Bayes' Theorem: Application in Consecutive Olympic Gymnastics Participant

Yuhan Xiao

Saratoga High School,20300 Herriman Ave, Saratoga, CA 95070
yhxiao82@gmail.com

Abstract
This paper aims to predict the chances of gymnasts' consecutive participation in the Olympic Games with the Bayes Theorem. The concept of conditional probability and the Bayes Theorem will be induced theoretically and demonstrated with idealistic examples of rolling the dice. The theories are subsequently applied along with standard distribution curves, historical statistics of consecutive Olympic Games participation, investigated injury rates among elite athletes, and age distributions of Olympic Games gymnasts to reach our predictive conclusion. Male and female gymnasts are discussed separately due to the differences between the apparatus. Overall, males tend to have a greater chance in consecutively participating in the Olympic Games than females.

Keywords— *Bayes Theorem, Sports Injury, Athlete Aging, Gymnastics*

1. Introduction

As the Olympic Games has just ended, it is reasonable for us to learn about the performances of athletes and look into the future. In this paper, we are focusing on the performances of Olympic gymnasts.

There are over millions of gymnasts around the world. Without any doubt, as one of the major athletic category, gymnastics has always been competitive. Given only around twenty gymnasts-artistic and rhythmic, female and male combined-can make it to each Olympics representing one country, gymnasts struggle to maintain good shape and polished techniques to become the one among the million. What is even harder is to consecutively qualifying for the Olympics.

Gymnasts face obstacles on their road of preparing for the Olympic Games. Injuries on the wrists, hands, feet, and ankles are not only threatening to the gymnasts' personal health, but also influence their participation in World Championships, World Cups, the Olympic Games and more. The injury rate for each gender will hence be utilized and applied in several sections.

Another significant hurdle-even bounded by rule-is age. According to rules set by the Federation Internationale de Gymnastique (FIG), stated on [1], gymnasts have to be at least 16 years old before the end of an Olympic calendar year in order to participate in the Olympics. Besides the lower age bound, there is also the impact on body conditions for relatively older gymnasts, making gymnastics age-sensitive.

Thus, we are going to restrict gymnasts' age qualification based on historical distributions.

2. Background

2.1. Introducing the Bayes Theorem

Throughout history, we have used the term" probability" to define the uncertainty within an event, or, the likeliness for an event to occur [2].

As the probability of an event can be determined either independent on its own or based on other factors, we can further extend the discussion of probability to the more specific concept of conditional probability.

Conditional probability is the probability of an event's occurrence, based on other-especially previous outcomes and evidence. A simple analogy of this might exist in crime investigation scenarios. When you are given a piece of evidence, for instance, a fingerprint, then there is a higher chance that person A is the criminal when his fingerprint matches the given fingerprint better than person B. In other words, his suspicion-probability of being the criminal-is influenced because of the fingerprint.

That is the essence of conditional probability, in short the probability of B given A has already occurred, which

A. El-Hashash et al. (Eds.): IEIT 2022, ACSR 100, pp. 154–159, 2023.
https://doi.org/10.2991/978-94-6463-058-9_28

we can denote as $P(B|A)$, whose arithmetic value can be calculated via

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \tag{1}$$

where $P(A)$ is non-zero and $B \cap A$ stands for the intersection of A and B. Since the intersection of sets are commutative, there is

$$P(B \cap A) = P(A \cap B) \tag{2}$$

Further, using Eqn.(1), Eqn.(2) can be rewritten as

$$P(A)P(B|A) = P(B)P(A|B) \tag{3}$$

which essentially yields,

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} \tag{4}$$

the basic form of Bayes' Theorem. Now, we can build up our understanding and interpretation of Bayes' Theorem with easy-to-understand ideal examples,

2.2. Demonstration of Bayes' Theorem: Rolling Dice

Suppose we are given two fair standard dices, each with six faces representing 1, 2, 3, 4, 5, and 6. Starting out simple, let's roll the two dices at the same instant, and consider P(product of the two numbers is 12).

When we roll the two dices, there are 6*6 = 36 different pairs of outcomes. Among the 36 pairs, there are four pairs: (2,6), (3,4), (4,3), and (6,2) that have a product of 12. Hence, by the basic literal definition of probability-how likely an event can occur-we can use the number of satisfying outcomes divided by the number of total outcomes to calculate the probability, that is $\frac{4}{36}$, further simplified to $\frac{1}{9}$. We can also take a look at the problem by verifying the basic definition of conditional probability stated in Section 1.1 Eqn.(1). Substitute event A with rolling the two dices and event B with getting a product of 12. Since the probability of A and B happening at the same time is $\frac{1}{9}$, and the probability of rolling the two dices is 1 (we are carrying out this action hundred percent sure), the probability of B given A is $\frac{1}{9}$.

Now that we become familiar with the conditional probability through a one-step action, we can step forward and think of a two-step action, again corresponding with the two events in our equations.

Label them as dice X and dice Y, but still consider them standard and fair. What is the probability of P(X yields a 4—product of X and Y is larger than 12)?

Following a similar path from above, the event A here should be product of X and Y is larger than 12, while event B is X giving the number 4. Since the two events are relevant and not independent, conditional probability and Bayes' Theorem must be taken into account. In order to apply the Eqn.(4), we need to derive the value of the three components on the right.

For P(A), we are looking for the probability of the product of X and Y being larger than 12. The new pairs that satisfy this condition are required to:

1. have both elements no less than that of a previously listed pair of product 12

2. have at least one element exceeding the corresponding value of a product-12 pair

(2,6), (3,4), (4,3), and (6,2) have a product of 12, as mentioned earlier. With the two requirements, the only pairs exceeding them are: ($X = 3, Y = 5,6$), ($X = 4, Y = 4,5,6$), ($X = 5, Y = 3,4,5,6$), and ($X = 6, Y = 3,4,5,6$), leading to a total of 13 pairs. Out of the 36 total outcomes, there is a $\frac{13}{36}$ probability that the product is larger than 12.

For P(B), since X is a fair dice and 4 is among the six equally-likely outcomes, the probability should be $\frac{1}{6}$.

Coming to $P(A|B)$, we are given the condition where event B has taken place and X yields a 4. There are 6 possible pairs given this condition, and only (4,4), (4,5), and (4,6) have a product larger than 12.

$P(A|B)$ is easily seen to be $\frac{1}{2}$.

Putting together the three components and plugging in Eqn.(4), we have

$$P(B|A) = \frac{\frac{1}{6} * \frac{13}{36}}{\frac{1}{2}} = \frac{13}{108} \tag{5}$$

which is our answer. Verification of our answer can be completed either through listing all cases or utilizing the conditional probability equation.

Questions might arose around Bayes' Theorem after setting up some simple scenarios and calculations. It might appear as an extraneous theorem which can be totally covered by the basics of conditional probability. Nevertheless, as more variations and errors come into account in real life, Bayes' Theorem have an above expected contribution to predicting results and calculating conditional probabilities.

3. Discussion: Investigating the Olympic Gymnasts

In this application section of the paper, we will assess the influence of age and injury on consecutive participation of Olympic gymnasts.

3.1. Methodology

With collected data from [3] of the 1996 to 2016 Olympics, along with injury rates and age distributions, we will use projected ages and injury rates to estimate consecutive qualification probabilities of

Olympic gymnasts by applying Bayes Theorem. Then we will compare with the past, realistic consecutive Olympic participation rate for further discussion.

3.2. Restricting Factors: Injuries and Age

As mentioned above in the introduction, two of the major factors in preventing the athletes from continuously participating in the Olympic Games are injuries and age. Among which the factor of age has always been a topic of interest, for instance, and been illustrated in [4]. Evaluating these two factors provide us with some knowledge about the basic qualifications of gymnasts.

Before getting into the topic, two pieces of information on gymnasts should be shared. Using artistry gymnastics as our application, male gymnasts practice on six apparatus. Female gymnasts, however, focus on four completely different apparatus, floor exercise, uneven bars, balance beam, and vault. These professional terms are not our topic today, but they yield the important implication. That is, male and female gymnastics should have different body condition requirements, thereby they have different best-fit age range and rate of injuries.

Given the various differences, we should discuss males' and females' consecutive Olympic participation rate separately.

The two factors we are interested in are ages and injuries. Without losing overall generality, we can denote" age suits consecutive Olympic participation" as event X, and" encounter participation-affecting injuries" as event Y.

According to [5], there are 98 Tokyo Olympics male gymnasts. With the list of these gymnasts' information, their age distribution is graphed in the form of a histogram, displayed in figure 1 below.

After applying one-variable statistics, the mean, median, mode, and standard deviation of their

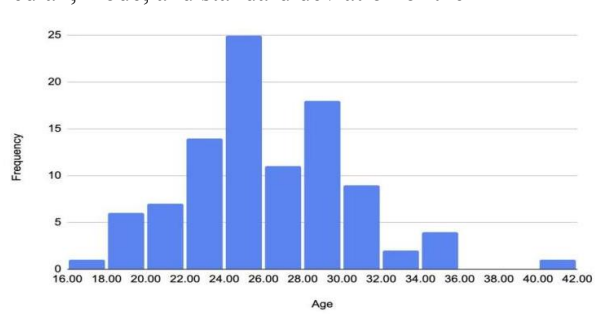

Figure 1: Age distribution among Male Gymnasts from Tokyo Olympics

ages are respectively yielded as 25.69, 25, 25, and 4.14. Given the result in [3] that historical age of male consecutive Olympic participants ranged from 16.8 to 39.5 years old in the 20 years period, we can estimate the probability of event X. Notice that the two endpoints of the range were with respect to each of their Olympic years, so we have to shift everyone's age by 3 to get their age at Paris 2024, because Tokyo 2020 was postponed for a year.

Hence, we are looking for the proportion of male gymnasts who are currently aged 16.0 to 36.5 years old, because the lower bound is restricted to 16. For the purpose of determining a portion, we should use the normal distribution curve, centered at 25.69 with a standard deviation of 4.14. The curve is visualized in figure 2.

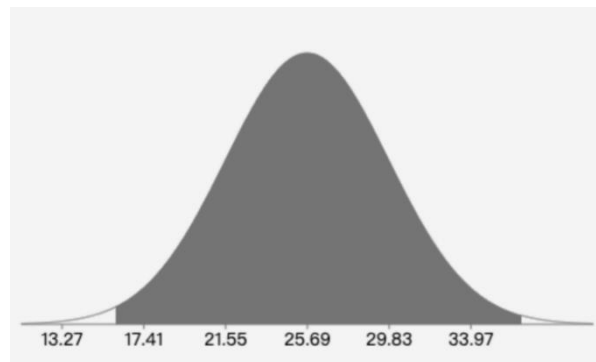

Figure 2: Normal Curve Corresponding with Male Gymnasts' Age Distribution.

The shaded region is what we are trying to estimate, which can be calculated through,

$$P(16.0 <= X <= 36.5) = P(X <= 36.5) - P(X <= 16.0) = 0.9955 - 0.00963 \quad (6)$$

which yields the result that the probability of the Tokyo Olympics male gymnasts to be suitable in age for a consecutive Olympic Games is

$$P(16.0 <= X <= 36.5) = 0.9859 \quad (7)$$

The last element that we need in order to apply the Bayes is P(male consecutive participation). Denote it as event A. In this case, we will obtain the data through [3]. In the article's results, it is given that respectively 277, 104, 28, 6, 1 male gymnasts participated in one, two, three, four, and five Olympic Games. Hence,

$$P(A) = \frac{104 + 28 + 6 + 1}{277 + 104 + 28 + 6 + 1} = 0.3341 \quad (8)$$

Now we can apply equation 4 of the Bayes Theorem to calculate the posterior probability of male gymnasts participating in the Olympic Games consecutively given event X has happened. The data should be incorporated into

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)} \quad (9)$$

which is further calculated as

$$P(A|X) = \frac{(1.0)(0.3341)}{(0.9859)} = 0.3389 \quad (10)$$

Note that P(X|A)'s value takes 1 because we used the full range of historical age when we were estimating the age fit.

Similarly, we can predict female gymnasts' consecutive participation rate based on their age distribution. Referring to [5], there are also 98 Tokyo Olympics female gymnasts, and their age distribution is graphed in a histogram, displayed in figure 3 below.

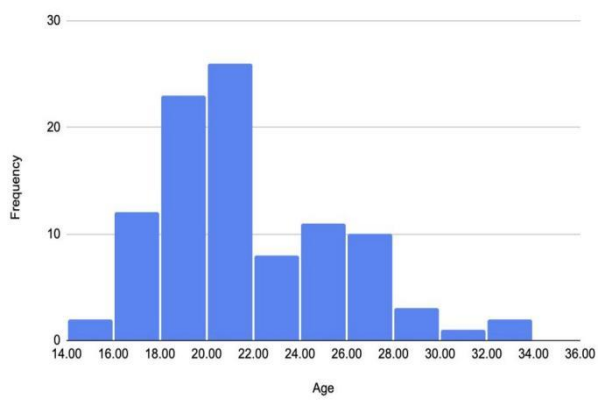

Figure 3: Age distribution among Male Gymnasts from Tokyo Olympics

The mean, median, mode, and standard deviation of the female gymnasts' ages are respectively 21.16, 20.5, 18, and 3.85. Given the result in [3] that historical age of female consecutive Olympic participants ranged from 15.3 to 41.2 years old in the 20 years period, we shall look for the current age range of 15.00 to 38.2 years old (the 15-year-old's are qualified because their birthday is within the calendar year). For a distribution mostly concentrated in the middle, we can use the normal curve, centered at 21.16 with a

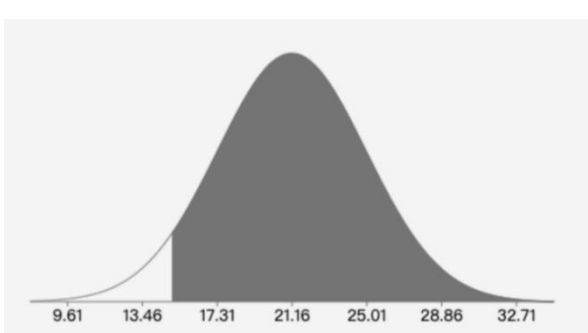

Figure 4: Normal Curve Corresponding with Female Gymnasts' Age Distribution.

standard deviation of 3.85, to estimate the proportion of the sample that fits into the range. Such a curve is visualized in figure 4.

The shaded region is our targeted area, which can be obtained through,

$$P(15.0 <= X <= 38.2) = P(X <= 38.2) - P(X <= 15.0) \quad (11)$$

yielding the result that the probability of the Tokyo Olympics female gymnasts to be suitable in age for a consecutive Olympic Games is

$$P(15.0 <= X <= 38.2) = 0.9452 \quad (12)$$

We need the very last element of P(female consecutive participation) to apply the Bayes Theorem. Again, to hold consistency, denote it as event B. In the article's results, it is given that respectively 408, 70, 11, 0, 1, 1 female gymnasts participated in one, two, three, four, five, and six Olympic Games. Hence,

$$P(B) = \frac{70+11+1+1}{408+70+11+0+1+1} = 0.1690 \quad (13)$$

After applying equation 4 of the Bayes Theorem to

$$P(B|X) = \frac{P(X|B)P(B)}{P(X)}, \quad (14)$$

the prediction for female gymnasts using age restrictions is

$$P(B|X) = \frac{(1.0)(0.1690)}{(0.9452)} = 0.1788 \quad (15)$$

The Bayes Theorem allows us to gain more insight about age's impact on gymnasts of different genders' continuous participation in competitions. Now that we are familiar with the methodology, we can replicate the routine and utilize the theorem similarly on rate of injuries.

Combined given data from [6] as well as [7] are sufficient for us to predict consecutive participation using injuries. The three elements that we need to calculate, by incorporating the condition and the desired result into Eqn.4, are P(no significant injury—consecutively participating in the Olympic), P(no significant injury), and P(consecutive participation), corresponding to notations P(Y—A) and P(Y—B), P(Y), and P(A) and P(B) for males and females.

We will simplify such a model by making the assumption that the difference between P(no injury | consecutively participating in the Olympic) and P(no injury | participating in the Olympic) is negligible. Hence, elite male gymnasts injury rate is used for probability of male injury, while Olympic Games male gymnast injury rate is applied to the conditional probability of getting injured if the gymnast consecutively participates in the Olympic Games. We will find the composite of injury rate for no-injury rate.

The prediction based on historical injury rate of male gymnasts is therefore

$$P(A|Y) = \frac{P(Y|A)P(A)}{P(Y)} = \frac{(0.9912)(0.3341)}{(0.99122)} = 0.3341 \quad (16)$$

and the injury rate of female gymnasts is similarly derived with

$$P(B|Y) = \frac{P(Y|B)P(B)}{P(Y)} = \frac{(0.9909)(0.1690)}{(0.9906)} = 0.1691. \quad (17)$$

Because the three studies for Olympic female gymnasts injury rate vary, we take the middle value of 9.1 to minimize the deviations.

4. Conclusions

Using the Bayes Theorem, we have reached the data of 33.9% as the probability of male gymnasts participating in consecutive Olympic Games, given the athletes' age fits historical range of consecutive participants. This is only around 1.5% away from the independent consecutive participation chance of 33.4%. For female gymnasts, we predict their chance of consecutively participating in Paris as 17.9%, being 5.9% from the independent value of 16.9%. Comparing the two pairs of values, it is reasonable to conclude that the factor of age has a stronger impact on female gymnasts than on male gymnasts. With the same idea, we explored the consecutive participation possibility by filtering injury rate. Similar results come up as 33.4% for male and 16.9% for female. Female gymnasts are again less likely to participate in the Olympic Games consecutively than male gymnasts. Additionally, the probability calculated through Bayesian method only result in small changes from that without the Bayes.

However, there can still be improvements to the methodology design. The first one lies in the Olympic Games we have chosen to perform predictions upon. Rio 2016 might have been a better target as we would have been able to reflect on the predicted values with Tokyo 2020 passing. The second potential improvement is the data collected from different studies. It would have been better if there have been studies on the exact topic that we desire, such as the injury rate among consecutive Olympic Games participants.

Despite all, the takeaway from this predictive process is: the outcome of historical and previous events do play important roles in future incidents. Starting from here, the very basic application in athletic activities, the Bayes Theorem can contribute to our daily lives outstandingly in even larger samples, even broader fields.

References

[1] Wikipedia contributors. Age requirements in gymnastics — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Age_requirements_in_gymnastics&oldid=1038533900, 2021. [Online; accessed 26-August-2021].

[2] Wikipedia contributors. Probability — Wikipedia, the free encyclopedia, 2021. [Online; accessed 10-August-2021].

[3] Sunˇcica Kalinski, Almir Atikovic, and Igor Jelaska. Gender differences in consecutive participation in artistic gymnastics at the olympic games from 1996 to 2016. *Science of Gymnastics Journal*, 10, 03 2018.

[4] Almir Atikovic, Sunˇcica Kalinski, and Ivan ˇcuk. Age trends in artistic gymnastic across world championships and the olympic games from 2003 to 2016. *Science of Gymnastics Journal*, 9:251–263, 10 2017.

[5] Wikipedia contributors. List of gymnasts at the 2020 summer olympics — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=List_of_gymnasts_at_the_2020_Summer_Olympics&oldid=1039803308, 2021. [Online; accessed 26-August-2021].

[6] Robert W Westermann, Molly Giblin, Ashley Vaske, Kylie Grosso, and Brian R Wolf. Evaluation of men's and women's gymnastics injuries: a 10-year observational study. *Sports health*, 7(2):161–165, 2015.

[7] Roger Edmund Thomas and Bennett Charles Thomas. A systematic review of injuries in gymnastics. *The Physician and sportsmedicine*, 47(1):96–121, 2019.