Proceedings of the 2nd International Conference on Applied Physics and Mathematical Modeling
DOI: 10.54254/2753-8818/53/20240184

Statistical mathematics in video games

Zeming Li

RCF Experimental School, Beijing, 100028, China

lizeming070215@yeah.net

Abstract. Statistics play a crucial role in game design and development, with statistical mathematics being an essential element that developers must carefully consider. The paper introduces how statistical elements can be utilized to develop a creative gameplay mechanism that precisely aligns with the general requirements of players. Since the player's gaming experience is considered the most important factor in a game, developers should be open to modifying their game based on feedback from dissatisfied players, which will enable the development team to provide an enhanced experience for all players. For multiplayer online games, the matching mechanism, which determines the teammates and opponents for each player, is just as crucial as the game's basic gameplay mechanism. Through statistical classification methods, we can improve the fairness of this matching process by analyzing players' previous behavior and labeling them with their level. Overall, we can conclude that statistics can be used to design games by establishing a general framework and incorporating compelling elements. Additionally, statistics are essential for balancing game attributes and facilitating the perfect matching process.

Keywords: Video games, statistics, classification, CS2, League of Legends.

1. Introduction

Statistic mathematics is commonly used in game design and development. While game designing is a commonly accessible process, there are only a limited amount of essays and articles about the mathematical algorithms involved in game design. However, the statistical mathematical are essential elements that should be carefully considered by developers.

In order to enrich players' gaming experience and promote creativity in game mechanisms, the utilization of statistics in game design is becoming more prevalent and important. The paper introduces how this theory could contribute to the development of creative gameplay mechanisms. Since the player's game experience is considered the most important factor of a game, developers could alter their game based on feedback from unsatisfied players so that the development team could offer a better experience for players. For multiplayer online games, the matching mechanism, which determines the teammates and opponents for each player, is just as important as their basic gameplay mechanism. Through statistical classification methods, we can improve the fairness of this matching process by analyzing players' previous behavior and labeling them with their level. This paper aims to offer points of reference to game developers to encourage them to improve the game mechanism. With theoretical support, game designers would provide better games to players.

DOI: 10.54254/2753-8818/53/20240184

2. The prevalence of statistics in game design

First of all, it is said that video games are a popular way of amusing oneself, no matter one's age, gender, or job. According to statistics, the number of video gamers continuously increases year by year, and in particular, it is projected that there will be 2.58 billion gamers worldwide in 2024, which represents an increase of 0.16 billion compared to the previous year (2023). Furthermore, it is predicted that the total number of gamers will reach 3.02 billion by 2029 [1]. When compared to the estimated world population of 8.1 billion, video gamers comprise a large proportion of our total population, and games have gained increased attention as the player base continues to grow.

In order to improve the game performance and maximize players' gaming experience, game development teams consider the use of statistical methods due to their diverse and valuable applications. For game design, it is important to plan what their game mechanism is going to be and how to attract more players before the team actually starts coding their game. So, the primary use of statistics is to collect feedback on previous successful and unsuccessful games, and then analyze the factors that contribute to their popularity, allowing development teams to draw inspiration for designing games. After determining the general orientation of their game, development teams may involve other data sets to deal with their basic game mechanism, for example, matching players by analyzing players' previous performance. To make sure the developed game meets players' requirements, the development teams must focus on players' feedback and adjust some improper aspects of the game. In this way, developers could further upgrade their game performance over the long term, leading to a significant improvement in players' experience.

3. The use of statistics in designing process

3.1. Setting general orientation

There are numerous games in the world, each with its own unique and attractive features, and these distinguishing characteristics set them apart from others and appeal to users who are interested in their gameplay. However, despite their distinctive gameplay strategies, these games have some similarities as well. Based on these similarities, we can sort games into several main categories: shooter games, action-adventure games, survival games, battle royale games, racing games, role-playing games(RPG), simulation games, strategy games, casual games, sports games, and puzzle games.

Each of these general genres of games has its loyal fan base, but the level of interest from players varies across each genre. The following diagram illustrates the popular game genres preferred by US gamers in 2022 (Figure 1) [2].

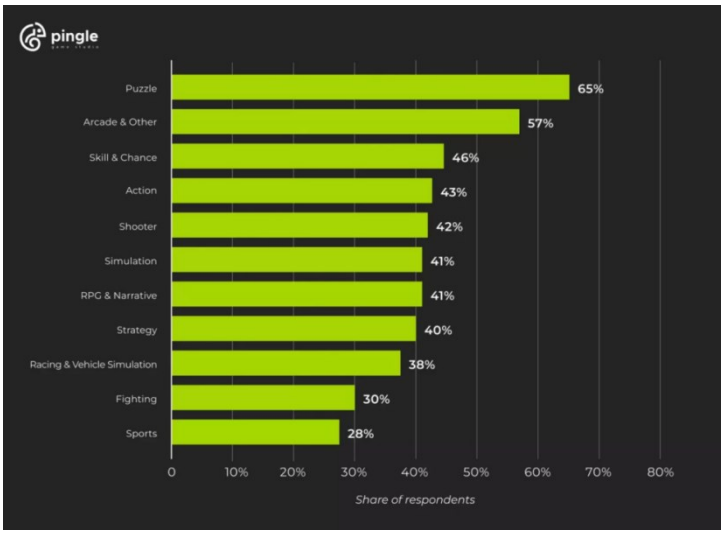

Figure 1. the popular game genres preferred by US gamers in 2022 [2].

DOI: 10.54254/2753-8818/53/20240184

Setting general orientation in game planning is quite similar to the survey that is done before entrepreneurship. The development teams do not have to choose the genre of game that is most welcome, but they must choose one or more genres to define the characteristics of their game so that there would not be any misunderstanding about their game's position. Additionally, conducting statistical surveys to establish a general orientation can not only clarify a game's plan, but also ensure that there is a substantial player base who may be interested in their game. Consequently, the development teams will be able to offer greater products with a better understanding of their project and increased confidence in their product.

When the development teams have already determined their game's genre, they need to make plans for the game's content. The game user research is a suitable and effective statistical method to collect demanding data.

The Games User Research focuses on understanding players' behavior, interactions, and experiences in video games. Researchers use methodologies like observations, interviews, and surveys to gather valuable data. This data helps improve games, remove bugs, and increase player experience [3]. By applying this game's user research, researchers would figure out whether people who never played this game before are interested in their idea or not, and they could also check if players respond to their game ideas as they wish. This type of research helps developers to learn from the gamer's perspective and promotes the game to meet players' requirements properly.

3.2. Designing gameplay mechanism

3.2.1. Training models. An extraordinary game should feature innovative and engaging gameplay mechanics that appeal to players. Gamers typically seek out challenges and derive satisfaction from solving problems, therefore it is important to incorporate some difficult tricks in games. Take board games like chess as examples, the game could be designed to be played online, allowing players to engage in player versus player (PvP) games, but the game could also feature built-in console games such as challenges against artificial intelligence (AI) or a series of levels that progressively increase in difficulty, known as player versus environment (PvE) games. Therefore, players are more likely to invest additional time in this game, resulting in a significantly enhanced gaming experience.

In order to make those built-in console games more interesting and challenging, it is necessary to involve statistics to help with the game design [4]. In the game-designing process, statistical analysis can be used to collect data on how players interact with the game and use this data to make predictions about how the game will be played in the future [5]. Continuing with the chess game example, the developers could create a specific dataset to store a great amount of data on PvP matches, and then use this data to train a model that could analyze the positive and negative effects of any single step in a game and learn from this feedback. Finally, the model would be able to play with human players by selecting the choice that is most likely to generate positive results according to the player's last step. As a result, the developers could create a model that could imitate human players.

While the development teams could train an AI to play with actual gamers, the difficulty of the game must be carefully addressed. It is highly unreasonable to assume that every player possesses similar abilities, as there will inevitably be individuals who have dedicated extensive time to practicing chess, alongside others who may not even grasp the rules of the game. Therefore, the developers should offer different levels of challenges to make sure that neither gamers nor trained AI overwhelms the other side. So when players download this chess game, they could choose to play with an opponent with a similar level to practice or choose to play with an opponent with a higher level to challenge themselves. In addition, by providing different levels of opponents, players would be encouraged to practice more to enhance their skills and challenge themselves with higher-level AI, so players would have a better game experience and gain a huge sense of achievement when they finally defeat an opponent after their hard work [6].

As long as the development teams could use statistics to train a model, they could use statistics to provide different levels of difficulty by altering the quantity and quality of the training dataset. The

DOI: 10.54254/2753-8818/53/20240184

more data has been trained and the higher the quality of data input, the higher the ability of the model. Therefore, the development team could control the ability of a model by limiting the input training data.

3.2.2. Discrete mathematics and probability. If the game components are designed to remain the same throughout the game, then gamers would easily get used to the game's basic rules and lose interest in playing. As a consequence, the random elements are a crucial part of a game. Statistics reveal the fundamental logic algorithms in analyzing probabilities of an event, which means it could be used to organize and implement random variables in a game system [7]. Moreover, probability factors into random event occurrence and AI decision-making, introducing an element of unpredictability that can enhance gameplay and increase replay value [8].

For instance, 2048, the casual mini-game, has a simple gameplay mechanism: synthesize two blocks of identical numbers to get a single block with the sum of the previous blocks. Gamers start with 2 and finally reach 2048, which is the goal of the game. This game may appear simple and uninteresting, leading some players to not invest much time in it. However, many gamers actually spend a significant amount of time on this seemingly basic game due to the unpredictable nature of the random elements and the sudden appearance of new blocks. Whenever a player finished a step, a new block would be added to the board with a random number and random position. This makes the game much more complicated and interesting because it makes players' operations more difficult since there is only limited space on the game board. With random elements, even a very simple game could be attractive.

4. The use of statistics in maintaining the process

4.1. Balancing game characters

Almost all of the successful games that are prevalent and popular include plenty of characters that are relevant to the essence of the gameplay mechanism. These characters might be figures, weapons, or abilities, but they must have distinctive features that can be used to distinguish themselves from others, as they all belong to the same category, they cannot be far more powerful than the others. For example, in the game League of Legends, a multiplayer online battle arena video game developed and published by Riot Games, there are a variety of champions that have unique abilities and powers, and each player in the same match could select one champion they would like to use in this match. Since League of Legends is a MOBA game that has five players on one side and another five players on the opposite side, there must not be any champion with overpower because it would cause an imbalance in the game and significantly affect other players' game experience negatively. The development teams aim at encouraging gamers to practice more to improve their skills rather than just simply picking some champions that do not need any skill but have outstanding performance. Therefore, game balancing is very important for assuring players' game experience.

On the contrary, there must be a standardized method for determining whether a champion is unbalanced and for measuring the champion's performance in a single game. Gamer's feedback is absolutely a good way of gathering information about balancing problems because the champion that is complained about most players definitely causes a negative game experience for other players. Based on gamers' feedback, the development team can easily identify any overpowered elements and make appropriate adjustments to reduce their impact. However, it is important to note that comments from gamers are inherently subjective, as they often express their emotions. Therefore, it is common to encounter claims of a certain champion being overwhelmingly powerful when in fact, that champion may be relatively weaker compared to others.

In order to carry out a method that can give game managers reasonable and objective feedback, statistics should be employed. If the official managers of the game record the data of champions in matches, they can get a dataset about different aspects of champions. These measurements, although undoubtedly helpful, need to be examined closely to ensure that no variables are being discounted [9]. To make a precise judgment about a champion, the dataset used as a reference must be as

DOI: 10.54254/2753-8818/53/20240184

comprehensive as possible, which means it should include the most important data about the performance of a champion such as win rate, pick rate, ban rate, damage, and so on [10]. Through these data, the managers could figure out the overpowered champions easily as those overpowered champions always have abnormal win rates, pick rate, and ban rates. In contrast, as there are champions that need to be weakened, there are champions that need to be strengthened too. They are also quite obvious because of their low win rate and pick rate.

4.2. Balancing matching mechanism

In most online games, players are divided into teams so that players are encouraged to actively communicate with their teammates and practice collaborating skills. Apparently, this mechanism significantly makes the game much more fun since each time players meet different players it makes the game more uncertain. However, the majority of online multiplayer game players often express dissatisfaction with the unfair matching mechanism, which results in one team being significantly more powerful and consequently having a negative impact on the overall gaming experience.

4.2.1. Using statistics to make matching more reasonable. Statistical methods are common and prevalent because they offer theoretical support for matching players. Basically, we could classify players based on their gaming performance so that the players would be divided into different groups, like higher skills, normal skills, and lower skills. And then, the matching mechanism would match players according to their labels, for example, for those 5v5 games such as CSGO, League of Legends, maybe each team would have one higher-skilled player and two normal-skilled players, and two lower-skilled players. Such kind of matching is actually quite reasonable because players with lower personal skills have to gain more support from their teammates while those with high personal skills can carry the team and lead them to victory. In this way, the matching process would be much more fair than matching players completely random.

Although the matching mechanism sounds simple, the most difficult part of building the matching mechanism is to evaluate and sort gamers into groups. Firstly, the designers must find out the most relevant factors in their games, and then collect data on these factors. For instance, the most important factors in FPS games like CSGO are players' killing numbers, death numbers, damage caused per round, and so on. Therefore, the matching program would look at the player's data in these aspects in previous games, and calculate the overall result that defines how great a player is and how many contributions a player makes by specified mathematical algorithms. In particular, if we look at the dataset of CSGO professional players' gaming performance, we can clearly conclude that the most important factors are their personal KDA and KAST (a quantity that reveals the overall contribution to the game), so we can use statistical methods of classification to label players with their levels of skills [11].

Unlike these classic FPS games that apply the same standard to all players, there are lots of games that require different standards for different players. For instance, the typical MOBA game, League of Legends, also matches players into two teams with 5 players on each side, but these 5 players who belong to the same team have completely different jobs since they are assigned different positions such as top, jungle, mid, bottom, and support. Similarly, the most important factors that determine the level of labels of players' skills can be concluded as the number of enemies killed, number of deaths, number of assistance, and so on, What makes this way of distinguishing players different is that every position has different roles. For example, the support player aims at offering help for teammates and offering debuffs for enemies, so the support player may not be able to have a great number of enemies killed, and if the support player takes a large proportion of the number of enemies killed by the whole team, there might be negative impacts because other teammates could not get enough gold from killing, thus unable to make proper contributions to the team. Therefore, the designers should provide different ways of evaluating players' skills and one of the possible solutions is to weigh different factors while calculating the overall impact. If we look at the datasets of LEC professional players of League of

DOI: 10.54254/2753-8818/53/20240184

Legends in 2022 and 2021, we could figure out the difference in data for different positions and create a proper mathematical algorithm to calculate whether a player is skillful or not [12].

4.2.2. Ways to improve the accuracy of statistical classification. The reasons and methods for creating a classification system are easy to understand, but the model we created could only make judgments to some extent, and there is lots of players who are not satisfied with the current matching mechanism. Therefore, it is important to improve the accuracy of classification.

Firstly, it is evident that we can improve the accuracy by offering a greater amount of samples thereby enabling the model to better evaluate players' skills. For example, offering a wider range of perspectives such as win rate could contribute to a more comprehensive assessment of players' abilities [13]. While this approach may prove highly effective when dealing with limited data, its effectiveness may diminish as the overall sample size increases due to overfitting. To address the issue of overfitting and enhance generalization errors, designers may consider incorporating additional dropout layers into the training process. These layers would randomly deactivate nodes, functioning similarly to a filter. Dropout has the effect of making the training process noisy, forcing nodes within a layer to probabilistic take on more or less responsibility for the inputs [14]. Therefore, the overfitting problem could be reduced and the model could give more accurate predictions.

Secondly, designers could improve the quality of the input dataset by data preprocessing. Missing values could lead to improper training results, so we must do data cleaning to remove duplicate records and handle missing values. Then, normalization and standardization are needed to prevent the dominance of certain features over others, due to scale differences, by re-scaling numeric values [15]..

5. Conclusion

Statistical methods cover many aspects of game design and perfection. With the support of statistics, the game development team could figure out the general orientation and attributes of their game by analyzing the market survey results. Furthermore, statistical elements are necessary when training machine learning, as this process always involves a large amount of training and testing datasets. By adding a well-trained model and random elements, the game would be much more joyful for players by enriching their gaming experience, which could also contribute to the longevity of the game. Although designers tried their best to avoid any factor that might negatively affect players' experience, some factors cannot be found until the game has been published. To deal with these problems, the game development team could use statistical tools to evaluate the character to find whether it is excessively powerful. If it surpasses the standard by a significant margin, further measures to weaken it would be necessary. For multiplayer online games, their matching mechanisms have always been criticized by players. In order to improve matching performance, developers should take steps to improve the accuracy and flexibility of their classification system so that the system would grade players more reasonably. This paper suggests several ways to involve statistics in video game development. However, due to a lack of data, the practical evidence has not been included. For further discussion, researchers could find more authoritative sources of data and conduct mock training and testing, or categorize different types of players. This article does not fully explain how machine learning works, but the training model is a complicated process that is worth exploring in depth. Furthermore, the video game market survey result is briefly introduced at the beginning of the article, and each of the general genres has its representative game, so the factors that make the game outstanding are great topics to discuss.

References

[1] Clement, J. (2024) Global video game users 2029, statista. https://www.statista.com/statistics/748044/number-video-gamers-world/

[2] Solod, T. (2024) What's the most popular gaming genre in 2023?, Pingle Studio. https://pinglestudio.com/blog/industry-news/whats-the-most-popular-gaming-genre-in-2023

DOI: 10.54254/2753-8818/53/20240184

[3] Drachen, A., Mirza-Babaei, P., & Nacke, L. E. (Eds.). (2018). Games user research. Oxford University Press.
[4] Xue, L. (2022) 'Application of artificial intelligence in digital games based on mathematical statistics', Mobile Information Systems, 2022, pp. 1–11. doi:10.1155/2022/7145588.
[5] Bailey, E., & Miyata, K. (2019). Improving video game project scope decisions with data: An analysis of achievements and game completion rates. Entertainment Computing, 31, 100299.
[6] Teixeira, C.S. et al. (2019) 'Board games: A tool in the process of microanatomy and embryology education and learning', The FASEB Journal, 33(S1). doi:10.1096/fasebj.2019.33.1_supplement.604.6.
[7] Shen, J., Wu, K. and Liu, Y. (2023) 'Application of big data to common statistical methods based on Game Systems', Asian Journal of Probability and Statistics, 22(4), pp. 49–59. doi: 10.9734/ajpas/2023/v22i4492.
[8] Morschheuser, B., Hassan, L., Werder, K., & Hamari, J. (2018). How to design gamification? A method for engineering gamified software. Information and Software Technology, 95, 219-237.
[9] Becker, A., & Görlich, D. (2020). What is game balancing?-an examination of concepts. ParadigmPlus, 1(1), 22-41.
[10] Cano, R. D. (2022). Stadistics and stats of characters in video games. Universitat Politècnica de Catalunya. Centre de la Imatge i la Tecnologia Multimèdia, Bachelor Thesis, http://hdl.handle.net/2117/394007
[11] Azam. S. (2022) CSGO Pro Players Dataset. Kaggle, https://www.kaggle.com/datasets/sadmadlad/csgo-pro-players-dataset
[12] Pompas. J. (2022) LEC Regular Season 2022/2021/LOL. Kaggle. https://www.kaggle.com/datasets/jordipompas/lec-regular-season-2021
[13] Kumar, A.S. et al. (2024) 'Improving learning-based birdsong classification by utilizing Combined Audio Augmentation Strategies', Ecological Informatics, 82, p. 102699. doi:10.1016/j.ecoinf.2024.102699.
[14] Moradi, R., Berangi, R., & Minaei, B. (2020). A survey of regularization strategies for deep models. Artificial Intelligence Review, 53(6), 3947-3986.
[15] Yin, M., Wortman Vaughan, J., & Wallach, H. (2019, May). Understanding the effect of accuracy on trust in machine learning models. In Proceedings of the 2019 chi conference on human factors in computing systems (pp. 1-12).