Stat 624 Project

Matthew Morgan

2023-03-31

Introduction

One of the key aspects within the field of statistics is the estimation of desired quantities. In order to estimate these quantities, a multitude of estimation methods exist. The goal of this project is to explore different estimation methods and compare their effectiveness using various metrics. The exploration of different estimation methods is important because it helps to verify their validity, while the comparison of different estimation methods showcases the advantages and disadvantages of each relative to the others.

This project will explore and compare three estimators: the maximum likelihood estimate, the method of moments estimate, and the Bayes estimate under squared error loss. The exploration and comparison of these three estimators will be two-fold. First, a simulation study will be conducted for each estimator under various settings to evaluate its effectiveness in capturing the truth. Second, the estimation methods will be applied to a real dataset in an effort to estimate the parameters of the selected sampling distribution.

The specific dataset analyzed in this project consists of Lionel Messi's goals scored by season in the domestic league in which he played. Only completed seasons were considered, resulting in 18 seasons worth of data scraped from FBref.com. Since the data is count data, the negative binomial distribution was selected for this project as it is used to model count data and matched the domain of the data. The negative binomial distribution was used in both the simulation study and the application to the real dataset.

The hope is that the exploration of estimators will be useful for various potential areas. Specifically, in the case of considering a negative binomial distribution for a dataset, a more informed decision can be made about which estimator would be most effective. Additionally, when estimating parameters of a distribution, one key decision is whether to use a Bayesian or a frequentist framework. This project explores both methodologies so that the reader can be informed about how parameter estimation works in each framework. Finally, the overall consideration of estimators is important in any type of analysis. The practice of considering different estimators and choosing the most effective one is critical, and the general methodology presented in this project could be applied to any analysis that includes the estimation of parameters.

Methodology

The negative binomial distribution considered in this project is governed by a size parameter r and a probability parameter p. The density of the negative binomial distribution is

$$p(y) = \frac{\Gamma(y+r)}{\Gamma(r)y!} p^r (1-p)^y$$

for $y = 0, 1, 2, ..., r > 0$ and $0 < p \leq 1$.

The mean is $\mu = r(1-p)/p$ and variance $r(1-p)/p^2$.

Each estimator considered in this project was calculated using a certain methodology. For the maximum likelihood estimates of the parameters of the negative binomial distribution, the methodology was to maximize the negative binomial likelihood. The r and p values that maximize the likelihood of the negative binomial distribution are then the respective maximum likelihood estimates of those parameters. In practice, the log-likelihood of the negative binomial distribution was maximized for computational stability. The maximization

of the log-likelihood was done by implementing the Nelder and Mead (1965) method of optimization to the log-likelihood function.

For the method of moments estimates for the parameters of the negative binomial distribution, the methodology. was to find the method of moments equations for both parameters. This was done by beginning with the following equations that use the mean of the negative binomial distribution, $E(y)$, and the variance of the negative binomial distribution, $V(y)$:

$$\frac{E(y)}{V(y)} = \frac{r(1-p)}{p} \times \frac{p^2}{r(1-p)} = p$$

$$\frac{E(y)^2}{V(y)} = \frac{r^2(1-p^2)}{p^2} \times \frac{p^2}{r(1-p)} = r(1-p)$$

Next, these equations were solved for r and p:

$$p = \frac{E(y)}{V(y)}$$

$$\frac{E(y)^2}{V(y)} = r\left(1 - \frac{E(y)}{V(y)}\right) \quad \frac{E(y)^2}{V(y)} = r\left(\frac{V(y) - E(y)}{V(y)}\right) \quad r = \frac{E(y)^2}{V(y)} \times \frac{V(y)}{V(y) - E(y)} \quad r = \frac{E(y)^2}{V(y) - E(y)}$$

Finally, the sample mean $\bar{y}$ and the sample variance s_y^2 were substituted into these equations for r and p to obtain their method of moments estimators:

$$\hat{p} = \frac{\bar{y}}{s_y^2} \qquad \hat{r} = \frac{\bar{y}^2}{s_y^2 - \bar{y}}$$

For the Bayes estimate under squared error loss of the parameters of the negative binomial distribution, the methodology was to implement a Markov Chain Monte Carlo algorithm to obtain posterior samples for r and p and then the mean of those posterior samples for r and p would be the Bayes estimates under squared error loss for those respective parameters. The priors for r and p were the following:

$$p(r) \sim \text{Gamma}(.01, .01)$$

$$p(p) \sim \text{Beta}(.1, .1)$$

The size parameter r had Gamma prior with a shape of .01 and a rate of .01 placed on it. A Gamma distribution was chosen as its support matched the support of r. A shape of .01 and a rate of.01 were chosen so that the Gamma prior would be a relatively non-informative prior overall.

The probability parameter p had a Beta prior with shape parameters .1 and .1 placed on it. A Beta distribution was chosen as this would result in the full conditional distribution of p being conjugate and would allow for greater efficiency in the MCMC algorithm. Shape parameters of .1 and .1 were chosen so that the prior mean of p would be .5 and have a wide enough variance so that the data would have more influence in the posterior overall.

The details of the MCMC algorithm used to obtain posterior samples for r and p was the following:

Posterior samples for this analysis were obtained using an MCMC algorithm. The specific details of the MCMC algorithm for this analysis were the following:

Let J be the desired number of samples to be obtained from the joint posterior distribution $p(r, p|\mathbf{y})$.

Let $\boldsymbol{\Theta} = (r, p)$ be the vector containing the parameters of interset.

Denote the jth sample by $\boldsymbol{\Theta}^{(j)} = \left(r^{(j)}, p^{(j)}\right)$.

2

1. Begin with an initial state $\boldsymbol{\Theta}^{(0)}$.

2. Obtain the next sample $\boldsymbol{\Theta}^{(j+1)}$ by updating each parameter in the vector.

   - Obtain $r^{(j+1)}$ via a Gaussian Metropolis-Hastings random walk where a proposal, r^*, is generated by sampling from $\mathcal{N}(r^{(j)}, \epsilon)$, a Gaussian proposal distribution centered around $r^{(j)}$ with variance ϵ. r^* is then accepted with probability $\min\left(1, \frac{p(r^*, p^{(j)}|\mathbf{y})}{p(r^{(j)}, p^{(j)}|\mathbf{y})}\right)$

   - Obtain $p^{(j+1)}$ via Gibbs sampling from its full conditional distribution

     - $p \sim \text{Beta}\left(.1 + r^{(j+1)}n, .1 + \sum_{i=1}^{n} y_i\right)$ where n is the number of observations.

3. Repeat the above step J times.

As mentioned earlier, in order to evaluate the effectiveness of the estimators, a simulation study was conducted. Specifically, this simulation study was used to determine the bias of the point estimates, the mean squared error of the point estimates, and the coverage of the uncertainty intervals for each of the estimators under a variety of true underlying parameters and samples sizes.

For the maximum likelihood estimators the bias and mean squared error of the point estimates were obtained by repeatedly generating a data set of the desired sample size using the true underlying parameters and then calculating the maximum likelihood estimates for . The bias and squared error were calculated for each generated data set and then the average bias and mean squared error metrics were obtained by taking the average of all the calculated biases and squared errors. Additionally, the average point estimate which was calculated by taking the average of the point estimates for each generated data set was also included.

To calculate coverage, uncertainty intervals were obtained through repeated bootstrapping. For each bootstrap sample, an uncertainty interval was calculated for the maximum likelihood estimators. It was then checked whether each of the uncertainty intervals contained the true underlying parameters. Furthermore, the widths of the uncertainty intervals for each bootstrap procedure were recorded as well. The coverage for each of the maximum likelihood estimators was calculated by taking the average of the indicators indicating whether or not the uncertainty intervals contained the true underlying parameters. Finally, the average uncertainty interval width was calculated by taking the average of the interval widths for each bootstrap procedure.

The metrics for the method of moments estimators were obtained in the same way the metrics for the maximum likelihood estimates were obtained, by performing a simulation study and repeated bootstrapping.

For the Bayes estimators under squared error loss, the averge bias and mean squared error of the point estimates were obtained through the use of the same MCMC algorithm outlined previously. The bias and squared error of the point estimates were calculated for each posterior draw and then the average bias and mean squared error of the point estimates were calculated by taking the average of the biases and squared errors for each posterior draw.

To calculate coverage, uncertainty intervals were obtained through repeated MCMC sampling. For each MCMC chain, an uncertainty interval was calculated for the Bayes estimators under squared error loss. For each MCMC chain, it was then checked whether each of the uncertainty intervals contained the true underlying parameters. Coverage was then calculated as the average of these indicators indicating whether or not the uncertainty intervals contained the true underlying parameters. Again, uncertainty interval width was recorded for each MCMC chain and the average uncertainty interval width for each Bayes estimator under squared error loss was computed.

Overall this simulation study helped to accomplish the research goal of evaluating and comparing estimators as the strengths and weaknesses of these estimators became more evident upon an examination of the calculated metrics in various settings.

The final step of this project involved applying the estimation methods to the data set of the goals that Lionel Messi scored each domestic league season mentioned earlier. While the true parameters are unknown in this application to real data, as a more robust evaluation of the estimators, uncertainty intervals were determined for each estimator. Again, this involved using bootstrapping to calculate uncertainty intervals

for the maximum likelihood and method of moments estimators and using MCMC sampling to calculate uncertainty intervals for the Bayes estimators under squared error loss.

Simulation Study

A simulation study was used to evaluate and compare the three estimators considered in this project. Three different settings were used for the true underlying parameters of the negative binomial distribution in this simulation study. The first setting considered was $r = 5$ and $p = .1$. The second setting considered was $r = 10$ and $p = .5$. The third setting considered was $r = 25$ and $p = .9$. For each of these three sets of true underlying parameters, three different sample sizes of $n = 10$, $n = 100$, and $n = 1,000$ were used.

For both the maximum likelihood estimators and the method of moments estimators, the average point estimates, the average bias of the point estimates, and the mean squared error of the point estimates were calculated from 10,000 randomly generated negative binomial data sets each with sample size n and true underlying parameters r and p. The coverage of the uncertainty intervals and the average uncertainty interval widths were calculated from repeating the prodecure of taking 1,000 bootstrap samples of a single negative binomial data set with sample size n and true underlying parameters r and p 100 times.

For the Bayes estimators under squared error loss, the average point estimates, the average bias of the point estimates, and the mean squared error of the point estimates were calculated from an MCMC chain of 10,000 posterior samples which used a negative binomial data set with sample size n and true underlying parameters r and p. The coverage of the uncertainty intervals and the average uncertainty interval widths were calculated from 100 MCMC chains of 1,000 posterior samples where each MCMC chain used a negative binomial data set with sample size n and true underlying parameters r and p.

Presented below are the results of the simulations studies. Each table contains the following quantites: average point estimate, average bias of the point estimate, mean squared error of the point estimate, coverage of the uncertainty interval, and average uncertainty interval with. Each table includes these quantities for one estimation method, one setting of true underlying parameters, and all three samples sizes considered.

Table 1: MLE Effectiveness (r = 5, p = .1)

	Estimate	Bias	MSE	Coverage	Width
r					
n = 10	8.459917	3.459917	238.992595	0.71	227.390395
n = 100	5.191418	0.191418	0.738390	0.94	3.427779
n = 1000	5.015039	0.015039	0.059922	0.95	0.966595
p					
n = 10	0.143249	0.043249	0.009950	0.73	0.443805
n = 100	0.103323	0.003323	0.000251	0.93	0.062345
n = 1000	0.100267	0.000267	0.000021	0.95	0.018131

Table 1 contains the results of the simulation study for the maximum likelihood estimators when the true underlying parameters were $r = 5$ and $p = .1$. An examination of this table reveals that the average bias of the point estimates, the mean squared error of the point estimates, and the average width of the uncertainty intervals all decreased as the sample size increased. The coverage of the uncertainty intervals increased as the sample size increased. It should also be noted that the initial values used were $r = 1$ and $p = .5$.

Table 2: MoM Effectiveness (r = 5, p = .1)

	Estimate	Bias	MSE	Coverage	Width
r					
n = 10	7.377789	2.377789	174.194744	0.88	97.456169
n = 100	5.162747	0.162747	0.828355	0.92	3.620833
n = 1000	5.016135	0.016135	0.076563	0.96	1.056754
p					
n = 10	0.134404	0.034404	0.008317	0.74	0.516899
n = 100	0.102844	0.002844	0.000279	0.91	0.064830
n = 1000	0.100277	0.000277	0.000027	0.96	0.019662

Table 2 contains the results of the simulation study for the method of moments estimators when the true underlying parameters were $r = 5$ and $p = .1$. An examination of this table reveals that the average bias of the point estimates, the mean squared error of the point estimates, and the averge width of the uncertainty intervals all decreased as the sample size increased. The coverage of the uncertainty intervals for r increased or stayed the same the sample size increased. The coverage of the uncertainty intervals for p increased when the sample size increased from 10 to 100 but slightly decreased when the sample size increased from 100 to 1,000. Since the coverage values are close to each other between the two cases, this behavior is not of concern.

Table 3: Bayes Effectiveness (r = 5, p = .1)

	Estimate	Bias	MSE	Coverage	Width
r					
n = 10	5.898159	0.898159	8.422607	0.95	7.777220
n = 100	4.917712	−0.082288	0.691197	0.99	3.846422
n = 1000	5.147833	0.147833	0.126247	0.98	2.343979
p					
n = 10	0.113540	0.013540	0.002583	0.96	0.149659
n = 100	0.103055	0.003055	0.000287	0.98	0.070896
n = 1000	0.102664	0.002664	0.000059	0.96	0.042824

Table 3 contains the results of the simulation study for the Bayes estimators under squared error loss when the true underlying parameters were $r = 5$ and $p = .1$. An examination of this table reveals that the average bias of the point estimates, the mean squared error of the point estimates, and the average width of the uncertainty intervals all decreased as the sample size increased. The coverage of the uncertainty intervals for r and p increased as the sample size increased from 10 to 100 and then decreased slightly when the sample size increased from 100 to 1,000. Since the coverage values are close to each other between the two cases, this behavior is not of concern. It should also be noted that the proposal variance used in this simulation study was $\epsilon = 0.1$ and each chain used an initial guess of $r = 1$ and $p = .5$.

Table 4: MLE Effectiveness (r = 10, p = .5)

	Estimate	Bias	MSE	Coverage	Width
r					
n = 10	125.391615	115.391615	80, 109.619527	0.78	1,078.091037
n = 100	11.319368	1.319368	22.334111	0.92	28.666398
n = 1000	10.099334	0.099334	0.881551	0.95	3.674931
p					
n = 10	0.630026	0.130026	0.073032	0.78	0.587296
n = 100	0.516500	0.016500	0.006286	0.93	0.302385
n = 1000	0.501493	0.001493	0.000533	0.94	0.089800

Table 4 contains the results of the simulation study for the maximum likelihood estimators when the true underlying parameters were $r = 10$ and $p = .5$. An examination of this table reveals that the average bias of the point estimates, the mean squared error of the point estimates, and the averge width of the uncertainty intervals all decreased as the sample size increased. The coverage of the uncertainty intervals increased as the sample size increased. One point of concern is that in the setting, the effectiveness of the maximum likelihood estimator for r was poor in the case when a sample size of 10 was used. It should also be noted that the initial values used were $r = 1$ and $p = .5$.

Table 5: MoM Effectiveness (r = 10, p = .5)

	Estimate	Bias	MSE	Coverage	Width
r					
n = 10	Inf	Inf	Inf	0.98	379.819842
n = 100	11.084154	1.084154	31.160994	0.93	21.354907
n = 1000	10.090876	0.090876	0.916743	0.96	3.720235
p					
n = 10	0.647633	0.147633	0.180983	0.71	2.488011
n = 100	0.511720	0.011720	0.006224	0.91	0.297671
n = 1000	0.501189	0.001189	0.000562	0.95	0.092026

Table 5 contains the results of the simulation study for the method of moments estimators when the true underlying parameters were $r = 10$ and $p = .5$. An examination of this table reveals a disadvantage of the method of moments estimators when a small sample size is used. When a small sample size is used, there is a greater probability that the sample mean could equal the sample variance. When this happens, the method of moments estimator for r is dividing by 0 and hence estimates r to be ∞. It can be seen that this happened for at least one of the generated data sets when a sample size of 10 was used. The fact that the method of moments estimator for r can estimate r to be ∞ highlights the instability of the method of moments estimators when the sample size is small.

Other than this, the behavior of the method of moments estimators seemed normal. Besides the method of moments estimator for r when a sample size of 10 was used, the average bias of the point estimates, the mean squared error of the point estimates, and the averge width of the uncertainty intervals all decreased as the sample size increased. The coverage of the uncertainty intervals increased as the sample size increased.

Table 6: Bayes Effectiveness (r = 10, p = .5)

	Estimate	Bias	MSE	Coverage	Width
r					
n = 10	5.578742	−4.421258	27.886742	0.68	10.126270
n = 100	10.961432	0.961432	8.864200	0.91	8.306092
n = 1000	9.725034	−0.274966	0.882840	0.87	5.046111
p					
n = 10	0.339110	−0.160890	0.041359	0.78	0.370694
n = 100	0.492930	−0.007070	0.004062	0.90	0.224258
n = 1000	0.493199	−0.006801	0.000612	0.86	0.143953

Table 6 contains the results of the simulation study for the Bayes estimators under squared error loss when the true underlying parameters were $r = 10$ and $p = .5$. An examination of this table reveals that the average bias of the point estimates, the mean squared error of the point estimates, and the average width of the uncertainty intervals all decreased as the sample size increased. The coverage of the uncertainty intervals for r and p increased as the sample size increased from 10 to 100 and then decreased slightly when the sample size increased from 100 to 1,000. Since the coverage values are close to each other between the two cases, this behavior is not of concern. It should also be noted that the proposal variance used in this simulation study was $\epsilon = 0.1$ and each chain used an initial guess of $r = 5$ and $p = .7$.

Table 7: MLE Effectiveness (r = 25, p = .9)

	Estimate	Bias	MSE	Coverage	Width
r					
n = 10	320.641447	295.641447	216,544.819602	0.74	232.644759
n = 100	93.707476	68.707476	20,268.851116	0.92	3.564163
n = 1000	32.968499	7.968499	673.348862	0.96	0.987542
p					
n = 10	0.878889	−0.021111	0.030024	0.75	0.476190
n = 100	0.895366	−0.004634	0.008103	0.93	0.065000
n = 1000	0.901138	0.001138	0.001562	0.94	0.018490

Table 7 contains the results of the simulation study for the maximum likelihood estimators when the true underlying parameters were $r = 25$ and $p = .9$. An examination of this table reveals that the average bias of the point estimates, the mean squared error of the point estimates, and the averge width of the uncertainty intervals all decreased as the sample size increased. The coverage of the uncertainty intervals increased as the sample size increased. It should also be noted that the initial values used were $r = 1$ and $p = .5$.

Table 8: MoM Effectiveness (r = 25, p = .9)

	Estimate	Bias	MSE	Coverage	Width
r					
n = 10	Inf	Inf	Inf	0.80	Inf
n = 100	-17.847437	-42.847437	$991,929.728385$	0.96	392.520914
n = 1000	38.410654	13.410654	$175,269.633431$	0.96	334.442028
p					
n = 10	1.127488	0.227488	0.512312	0.87	Inf
n = 100	0.919208	0.019208	0.018555	0.90	0.530266
n = 1000	0.901951	0.001951	0.001715	0.94	0.159103

Table 8 contains the results of the simulation study for the method of moments estimators when the true underlying parameters were $r = 25$ and $p = .9$. An examination of this table again reveals the earlier mentioned disadvantage of the method of moments estimators when a small sample size is used. Although in this case, the problem is more pronounced as the average width of the uncertainty interals for r and p were ∞ when a sample size of 10 was being used. Furthermore, this table reveals another disadvantage of the method of moments estimators. Looking at the method of moments estimator for r when a sample size of 100 was used, it can be seen that the average point estimate is negative, and this is outside of the parameter space for r. When the sample variance is less than the sample mean, the method of moments estimator for r is dividing by a negative number and hence estimates r to be negative. Seeing as this happened enough that the average point estimate for r across the 10,000 generated data sets was negative, this showcases another and possibly major instability of the method of moments estimator.

Besides these noted anomalies, the method of moments estimator for r when a sample size of 1,000 was used produced a valid average point estimate but a large mean squared error and a large average uncertainty interval width. Besides the average uncertainty interval width for p when a sample size of 10 was used, the method of moments estimator for p behaved normally. The average bias of the point estimates, the mean squared error of the point estimates, and the averge width of the uncertainty intervals all decreased as the sample size increased. The coverage of the uncertainty intervals for r and p increased as the sample size increased.

Table 9: Bayes Effectiveness (r = 25, p = .9)

	Estimate	Bias	MSE	Coverage	Width
r					
n = 10	15.152124	-9.847876	403.865529	0.51	25.801555
n = 100	14.454779	-10.545221	184.042686	0.73	26.292863
n = 1000	23.890839	-1.109161	75.906363	0.63	14.648095
p					
n = 10	0.666249	-0.233751	0.104281	0.64	0.484278
n = 100	0.785262	-0.114738	0.021406	0.76	0.190139
n = 1000	0.890874	-0.009126	0.001117	0.65	0.069222

Table 9 contains the results of the simulation study for the Bayes estimators under squared error loss when the true underlying parameters were $r = 25$ and $p = .9$. An examination of this table reveals that the average bias of the point estimates, the mean squared error of the point estimates, and the average width of the uncertainty intervals all decreased as the sample size increased. The coverage of the uncertainty intervals for r and p increased as the sample size increased from 10 to 100 and then decreased when the sample size increased from 100 to 1,000. This could indicate that for case when a sample size of 1,000 was used, the MCMC

algorithm was not exploring the parameter spaces well. Perhaps an alternative proposal variance could have been chosen to try and improve the coverage in this case where the sample size was 1,000. Although for this simulation study, a constant proposal variance of $\epsilon = 1$ was used. Overall, the fact that the proposal variance may need to be adjusted for different sample sizes is evidence that the MCMC algorithm is sensitive to the proposal variance chosen. It should also be noted that each chain used an initial guess of $r = 20$ and $p = .5$.

For the first set simulation studies where the true underlying parameters were $r = 5$ and $p = .1$, all three estimation methods performed similarly as each method did better as the sample size increased. When comparing the three estimation methods, the maximum likelihood estimators outperformed the other two methods with respect to average bias and mean squared error. The Bayes estimates under squared error loss had the highest coverage and smallest average uncertainty interval width when a sample size of 10 was used.

For the second set of simulation studies where the true underlying parameters were $r = 10$ and $p = .5$, the instability of the method of moments estimators with small sample sizes became evident. While both the Bayes estimators under squared error loss and the maximum likelihood estimates performed similarly and did better as the sample size increased, the maximum likelihood estimators seemed to have performed better on average than the Bayes estimators under squared error loss did.

For the third set of simulation studies where the true underlying parameters were $r = 25$ and $p = .9$, the instability of the method of moments estimators with small sample sizes again became evident and further instability was showcased when r was estimated to be negative. Again, both the Bayes estimators under squared error loss and the maximum likelihood estimates performed similarly and did better as the sample size increased. The maximum likelihood estimators outperformed the Bayes estimators under squared error loss with respect to the uncertainty intervals as the maximum likelihood estimators had higher coverage and smaller average uncertainty interval widths. On the ohter hand, the Bayes estimators under squared error loss outperformed the maximum likelihood estimators with respect to average bias and mean squared error as the Bayes estimators under squared error loss had lower average biases and lower mean squared errors.

There are some important things to note overall about the three estimation methods. The maximum likelihood simulation studies had the greatest computational burden overall as those simulation studies took the longest to run. Overall, the performance of the maximum likelihood estimators did not appear to be very dependent on initial guesses of the parameters. On the other hand, the performance of the Bayes estimators under squared error loss did appear to depend on the initial guesses of the parameters as well as the proposal variance. Finally, the general instability of the method of moments estimators is a concern and should be considered in future implementation.

Results

Lionel Messi is considered by many to be the greatest soccer player of all time. Figure 1 is a histogram of the goals Messi has scored each season for the domestic league he played in.

Figure 1: Lionel Messi's Domestic League Goals Each Season
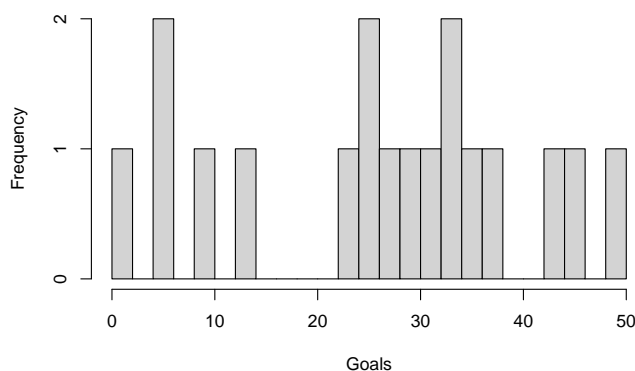
On average, Messi scored 26.67 domestic league goals per season. The variance of his domestic league goals acreess the 18 seasons considered was 205.06. The minimum number of domestic league goals he scored in a season was 1 and the maximum number of domestic league goals he scored in a season was 1.

As with the simulation study, the data was assumed to follow a negative binomial distribution. With this assumption, the goal was to estimate the size and probability parameters of the negative binomial distribution when it was applied to the data set of Lionel Messi's domestic league goals each season. First, the estimates were obtained using the maximum likelihood estimation procedure. The maximum likelihood estimate of the size parameter was $\widehat{r_{MLE}} = 2.4577079$ and the maximum likelihood estimate of the probability parameter was $\widehat{p_{MLE}} = 0.0843818$. Then, 10,000 bootstrap samples were taken to quantify the uncertainty of the maximum likelihood estimates.

Table 10: MLE

	Estimate	Lower	Upper
r	2.45770788	1.20231753	10.9840619
p	0.08438181	0.04838526	0.2623053

Table 10 shows the maximum likelihood estimates of the size and probability parameters along with the 95% uncertainty intervals associated with those estimates that were obtained from the 10,000 bootstrap samples.

Next, the estimates were obtained using the method of moments estimation procedure. The mthod of moemts estimate of the size parameter was $\widehat{r_{MoM}} = 3.9862241$ and the method of moments estimate of the probability parameter was $\widehat{p_{MoM}} = 0.130044$. Again, 10,000 bootstrap samples were taken to quantify the uncertainty of the method of moments estimates.

Table 11: MoM

	Estimate	Lower	Upper
r	3.986224	1.93367520	14.138618
p	0.130044	0.08129934	0.311603

Table 11 shows the method of moments estimates of the size and probability parameters along with the 95% uncertainty intervals associated with those estimates that were obtained from the 10,000 bootstrap samples.

Finally, Bayes estimates under squared error loss were obtained using the MCMC procedure outlined in the Methodology section. 10,000 MCMC samples were obtained using a proposal variance of $\epsilon = 1$ and an initial guess of the parameters that was $r = 1$ and $p = .2$. The resulting Bayes estimate under squared error loss of the size parameter was $\widehat{r_{Bayes}} = 2.3513452$ and the resulting Bayes estimate under squared error loss of the probability parameter was $\widehat{p_{Bayes}} = 0.080247$.

With the 10,000 posterior samples, the trace plots of the size and probability parameters were the following:
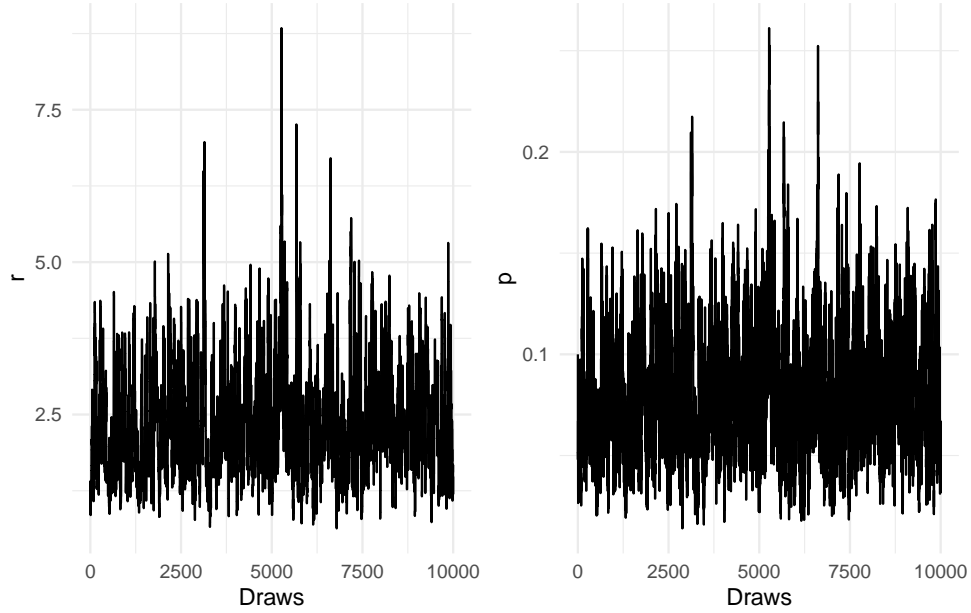
Figure 2: Trace Plots of r and p

Figure 2, the trace plots of r and p indicated that the posterior chains for r and p had good mixing and with this, the assumption was made that the posterior chains for r and p had approximately converged. With approximate convergence of the posterior chains, the Bayes estimates under squared error loss of r and p were considered to be valid estimates.

Table 12: Bayes

	Estimate	Lower	Upper
r	2.35134524	1.03282724	4.4975868
p	0.08024704	0.03296866	0.1498693

Table 12 shows the Bayes estimates under squared error loss of the size and probability parameters along with the uncertainty intervals associated with those estimates that were obtained from the 10,000 posterior samples.

The estimated mean of Messi's domestic league goals each season using the maximum likelihood estimates came out to be 26.6683321. The estimated variance of Messi's domestic league goals each season using the maximum likelihood estimates came out to be 316.0436254.

The estimated mean of Messi's domestic league goals each season using the method of moments estimates came out to be 26.6666667. The estimated variance of Messi's domestic league goals each season came out to be 205.0588235. It should be noted that the fact that the method of moments estimators exactly estimated the actual mean and variance of Messi's domestic league goals each season was expceted as the method of moments estimators are constructed off of the mean and variance of the data.

The estimated mean of Messi's domestic league goals each season using the Bayes estimates under squared error loss came out to be 26.9499889. The estimated variance of Messi's domestic league goals each season using the Bayes estimtaes under squared error loss came out to be 335.8378071. The key thing to ntoe about the Bayes estimates under squared error loss if that their uncertainty intervals for r and p were the smallest when compared to the other two methods.

Conclusion

The goal of this project was to explore and compare three different estimation methods: the maximum likelihood estimator, the method of moments estimator, and the Bayes estimate under squared error loss. These methods were compared using the negative binomial distribution to estimate the size and probability parameters. When applied to a real dataset of Lionel Messi's domestic league goals, all three methods effectively estimated the mean, and the Bayes estimates under squared error loss had the least amount of associated uncertainty. Simulation studies were conducted under various settings to evaluate the effectiveness of the three estimation methods, which revealed the strengths and weaknesses of each method. The maximum likelihood estimators performed well in all settings but had the largest computational burden. The method of moments estimators performed well with a large enough sample size and when the variance of the data was greater than the mean. The Bayes estimates under squared error loss performed well across all settings when a good initial guess of the parameters was used and the variance of the normal proposal distribution was well tuned.

The fact that each estimation method has its strengths and weaknesses leads to the conclusion that a decision about which estimation method to use is not an arbitrary matter. There are multiple considerations to take into account when choosing an estimation method, such as how much data is available, whether or not prior beliefs should be implemented, and whether there are closed form solutions for the estimation method. Overall, the decision about which estimation method to use in an analysis should be thoughtful and deliberate. When this is done, a more thorough and robust analysis is likely to result.