Quest Journals
Journal of Research in Applied Mathematics
Volume 7 ~ Issue 5 (2021) pp: 07-13
ISSN(Online) : 2394-0743 ISSN (Print): 2394-0735
www.questjournals.org

Research Paper

Estimation of accurate shot and goal probabilities in football matches with Geometric Distribution and Negative Binomial (Pascal) Distribution

Şenol Çelik
Bingöl University Faculty of Agriculture Biometry and Genetic Department Bingöl, Turkey

ABSTRACT
In this study, it was aimed to determine the accurate shot rates and goal probabilities in football matches played in the Super League in Turkey in the 2020-2021 season. Geometric and Negative Binomial distributions are used to achieve this. With the geometric distribution, the probability of the teams to get the first hit in the 3rd, 4th and 5th shots was calculated and also, the probability of reaching the first goal in the 3rd, 4th and 5th hits for accurate shots has been determined. With the Negative Binomial distribution, the probabilities of getting the second, third and fourth hits on the 5th, 6th and 7th shots were calculated. Again, the probabilities of finding the 2nd and 3rd goals in the 4th, 5th and 6th shots were determined. Using these probabilities, the expected value and variance values of the Geometric distribution and the Negative Binomial distribution were also calculated.

KEYWORDS: *Geometric distribution, Negative Binomial distribution, football, shooting*

Received 14 May, 2021; Revised: 28 May, 2021; Accepted 30 May, 2021 © The author(s) 2021. Published with open access at www.questjournals.org

I. INTRODUCTION

Spor Toto Super League is the highest level football league in Turkey, which started in 1959 and continues to this day. The super league, which has been played with 18 teams for many years, last competed in the 2020-2021 season and is a double-circuit system in which each team plays two games. The top-ranked team is the champion, while the last four teams are relegated to a lower league, while three teams in one lower league advance to the Super League to compete the following season. The 2020-21 season of the Super League consists of 40 weeks and 420 matches. At the end of the 2018–19 season, the league is ranked 10th in the UEFA countries rankings and sends 2 teams to the UEFA Champions League and 3 teams to the UEFA Europa League. A team that is the Turkish Cup winner and does not make the top 4 can also participate in the UEFA Europa League. Six teams have won the championship in the Super League so far. They are Galatasaray (22), Fenerbahce (19), Besiktas (16), Trabzonspor (6), Bursaspor (1) and Istanbul Basaksehir (1). The team that won the championship in the 2020-21 season is Besiktas [1].

Football is a goal game. To reach the goal, it is necessary to shoot a lot of shots or even hit accurate shots. When the hit shot increases, the likelihood of a goal increases. Teams that score more goals also have a good chance of winning the game. The probability of scoring the first accurate shot and the first goal is calculated by geometric distribution, and the probability of scoring 2nd, and 3rd goals and more in a large number of accurate shots is calculated by negative binomial distribution.

Studies have been carried out in various fields regarding geometric and negative binomial distribution. In one of these studies, Geometric and Pareto distributions were used when comparing the classical reward/punishment system and the damage severity with the reward/punishment system. As a result of the systems they created with the second-order loss function and net premium, it was concluded that the systems that take into account the damage frequency and damage severity produce fairer results [2]. In another study, two different results were determined and two loss functions were compared using the second degree and exponential loss function when creating a reward/penalty system using the Negative Binomial distribution [3]. Negative Binomial distribution was used in the optimal reward/penalty system created using the theory of creditworthiness, the bayesci approach and the expected value premium principle [4]. [5] determined the reward/penalty premium values for compulsory traffic insurance insurances in Turkey using the Negative Binomial distribution.

In this study, it is aimed to calculate the probability of 21 teams scoring in 40 super league games with accurate shots and accurate shots.

II. MATERIAL AND METHODS

Material

The material of the research consisted of the results of the football matches played every week in the Turkish Super League in the 2020-2021 season. Spor Toto Super League" was entered on the www.google.com website and browsed.

"https://www.google.com/search?q=spor+toto+s%C3%BCper+lig&sxsrf=ALeKk00w_x66fR1gk1A2gdmrvYXxJo7CnQ%3A1618086188694&ei=LAlyYKLwKYPzkgWr-" website was used [6]. On the relevant site, the matches of all teams throughout the season are entered every week and examined one by one. The total number of shots, accuracy and goals scored by the teams were recorded throughout the match. In the 2020-2021 season, 21 teams took part in the Turkish Super League. These are the teams of Besiktas, Fenerbahce, Galatasaray, Trabzon, Alanya, Antalya, Ankaragucu, Erzurum, Denizli, Fatih Karagümrük, Gaziantep, Genclerbirligi, Göztepe, Hatay, Istanbul Basaksehir, Kasimpasa, Kayseri, Konya, Rize, Sivas and Malatya. Each team has 2 matches with all teams, on their own court and on the opponent's field (away). Each team has 40 games during the season. Accuracy (%) in total shots fired by teams each week was calculated. Goal rate in hit-and-run (%) was calculated. According to these calculations, the accuracy rate (p) was calculated first, and the probability of how many accurate shots were scored on the total shots was calculated by negative binomial distribution. Again, the probability of the first accurate shot on any number shot was calculated by geometric distribution. The same was done for the calculation of the probability of the first goal scored in the accurate shot.

Method

Geometric Distribution

Geometric distribution relates to the first success in the xth trial and is a special version of the negative binomial distribution [7]. A Bernoulli trial with a probability of success p is repeated under the same conditions until an independent success is achieved [8]. The probability function of the geometric distribution is defined as follows [9].

$$P(X = x) = q^{x-1}p, \qquad x = 1,2, \dots$$

Here, p is the probability of success, that is, the probability of an undesired situation; and q is the probability of failure, that is, an undesired situation. q = 1-p.

The expected value for the geometric distribution is.

$$E(X) = \sum_{x=1}^{\infty} xpq^{x-1} = p\sum_{x=1}^{\infty} \frac{d}{dq} q^x = p\frac{d}{dq}\sum_{x=1}^{\infty} q^x = p\frac{d}{dq}\left(\frac{q}{1-q}\right) = \frac{1}{p}$$

Variance

$$V(X) = \frac{q}{p^2}$$

is defined as this [9]. In short, the number of attempts required to achieve initial success by continuing to conduct independent trials until they achieve initial success is geometrically random variable.

Negative Binomial (Pascal) Distribution

A Bernoulli trial with a probability of success in negative binomial distribution is considered to be conducted independently under the same conditions until r achieves success [8].

The probability function of this distribution is,

$$P(X = x) = \binom{x-1}{r-1} q^{x-r}p^r, \qquad x = r, r+1, r+2, \dots$$

where 0<p< 1 and r is an integer such that r≥1 [10].

If X has a Negative Binomial distribution presented by probability function, then expected value and variance as follow, respectively [9].

$$E(X) = \frac{r}{p}$$

$$V(X) = \frac{rq}{p^2}$$

III. RESULTS

The total shots, accurate shots and goals scored by the teams in all matches (40 games) in the 2020-2021 season are presented in Table 1. The accuracy rate and the goal rate in the accurate shot are also calculated and included in Table 1.

Table 1. Teams' shots, accuracy and goals

Order	Teams	Shots	Shots on target	Goal	Shots on target rate (%)	Goal ratio on shots on target (%)
1	Alanya	647	244	58	0.377	0.238
2	Antalya	368	144	41	0.391	0.285
3	Ankaragücü	425	162	46	0.381	0.284
4	Beşiktaş	554	212	89	0.383	0.420
5	Başakşehir	534	162	42	0.303	0.259
6	Denizli	414	145	38	0.350	0.262
7	Erzurum	471	165	44	0.350	0.267
8	Fenerbahçe	607	231	72	0.381	0.312
9	Galatasaray	569	238	80	0.418	0.336
10	Gaziantep	434	162	59	0.373	0.364
11	Gençlerbirliği	424	137	44	0.323	0.321
12	Göztepe	361	135	59	0.374	0.437
13	Hatay	464	176	62	0.379	0.352
14	Karagümrük	476	180	64	0.378	0.356
15	Kasımpaşa	458	182	47	0.397	0.258
16	Kayseri	413	143	35	0.346	0.245
17	Konya	500	173	49	0.346	0.283
18	Malatya	414	142	49	0.343	0.345
19	Rize	429	158	53	0.368	0.335
20	Sivas	455	168	54	0.369	0.321
21	Trabzon	511	191	50	0.374	0.262

NOTE: Teams are given in alphabetical order.

When table 1 was examined, Alanya was the team with the most shots with 647. Fenerbahce, Galatasaray and Besiktas, respectively. Their shots were 607, 569 and 554, respectively. Teams with a high total number of shots are also expected to have a high probability of accurate shots and scoring goals. The teams with the highest probability of accurate shots were Galatasaray, Kasimpasa and Antalyaspor, respectively. Their accuracy was 0.418, 0.397 and 0.391, respectively. Besiktas, who finished the league champion, had a 0.383 shot probability. The teams with the highest probability of scoring on the hit shot were Göztepe, Besiktas and Gaziantep, respectively. Their scoring odds were 0.437, 0.420 and 0.364, respectively. According to this information, when the accuracy of the shot probabilities is calculated, the accuracy rate parameter is p, the hit rate parameter is q, and the number of shots is x. In the 2020-2021 season, Besiktas won the super league and were p=0.383 and q=0.617 for the probability of shooting third, fourth, fifth or first shot in any order according to geometrical distribution. The probability of Besiktas hitting their first shot on their third shot is calculated as follows.

$$P(X = x) = q^{x-1}p$$
$$P(X = 3) = 0.617^{3-1} * 0.383 = 0.617^2 * 0.383 = 0.146$$

According to this result, there is a 14.6% chance that the champion Besiktas will hit the first accurate shot on his third attempt. Similarly, the same calculation is done for the probability of scoring goals. Since the probability of scoring for Besiktas is p=0.42, q=0.58, the probability of him scoring the first goal on his fourth hit,

$$P(X = 4) = 0.58^{4-1} * 0.42 = 0.58^3 * 0.42 = 0.082$$

that is, there is an 8.2% chance that Besiktas will score the first goal on their fourth shot. Besiktas' number of accurate shots expected to be until they score the first goal is E(X)=1/p=1/0.42=2.38.
In short, Besiktas needs to shoot an average of 2.38 hits before scoring the first goal.
When the probability of accurate shots and goals is calculated according to the negative Binomial distribution, for example if the probability of shooting 3 of the 5th total shot of the champion Besiktas is calculated,

$$P(X = x) = \binom{x-1}{r-1} q^{x-r} p^r$$

$$P(X = 5) = \binom{5-1}{3-1} 0.617^{5-3} 0.383^3 = 0.128$$

Here, it was seen that there is a 12.8% chance that Besiktas will shoot 3rd in total shooting. The possibility of them scoring their 2nd goal on the 4th shot,

$$P(X = 4) = \binom{4-1}{2-1} 0.58^{4-2} 0.42^2 = 0.178$$

Besiktas had a 17.8% chance of scoring for the second time on their 4th shot. The number of accurate shots that Besiktas is expected to take until they score the 2nd goal is E(X)=r/p=2=4.76. If Besiktas scores 2 goals, they should shoot an average of 4.76 or about 5 hits.
These calculations were made for all teams in various numbers for total shots, accurate shots and goals. Accurate shot probability calculations by geometric distribution are given in Table 2 and goal probabilities are given in Table 3. Accurate shot probability calculations according to negative binomial distribution are presented in Table 4 and goal probabilities are presented in Table 5.

Table 2. Probabilities of first accurate shots based on geometric distribution results

Order	Teams	First target on 3rd shot	First target on 4th shot	First target on 5th shot
1	Alanya	0.14631	0.09114	0.05677
3	Ankaragücü	0.14597	0.09033	0.05590
2	Antalya	0.14498	0.08825	0.05372
5	Başakşehir	0.14722	0.10256	0.07145
4	Beşiktaş	0.14583	0.09003	0.05558
6	Denizli	0.14787	0.09608	0.06243
7	Erzurum	0.14786	0.09606	0.06241
8	Fenerbahçe	0.14602	0.09045	0.05603
9	Galatasaray	0.14155	0.08234	0.04790
10	Gaziantep	0.14662	0.09189	0.05759
11	Gençlerbirliği	0.14804	0.10021	0.06783
12	Göztepe	0.14656	0.09176	0.05744
13	Hatay	0.14613	0.09070	0.05630
14	Karagümrük	0.14623	0.09093	0.05655
15	Kasımpaşa	0.14431	0.08696	0.05241
16	Kayseri	0.14798	0.09674	0.06325
17	Konya	0.14799	0.09679	0.06330
18	Malatya	0.14806	0.09727	0.06391
19	Rize	0.14697	0.09284	0.05865
20	Sivas	0.14691	0.09266	0.05845
21	Trabzon	0.14658	0.09179	0.05748

When Table 2 was examined, the teams with the highest probability of the first accurate shot on the 3rd shot were Malatya, Genclerbirligi and Konya, respectively. Their first hit shots were 0.14806, 0.14804 and 0.14799, respectively. The teams with the highest probability of first-shot shots on the 4th shot were Basaksehir (0.10256), Genclerbirligi (0.10021) and Malatya (0.09727), respectively. The teams with the highest probability of first-hit shots on the 5th shot are Basaksehir (0.07145), Genclerbirligi (0.06783) and Malatya (0.06391), respectively.

Table 3. First goal probabilities on accurate shots based on geometric distribution results

Order	Teams	1st goal on 3rd shots on target	1st goal on 4th shots on target	1st goal on 5th shots on target	E(X)	V(X)
1	Alanya	0.13813	0.10529	0.08027	4.207	13.491
3	Ankaragücü	0.14559	0.10425	0.07465	3.512	8.823
2	Antalya	0.14567	0.10419	0.07453	3.522	8.881
5	Başakşehir	0.14225	0.10537	0.07805	2.382	3.292
4	Beşiktaş	0.14132	0.08199	0.04757	3.857	11.020
6	Denizli	0.14271	0.10531	0.07771	3.816	10.744
7	Erzurum	0.14341	0.10517	0.07712	3.750	10.313
8	Fenerbahçe	0.14767	0.10164	0.06996	3.208	7.085
9	Galatasaray	0.14814	0.09835	0.06529	2.975	5.876
10	Gaziantep	0.14722	0.09361	0.05951	2.746	4.793
11	Gençlerbirliği	0.14800	0.10047	0.06820	3.114	6.581
12	Göztepe	0.13851	0.07798	0.04390	2.288	2.947
13	Hatay	0.14780	0.09573	0.06201	2.839	5.220
14	Karagümrük	0.14767	0.09516	0.06133	2.813	5.098
15	Kasımpaşa	0.14209	0.10539	0.07818	3.872	11.123
16	Kayseri	0.13961	0.10544	0.07963	4.086	12.607
17	Konya	0.14551	0.10430	0.07476	3.531	8.935
18	Malatya	0.14801	0.09694	0.06349	2.898	5.500
19	Rize	0.14814	0.09845	0.06543	2.981	5.906
20	Sivas	0.14800	0.10043	0.06815	3.111	6.568
21	Trabzon	0.14266	0.10532	0.07775	3.820	10.772

As shown in Table 3, the teams with the highest probability of first goals on the 3rd accurate shot were Rize, Galatasaray and Malatya, respectively. Their first goal odds on accurate shots were 0.14814, 0.14814 and 0.14801, respectively. The teams with the highest probability of scoring first goals in the 4th shot were Kayseri (0.10544), Kasimpasa (0.10539) and Başakşehir (0.10537), respectively. The teams with the highest probability of first goals in the 5th shot are Alanya (0.08027), Kayseri (0.07963) and Kasimpasa (0.07818), respectively. In addition, the average of the number of accurate shots that teams must shoot until they find the first goal, i.e. the expected value and variance, has been calculated.

Table 4. 2nd, 3rd and 4th accuracy shooting probabilities on different number of shots according to negative Binomial distribution results

Order	Teams	2nd shots on target 5th shot	3rd shots on target 6th shot	4th shots on target 7th shot
1	Alanya	0.13748	0.12962	0.09776
3	Ankaragücü	0.13772	0.13124	0.10005
2	Antalya	0.13813	0.13513	0.10575

5	Başakşehir	0.12446	0.09439	0.05727
4	Beşiktaş	0.13780	0.13183	0.10090
6	Denizli	0.13460	0.11786	0.08256
7	Erzurum	0.13461	0.11789	0.08260
8	Fenerbahçe	0.13769	0.13100	0.09971
9	Galatasaray	0.13776	0.14406	0.12051
10	Gaziantep	0.13720	0.12803	0.09558
11	Gençlerbirliği	0.12951	0.10462	0.06761
12	Göztepe	0.13725	0.12832	0.09597
13	Hatay	0.13762	0.13050	0.09900
14	Karagümrük	0.13755	0.13003	0.09834
15	Kasımpaşa	0.13823	0.13732	0.10914
16	Kayseri	0.13399	0.11598	0.08032
17	Konya	0.13395	0.11587	0.08018
18	Malatya	0.13346	0.11444	0.07850
19	Rize	0.13677	0.12593	0.09276
20	Sivas	0.13686	0.12633	0.09329
21	Trabzon	0.13724	0.12824	0.09587

According to the results obtained in Table 4, the teams with the highest probability of second-shot shots on the 5th shot were Kasimpasa (0.13823), Antalya (0.13813) and Besiktas (0.13780), respectively. The teams with the highest probability of third-hit shots on the 6th shot were Galatasaray (0.14406), Kasimpasa (0.13732) and Antalya (0.13513), respectively. The teams with the highest probability of 4th shot on 7th shot were Galatasaray (0.12051), Kasimpasa (0.10914) and Antalya (0.10575), respectively.

Table 5. 2nd, 3rd and 4th accuracy shooting probabilities on different number of shots according to negative Binomial distribution results

Order	Teams	2nd goal 4th shots on target	3rd goal 5th shots on target	3rd goal 6th shots on target	E(X)	V(X)
1	Alanya	0.09850	0.04683	0.05950	8.414	26.982
3	Ankaragücü	0.12402	0.07043	0.08405	7.024	17.647
2	Antalya	0.12443	0.07085	0.08447	7.043	17.762
5	Başakşehir	0.11064	0.05737	0.07083	4.764	6.584
4	Beşiktaş	0.17798	0.14943	0.14450	7.714	22.041
6	Denizli	0.11220	0.05881	0.07233	7.632	21.489
7	Erzurum	0.11473	0.06119	0.07478	7.500	20.625
8	Fenerbahçe	0.13808	0.08608	0.09875	6.417	14.170
9	Galatasaray	0.14939	0.10043	0.11112	5.950	11.751
10	Gaziantep	0.16086	0.11717	0.12416	5.492	9.587
11	Gençlerbirliği	0.14260	0.09160	0.10363	6.227	13.162
12	Göztepe	0.18160	0.15873	0.14893	4.576	5.895
13	Hatay	0.15619	0.11005	0.11880	5.677	10.439
14	Karagümrük	0.15751	0.11201	0.12030	5.625	10.195
15	Kasımpaşa	0.11008	0.05685	0.07029	7.745	22.245
16	Kayseri	0.10251	0.05018	0.06316	8.171	25.215

17	Konya	0.12364	0.07004	0.08367	7.061	17.869
18	Malatya	0.15322	0.10575	0.11543	5.796	11.000
19	Rize	0.14908	0.10002	0.11078	5.962	11.812
20	Sivas	0.14272	0.09175	0.10376	6.222	13.136
21	Trabzon	0.11204	0.05866	0.07217	7.640	21.545

As seen in Table 5, the teams most likely to score the second goal on the 4th accurate shot were Göztepe, Besiktas and Gaziantep, respectively. Their first goal probability in accurate shots is 0.18160, 0.17798 and 0.16086, respectively. The teams with the highest probability of scoring on the hit shot were Göztepe, Besiktas and Gaziantep, respectively. Their first goal probability in accurate shots is 0.15873, 0.14943 and 0.11717, respectively. The teams with the highest probability of scoring 3rd goals in the 6th accuracy were Göztepe, Besiktas and Gaziantep, respectively. Their first goal probability in accurate shots is 0.14893, 0.14450 and 0.12416, respectively. The E(X) and V(X) statistics show the number and variance of accurate shots teams are expected to take until they score the 2nd goal.

IV. CONCLUSION

In the 2020-2021 season, the overall accuracy rate in the Turkish Super League was 0.367. The goal ratio in hit shots was 0.312. It has been observed that Goztepe, Besiktas and Gaziantep teams are more likely to score 2 or more goals when a different number of accurate shots are fired. Considering that the geometric and negative binomial distribution is appropriate to be used to calculate goal probabilities in football matches, it is recommended to expand similar studies in the future.

REFERENCES

[1]. Anonymous, 2021. Süper Lig. https://tr.wikipedia.org/wiki/S%C3%BCper_Lig
[2]. Mert, M., Saykan, Y. 2005. On a Bonus-Malus system where the claim frequency distribution is Geometric and the Claim Severity Distribution is Pareto. Hacettepe Journal of Mathematics and Statistics, 34:75–81.
[3]. Durak, N. T., Erdoğan, K. 2010. Trafik sigortalarında ödül-ceza sistemi üzerine bir uygulama. Reasürör, (Ekim), 4–13.
[4]. Bülbül, S. E., Baykal, K. B. 2016. Optimal Bonus-Malus system design in motor third-party liability insurance in Turkey: Negative Binomial Model. International Journal of Economics and Finance, 8(8):205–211
[5]. Köse, A. 2018. Zorunlu trafik sigortalarının fiyatlandırılmasında Negatif Binom dağılımına göre ödül/ceza sistemi. Ordu Üniversitesi Sosyal Bilimler Araştırmaları Dergisi, 8(1):9-15.
[6]. Anonymous, 2021. Süper Lig.
[7]. https://www.google.com/search?q=spor+toto+s%C3%BCper+lig&sxsrf=ALeKk00w_x66fR1gk1A2gdmrvYXxJo7CnQ%3A1618086188694&ei=LAlyYKLwKYPzkgWr-
[8]. Sınıksaran, E. 2011. Teori ve Uygulamalarıyla İstatistiksel Yöntemler. Türkmen Kitabevi, İstanbul.
[9]. Öztürk, F. 2010. Olasılık ve İstatistiğe Giriş I. Gazi Kitabevi, Ankara.
[10]. Meyer, P. L. 1970. Introductory Probability and Statistical Applications. Addison-Wesley Publishing Company, USA.
[11]. Yates, R. D., Goodman, D. J. 1999. Probability and Stochastic Processes. A friendly introduction for electrical and computer engineers. John Wiley and sons, Inc., New York.