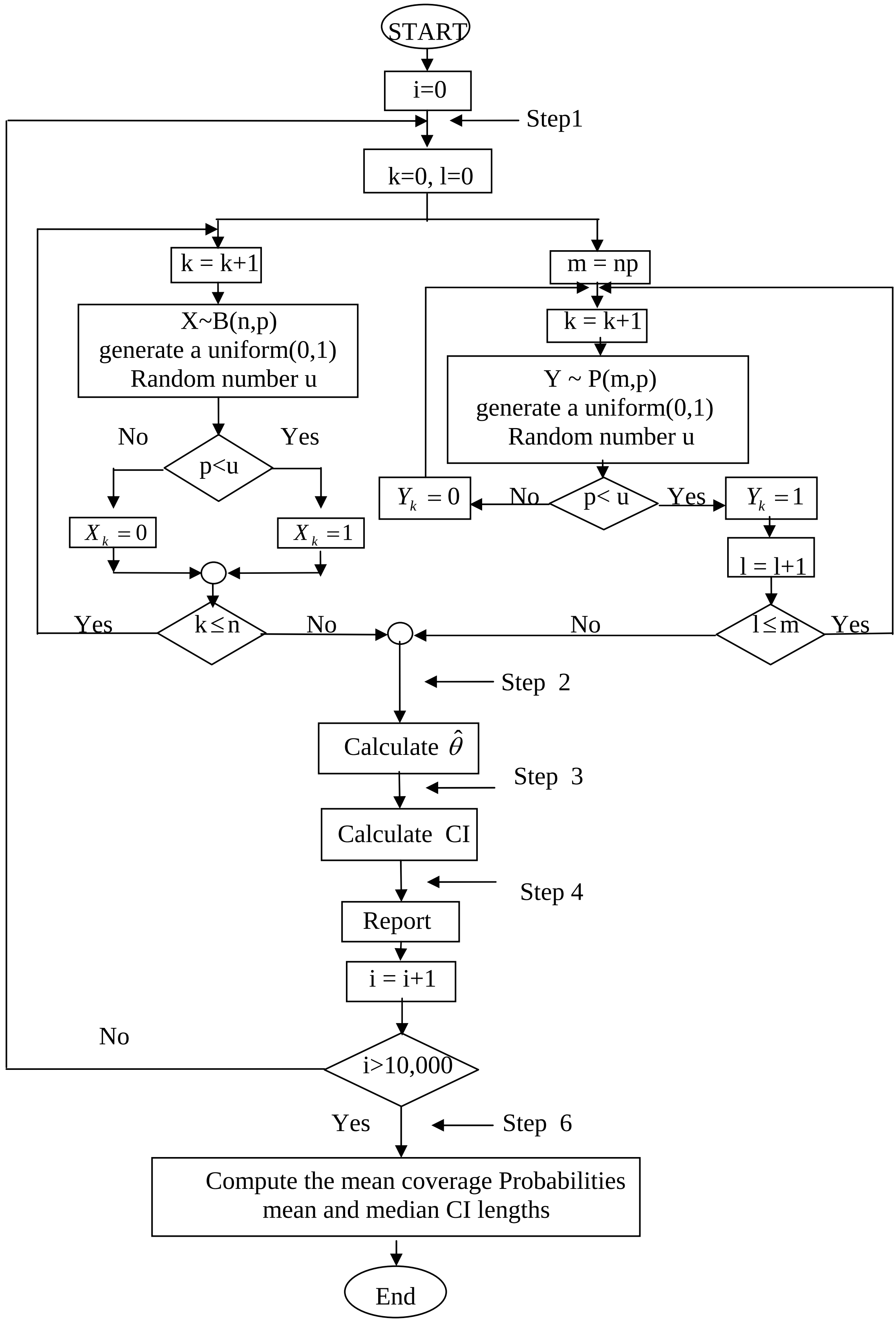

Figure 3.2
Procedure of CI using direct and inverse binomial sampling method when the true value of the variance is used.

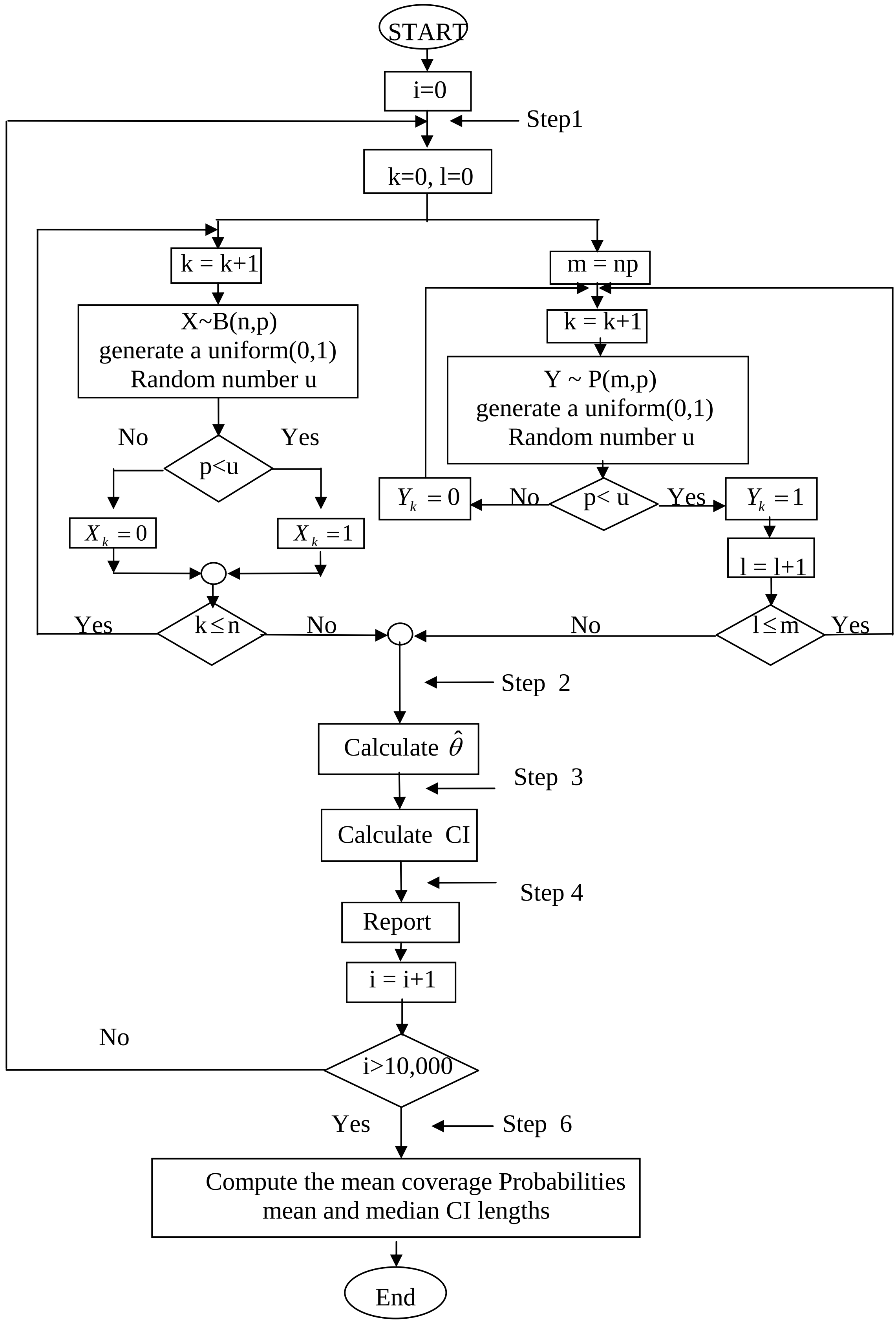

Figure 3.3
Procedure of CI using direct and inverse binomial sampling method when we fix the number of success as in the first experiment.

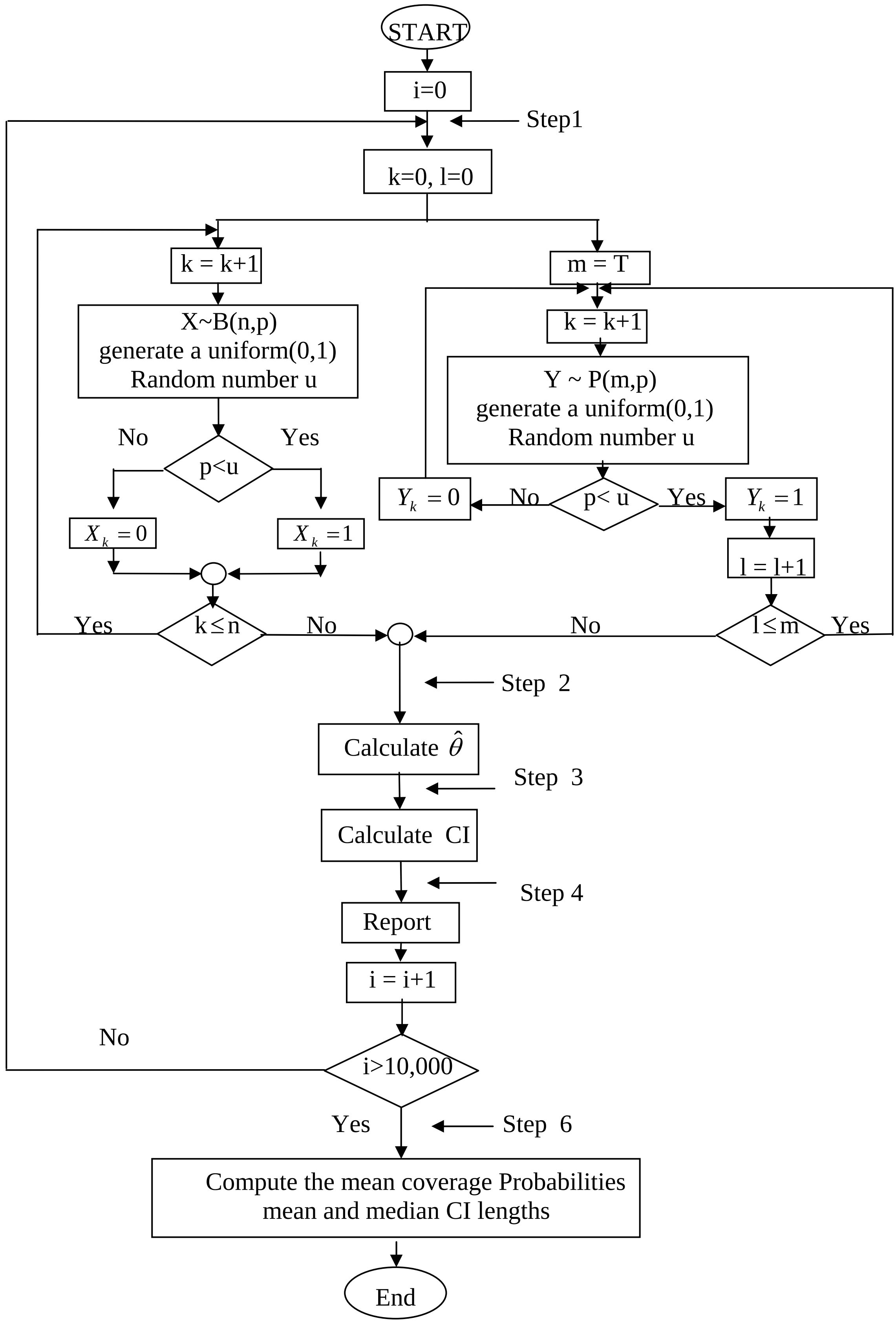

Figure 3.4
Procedure of CI using direct binomial sampling method

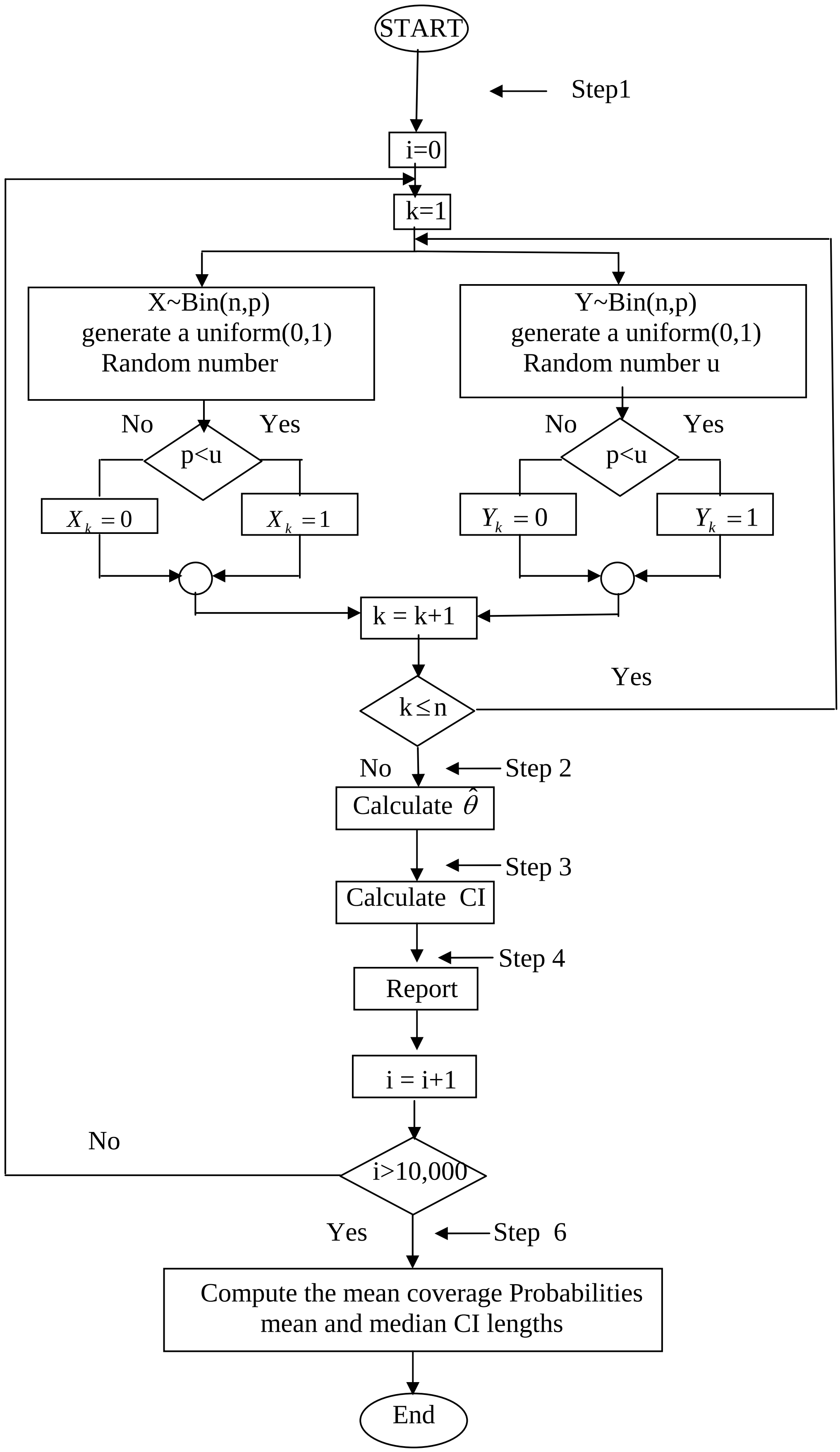

Figure 3.5
Procedure of CI using inverse binomial sampling method