Review Article

The Cricketer’s Performance Analysis in the Key of Mathematical and Statistical Terms

Aftab Alam[1], Ankit Kumar Goyal[2], Chaudhary Arunam Kumar[3]
1. Associate Professor
2. Assistant Professor
3. Research Scholar

Mathematics Department, Swami Vivekanand Subharti University, Meerut U.P

Abstract:
In This Paper is aimed primarily at the selectors of top-level cricket teams. An attempt is made to keep the statistics involve simple as possible so that all levels of selectors may apply this methodology to their teams. the analysis of both batsman and bowler performance outputs has assumed random performance. This assumption needs to be clarified before further progress can be made. The initial thrust of the paper is the identification of what constitutes form. Form can be likened to autocorrelation, in that an individual displays patterns or trends in performance over time. Therefore we will find some of the key mathematical and statistical terms, explaining their significance and how they are applied in the context of cricketer performance analysis.

Key words : Sportsmanship, Aggregate, Autocorrelation, Commentators, Economy Index.

Corresponding Author: Chaudhary Arunam Kumar, Research Scholar, Department of Mathematics, Keral Verma Subharti College of Science, Swami Vivekanand Subharti University, Meerut, UP- 250005

E-Mail : arunamchaudhary2016@gmail.com **Contact:** +91-90270 07194

Introduction

Cricket is a game of numbers. The very core of the sport is entwined with numerical values that translate ultimately to a match result. These sport statistics are a natural by-product of competitive sport and have been around along as contested sport has existed. Currently sport reporters and commentators bombard observers with a vast array of numerical values designed to describe an individual's performance at a particular skill. These added extras contribute to the entertainment value of professional sport. However, is this information of use to coaches and selectors of cricket teams?

Figure No.:01

This paper is aimed primarily at the selectors of top-level cricket teams. An attempt is made to keep the statistics involve simple as possible so that all levels of selectors may apply this methodology to their teams.

There are several key reasons for measuring and evaluating performance in team sport. Organisational Behaviour Theory proves particularly useful in drawing together sport statistics and selection. According to Greenberg and Baron (1997) to build high performance teams appropriate performance measures are required.

General Overview of Cricket Statistics

Cricket statistics are meticulously collated ball by ball. The vocabulary of the game continually refers to abstract statistical concepts such as average, aggregate and form - without divulging the secrets of what these mystical values contain. For the sake of simplicity, all values involved are reduced to one dimension. However, this leaves the cricket observer to assume and speculate as to the base values involved. The written media has recently taken to describing bowling performance by listing the number of wickets taken followed by the bowler's average.

Figure No.:02

This form is limited; the basis behind this statement is discussed later in the evolution of the bowling indices.

In recent times, an increasing number of studies.: have been undertaken to understand the statistical processes at work in the game of cricket. G.H. Wood and W.P. Elderton started the ball rolling in 1945, analysing individual batsmen in an attempt to find a general model that would describe individual scores. This is in accordance with the general trend, where most work to date has revolved around batsmen. This seems like an apparent contradiction, as the first skill taught to junior cricketers is how to bowl, for without bowlers the game cannot be played. However, with advent of one-day cricket and now Cricket Max, both geared towards entertainment, the game is becoming increasingly batsmen orientated. "Batsmen have always received the highest accolades. Most histories of cricket are written around them, with the bolwers regarded merely as a necessary evil." (Nigel Smith, 1994, p.177). The reasoning behind the domination of batsmen in statistical papers may be due to the perceived ease of evaluation.

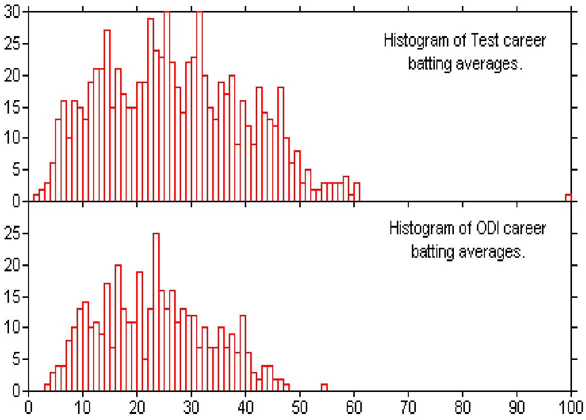

Figure No.:03

This leads to the definition of the statistics utilised in analysis of player performance, enabling a better understanding of the statistics involved.

Sport Statistics can be separated into two broad categories; Performance Indicators and Performance Outputs. A Performance Indicator is a quantitative measure that indicates individual performance in a particular facet of the game. These values are collated during the game in progress. Effectively, the game is dissected into small manageable slices, such that a numerical value can be assigned as a descriptive measure. An example of a Performance Indicator, associated with fielding performance, is Ground Ball Efficiency, defined as the number of times the ball isfielded cleanly divided bt the total number of times fielded. These values do not have a direct impact on the match figures.

Ground ball efficiency =

$$\frac{\text{no.of times the ball is fielded cleanly}}{\text{Total number of times fielded}}.$$

In contrast a Performance Output is a numerical expression detailing the direct result of participation in an event. For cricket these are summary measures detailed in a score book at the completion of an innings, such as score, wickets taken, overs bowled and so forth. As a consequence these values have a direct impact on the match figures-3 It stands to reason that these two categories are related in some manner. However, only performance outputs will be examined in this study, due to the ease of data collection and availability. Investigating the possible relationship between performance indicators and performance outputs will be analysed in future research. Statistically, assessing the performance of a batsman is relatively simple, as this can be given by a single variable, either runs scored in an innings, aggregate, average, or average contribution.

'Aggregate' refers to the total number of runs scored by the individual over a specified period of time. A player's 'Average' is then calculated by dividing the aggregate by the number of times the individual was dismissed during the specified time. 'Contribution' is the percentage of runs the individual provides the team total in an innings. Each value on its own can effectively describe performance.

Ideally, two dimensions need to be considered, one involving the players attacking ability, the other involving the ability to restrict runs. Kimber (1993) gives a graphical method for comparing bowlers. This utilises two dimensions; the attacking ability (strike rate) and the ability to restrict runs (economy rate). Bracewell (1998) proposed two independent normally distributed indices, based upon strike rate and economy rate, to describe performance. The first index deals with a bowler's ability to take wickets, the second with the ability to restrict runs. Both indices are evaluated using simple variations of formula that are already used, taken relative to the team performance.

The section dealing with assessing bowlers relies heavily on these indices. Having defined the performance outputs to be assessed it is necessary to discuss the relevance in a selection situation.

Selection Framework

With the wealth and quality of data available in cricket, it makes sense to utilise this quantitative information in the selection of individuals to maximise the formation of a collective unit (the team). The main assumption underpinning the work in this paper is that a player's natural ability is expressed by individual performance outputs collated following the completion of a match.

Statistics are not the only factors considered when selecting a team. However, Former New Zealand Coach Glenn Turner (1998) discusses the importance of statistics in choosing players in his book Lifting the Covers. In particular the second chapter reveals the emphasis placed on statistics in comparing and selecting individuals. In this instance it is used particularly to justify the non-selection of players, (Andrew Jones and Ken Rutherford) then to defend the selection of Lee Germon.

Since statistics are used to make and confirm selection decisions it is necessary to attempt to understand the nature of the data being generated by participation in sport. A greater understanding leads directly to better implementation and hopefully a competitive edge, for the selected team.

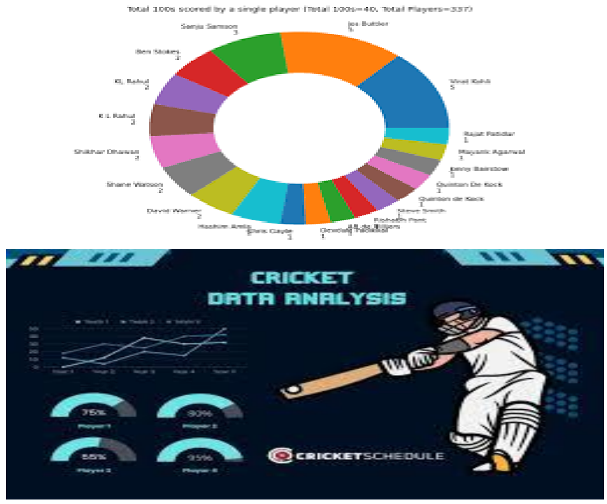

Figure No.:04

Former Australian captain, Richie Benaud, remarked on the simplistic nature of selection and the use of statistics stating, "All acaptain needs is the confidence that his bowlers are each capable of taking five wickets in an innings, his batsmen are capable of scoring a century and that everyone can field like Viv Richards (Benaud, 1995)." Obviously the captain deals with the players on the field and is not responsible for those selected to take the field, this lies in the hands of the selectors. The captain must believe that he has been given the best men to compete It then begomes the job of the selectors to ensure that the best combination of players available takes the field. If statistics are to be used in the selection process they must be meaningful, and secondly they must be used in an appropriate manner. This means that a relevant application of statistical methodology is that of monitoring individual ability.

The Application of Quality Control to Cricket

The idea of monitoring performance is as useful to the selector and the player as it is to the arm chair critic. An ideal method for monitoring an individual's performance is with control charts. The control chart is a useful tool in statistical process control. First developed by W.A Shewhart, the shewhart charts are widely accepted as standard tools for monitoring process of univariate independent and nearly normal measurements (Liu & Tang, 1996) Control charts have found frequent appiications in both manufacturing and non-manufacturing settings (Montgomery, 1997). With slight adjustments shewhart charts can be applied to cricket.

Provided the measurements of the individual's performance are reflective of quality, function, or performance then the nature of the 'thing' being measured has no bearing on the general applicability of control charts (Montgomery, 1997).

Montgomery (1997) discloses several reasons for the popularity of control charts. At least 3 draw direct parallels to cricket. Possibly most important is that control charts provide diagnostic information. This can identify flaws in technique, or the tendency for a player to struggle under certain conditions. Also control charts are proven at improving productivity, which translates to pushing a player and not allowing complacency.

In Cricket we are interested in selecting individuals that will maximise team performa~ce and ensure the best chance of victory. Whilst Cricket is a team sport, the nature of the game allows for individual aspects to stand out. Indeed, when we look at the possible selection of an individual, it is the performance outputs of the individual that is of primary concern. Therefore to ensure the right selections are made, it is important the right statistics are used.

Due to the awkward nature of bowling performance outputs, this leads to the evolution of the bowling indices.

These two independent, random, standard normal indices are a simple and effective way of allowing bowling performance to be measured from the post match statistics. They are more useful than the current convention used in the written media of quoting the number of wickets taken and the bowling average of an individual.

Utilising the assumption that an individual's worth is expressed via performance outputs, this thesis seeks to describe and understand the underlying statistical processes that shape our impression of player performance in the second chapter. Randomness is tested for and then distributional properties of the data are sought.

Major Contributions of this Study

It is expected that two extremes may exist, either form exists, or performance is random. If autocorrelation is present, then form exists. Intuitively performance would be considered random, due to the apparent lack of predictability of such a sport. "Uncertainty plays a large role in sports, and one can argue that the uncertainty associated with sports outcomes is one reason that sports are so popular (Stern, 1997, p19)." It has been shown that baseball is a game of chance (Cook, 19]7). An analysis of team tactics as related to the game of baseball and analysis of the annual World Series competition revealed that results were subject more to the laws of chance than the relative calibre of the competing teams. Taking a simplistic view of competitive sports suggests this may also be the case in cricket (Assuming everyone is equally able to compete, and that natural ability will differ, dependent on the pool of talent available). Logistically it would be ideal if performance is random. If this is the case then it is a relatively simple task to select the best individuals, provided that the sampling distributions to which the data belong are known.

In order to fully understand the summary statistics presented, and make effective use of the available information, the statistical distribution for each of the performance outputs needs to be known. Fulfilment of this requirement and that of randomness satisfies the most important assumptions regarding inference and quality control.

Bowling Performance for Individual Bowlers Of the individual disciplines, bowling is perhaps the hardest to evaluate quantitatively. A typical bowling analysis consists of four variables, Runs conceded. Maidens bowled, Overs bowled and Wickets taken. There is no easy way of interpreting these values independently. History plays a large part of how these statistics are perceived as does the game situation. This section briefly reviews the statistical methods for evaluating an individual's bowling performance.

Kimber (1993) proposed a two-dimensional graphical display for comparing bowlers in cricket based on strike rate (SR) and economy rate (ER), taking advantage of the relationship that these two values have with the Bowling Average (AV). SrxER=100AV These values are traditionally calculated as follows the Economy Rate (ER) is defined as the runs conceded per ball.

However, this relationship does not take into consideration the team situation, and other confounding variables that confront a bowler, such as the state of the game, combined with environmental factors, as these can have an impact on how the specific individual's involved, batsman and bowler, approach each delivery. As a brief example, a batsman 1s more likely to attack the bowler towards the end of the innings, with wickets remaining in a run chasethan a batsman trying to save the match by remaining not out 1n a last wicket partnership when a run chase is no longer viable. Strike Rate has an additional problem, in that if a player fails to take a wicket, a value for SR is not returned as the divisor is zero.

Thus SR is not suitable for evaluation on an innings by innings basis. This measurn could be calculated using all the match results for a season, but in terms of selection and monitoring a player's performance, it is too late to address an individual's worth at the end of the season. Thus only players who have taken wickets can have strike rate as a performance measure.

Bracewell (1998) detailed a novel way to evaluate individual bowling performance, incorporating SR and ER into two separate indices that considered relative performance to the team. This involved an attempt to form ratio's that took into account an individual 's performance in relation to the team performance. The Attack Ratio involved inverting SR for both team and individual so that wickets taken was no longer the denominator.

However it was found that as the number of overs bowled by an individual increased, the score for both indices tended to zero. This was because as a player bowls more and more overs (approaches 50%) this player is having a huge influence on the team performance. His performance therefore reflects the team performance very closely.

The final evolution of performance measures for bowlers involved multiplying the ratio's by a weighting factor related to time (overs). The problem described earlier was removed in this way.

In addition it was found the Attack index needed to be multiplied by a wicket weighting factor, defined in terms of w, the number of wickets taken in any innings. This index is therefore innings specific whereas the other measures are more general .

Investigation of Bowling Measures

Of the statistical analyses performed using cricket data, bowling is an area deficient in research. Only Kimber (1993) and Bracewell (1998) have examined how to measure an individual's bowling performance. Kimber sought to do this via a graphical display based on Strike Rate and Economy Rate, whereas Bracewell tried extending these values relative to the team.

Randomness

Very little research has been done on the aspect of randomness in an individual's performance in cricket. A distantly related team sport, baseball, was found to be essentially random (Cook 1977). There is anecdotal evidence supporting the claim that the role of an individual within a game is random, generally commenting on the apparent lack of predictability of cricket. Berkmann, (1990), Brittenden (1994) and Turner (1998) are just a small selection of cricket observers that subscribe to the unpredictability of cricket view Hunting through player biographies also reveals that those who play the game express this view

Danaher (1989) applied a Run's test to 6 English County Cricketers and found that none showed a significant runs pattern at the 5% significance level. The batsmen chosen were of varying batting ability but chosen because they were either top, or close to the top, of their team's batting averages list.

Kumar (1996) suggested that cricket is not by chance. However, this assertion was based solely upon over run rates in one-day cricket. The implications of this are manifested in the troublesome interrupted match rules. If over rates were random, then the simple Average Run Rate (ARR) rule would suffice, as this is based on the assumption that run rate of the batting side does not change during the innings. Instead, the resources available to a team play an important role in determining the outcome of a one-day match. One only needs to look to the Duckwort~ Lewis model (1996) to see the effect that time (overs in hand) and wickets in hand have in determining a batting side's capacity for team total. Team strategies also illustrate this point. As a simple illustration of batting capacity, this model accepts the fact that a side is more capable (or daring) of scoring runs when only 2 wickets have been lost, as opposed to being 8 down, with 10 overs remaining. The reasoning behind this is; with the loss of only 2 wickets, presumably the better batsmen are still available, and there are plenty of individual's remaining. Thus batsmen are more able to go after their shots, as the consequences to the team of their dismissal are not as great.

Whereas, a batting side with 8 wickets down needs to adopt a more cautious approach , as once a team is dismissed, there is no further chance of adding to the team total.

Distribution of Performance Measure

Pollard (1977) conceded "that a more elaborate model needs to be developed to describe the distribution of batsman's scores " This was due to the fact that previous results did not cater for the higher than expected frequencies of failures to score, compared to the theoretical models.

Bracewell (2)(1998) suggested a discrete version of a mixed exponential distribution for score and a relatively new concept in cricket statistics, contribution, based upon 5609 observations of individuals in the top 6 of the batting order from New Zealand domestic first class cricket. This involved separating the occurrence of zero and recalculating the mean to find the parameters of the distribution involving the non-zero values.

Economy Index As indicated in 2.1 0 Bracewell 's (2)(1998) economy index is:

ECONOMY INDEX= [ECONOMY RATIO x /OVERS]

Where Economy Ratio has been defined previously and overs is the number of overs bowled by the individual in the innings.

Standardising the above equation by first subtracting the mean (0.3657) and then dividing by the standard deviation (3.2016) gives a value to comparable to the standardised attack index. A 95% Confidence interval for the mean reveals that the population mean probably lies between 0.3851 and 0.3463. As zero is not contained within this confidence interval the value for the mean can not be ignored and must be included in the standardisation.

Similarly a 95% confidence interval for the population standard deviation shows that it probably falls between 3.2001 and 3.2035. Due to 1 not falling in this interval, the value given for standard deviation needs to be used to standardise the economy index.

If the probability distribution of a population from which the sample is gathered is known , then the probability distribution of the various statistics computed from the sample data can be determined (Montgomery, 1997). More importantly, it can be established what a player is expected to score and thus their progress can be monitored, which is especially relevant for team selection.

A population is a set of measurements that can be described by a set of numerical measures called parameters (Ott, Mendenhall , 1985). In most applications of statistics the parameters are not known but inferences about them are made using information contained in a sample.

For time series analysis it is assumed that for each time point t, Z1 is a random variable. Thus the behaviour of Zt will be determined by a probability distribution (Cryer). In this instance time t, refers to each innings and Z1 refers to a performance output. Previous studies have assumed that the data are independent, in that for each individual bowler the previous match result does not have a direct impact on the following match result. At first class level this is a safe assumption as it is presumed that players who reach this level have developed the necessary mental skills.

According to the analyses performed individual batting scores are best modelled by a mixed geometric distribution, mixed in two parts, the zero portion and non-zero portion. Batting contribution is best modelled by the negative binomial distribution (with r set equal to 1). The zero component of this distribution represents the zero portion of the score distribution.

It is important to note that the negative binomial distribution and the parameter from the contribution distribution, model the occurrence of zero amongst individual scores. This is an interesting phenomenon.

Obviously either score or contribution has to be mixed as both share the same number of zeros; unless the team continually scores exactly 100 in each innings, in which case the two distributions will be equivalent. That is the score is effectively the percentage contribution, as score is continually divided by 100 runs (the team total). Considering the case of scoring 0, the probability of this occurring is the same for both contribution and score. This is because for contribution O/y = 0, where y is the team score. Due to the sample mean representing the shape parameter, there is a difference between the shapes for the score and contribution distributions.

Normality Test Bracewell (3) (1998) hypothesized the bowling indices for individuals are normally distributed. To test this hypothesis a normality test needs to be performed. The normality test for the bowling indices involved the generation of a normal probability plot. The probability for the x-values (index) is calculated then plotted against a standard normal probability score. A least-squares line is fitted to the points. This forms an estimate for the cumulative distribution function from which the data for the population is drawn. The Anderson-Darling test for normality is used, which is an ECDF (empirical cumulative distribution function) based test.

Obviously either score or contribution has to be mixed as both share the same number of zeros; unless the team continually scores exactly 100 in each innings, in which case the two distributions will be equivalent. That is the score is effectively the percentage contribution, as score is continually divided by 100 runs (the team total). Considering the case of scoring 0, the probability of this occurring is the same for both contribution and score. This is because for contribution O/y = 0, where y is the team score. Due to the sample mean representing the shape parameter, there is a difference between the shapes for the score and contribution distributions.

Of the individual disciplines, bowling is perhaps the hardest to evaluate quantitatively. A typical bowling analysis consists of four variables, Runs conceded. Maidens bowled, Overs bowled and Wickets taken. There is no easy way of interpreting these values independently. History plays a large part of how these statistics are perceived as does the game situation. This section briefly reviews the statistical methods for evaluating an individual's bowling performance.

Kimber (1993) proposed a two-dimensional graphical display for comparing bowlers in cricket based on strike rate (SR) and economy rate (ER), taking advantage of the relationship that these two values have with the Bowling Average (AV).

SrxER=100AV

These values are traditionally calculated as follows the Economy Rate (ER) is defined as the runs conceded per ball.

$$\text{Economy rate} = \frac{\text{Total runs conceded}}{\text{Total balls faced}}$$

However, this relationship does not take into consideration the team situation, and other confounding variables that confront a bowler, such as the state of the game, combined with environmental factors, as these can have an impact on how the specific individual's involved, batsman and bowler, approach each delivery. As a brief example, a batsman 1s more likely to attack the bowler towards the end of the innings, with wickets remaining in a run chase, than a batsman trying to save the match by remaining not out 1n a last wicket partnership when a run chase is no longer viable. Strike Rate has an additional problem, in that if a player fails to take a wicket, a value for SR is not returned as the divisor is zero

Thus SR is not suitable for evaluation on an innings by innings basis. This measurn could be calculated using all the match results for a season, but in terms of selection and monitoring a player's performance, it is too late to address an individual's worth at the end of the season. Thus only players who have taken wickets can have strike rate as a performance measure.

Bracewell (1998) detailed a novel way to evaluate individual bowling performance, incorporating SR and ER into two separate indices that considered relative performance to the team. This involved an attempt to form ratio's that took into account an individual 's performance in relation to the team performance. The Attack Ratio involved inverting SR for both team and individual so that wickets taken was no longer the denominator. The ratios are defined as follows

Economy Ratio= (Opposition Total I Total Overs - Runs Conceded I Overs] Attack Ratio = [Wickets/Overs - Total wickets/Total Over's]

DISTRIBUTION FITTING

However it was found that as the number of overs bowled by an individual increased, the score for both indices tended to zero. This was because as a player bowls more and more overs (approaches 50%) this player is having a huge influence on the team performance. His performance therefore reflects the team performance very closely.

The final evolution of performance measures for bowlers involved multiplying the ratio's by a weighting factor related to time (overs). The problem described earlier was removed in this way

In addition it was found the Attack index needed to be multiplied by a wicket weighting factor, defined in terms of w, the number of wickets taken in any innings. This index is therefore innings specific whereas the other measures are more general .

The wicket weighting factor in the Attack Index is given by p(w), the probability of taking a certain number of wickets in an innings. Standardisation allows the indices to be compared on similar scales.

MEASURES OF CENTRAL TENDENCY AND DISPERSION

These are the fundamental building blocks for understanding any dataset.

Mean (μ or x¯):

Detail: The most common average. You sum all the values in a dataset and divide by the number of values.

Application in Cricket:

Batting Average: As discussed, this is the mean runs scored per dismissal.

- **Mean Bowling Speed:** Average speed of a bowler's deliveries.
- **Average Catches per Match:** For a fielder.

o **Why it matters:** Gives a quick snapshot of "typical" performance. However, it can be heavily influenced by outliers (e.g., a single very high score by a batsman can inflate their average if they haven't played many innings).

- **Median:**
  - **Detail:** The middle value when data is ordered. If there's an even number of values, it's the average of the two middle values.
    - **Application in Cricket:**
    - **Median Score:** If a batsman has scores of 10, 20, 50, 70, 100, the median is 50. If their scores were 0, 0, 5, 100, 150, the mean might be high but the median reveals a different story about consistency.

o **Why it matters:** It's robust to outliers. For highly skewed performance data (e.g., a bowler who takes many cheap wickets but sometimes gets hit for many runs), the median economy rate might give a more stable picture of typical performance than the mean.

- **Mode:**
  - **Detail:** The most frequently occurring value.
  - **Application in Cricket:**
    - **Most Frequent Dismissal Type:** For a batsman (e.g., "caught" might be their mode of dismissal).
    - **Most Common Score Interval:** What score range does a batsman most often fall into (e.g., 20-30 runs).
  - **Why it matters:** Useful for identifying common patterns or habits.
- **Standard Deviation (σ or s):**
  - **Detail:** It quantifies the amount of variation or dispersion of a set of data values around the mean. A low standard deviation indicates that the data points tend to be close to the mean, while a high standard deviation indicates that the data points are spread out over a wider range of values.
  - **Formula Intuition:** It's the square root of the average of the squared differences from the mean. Squaring the differences makes sure positive and negative deviations don't cancel out, and gives more weight to larger deviations. The square root brings it back to the original units.
  - **Application in Cricket:**
    - **Batting Consistency:** A batsman with an average of 40 and a standard deviation of 5 is very consistent. Another batsman with an average of 40 and a standard deviation of 20 is inconsistent (scores range widely).
    - **Bowling Line & Length:** A bowler with a low standard deviation in their delivery's landing spot is highly accurate.
    - **Speed Variation:** For a fast bowler, the standard deviation of their speed can indicate how much they vary their pace.
  - **Why it matters:** Provides crucial insight into a player's reliability and predictability. A player with a high mean but also high standard deviation might be a "match-winner" on their day, but also prone to failures.
- **Coefficient of Variation (CV):**

o **Detail:** It expresses the standard deviation as a percentage of the mean. This is unitless, making it ideal for comparing consistency across different metrics or players.

- **Formula:**
CV=(Standard Deviation/Mean)×100%
- **Application in Cricket:**
- **Comparing Consistency:** Is a batsman with an average of 50 and σ=10 more consistent than a bowler with an economy of 4.0 and σ=0.5? You can compare their CVs. The batsman's CV = (10/50)*100 = 20%. The bowler's CV = (0.5/4.0)*100 = 12.5%. The bowler is relatively more consistent in their economy.
- **Why it matters:** Allows for fair comparisons of variability when the means are very different, or when the units are different.

PROBABILITY AND DISTRIBUTIONS

Understanding the likelihood of events is fundamental to strategy.

- **Probability (P(E)):**
- **Detail:** A numerical measure of the likelihood that an event will occur, ranging from 0 (impossible) to 1 (certain).
- **Application in Cricket:**
- P (wicket on next ball): Based on bowler form, pitch, batsman.
- P(boundary on next ball): Based on batsman form, field setting.
- P (win | score X runs): The probability of winning a match given a certain score.
- **Why it matters:** Informs risk assessment, decision-making (e.g., whether to go for a risky shot or play safe), and strategic planning.
- **Probability Distributions:**
- **Normal Distribution:**
- **Detail:** The classic "bell curve." Many natural phenomena follow this distribution (e.g., heights, IQ scores). Characterized by its mean and standard deviation. Data points are symmetrically distributed around the mean.
- **Application in Cricket:**
- **Bowling Speed:** While individual balls might vary, a bowler's average speed over many deliveries often approximates a normal distribution.
- **Run Rate Distribution:** The distribution of run rates per over across a long period might show a normal distribution around the average.
- **Why it matters:** If data is normally distributed, we can make strong inferences. For example, roughly 68% of data falls within one standard deviation of the mean, 95% within two, and 99.7% within three. This helps predict ranges of performance.
- **Poisson Distribution:**
- **Detail:** Describes the number of events occurring in a fixed interval of time or space, given a known average rate of occurrence and that these events occur independently.
- **Application in Cricket:**
- **Number of Wickets in an Innings:** Given a team's average wicket-taking rate, you can model the probability of taking 0, 1, 2, etc., wickets in a particular spell or innings.
- **Number of Boundaries in an Over:** Predicting the likelihood of hitting a certain number of fours or sixes in an over.
- **Why it matters:** Useful for modeling discrete event counts, especially for rare events or events occurring at a constant average rate.
- **Binomial Distribution:**
- **Detail:** Models the number of "successes" in a fixed number of independent "trials," where each trial has only two possible outcomes (success/failure) and the probability of success is constant.
- **Application in Cricket:**
- **Catches Taken:** Out of 5 catching opportunities, what's the probability a fielder takes 3 catches (given their historical success rate)?
- **Dot Balls Bowled:** Out of 6 balls in an over, what's the probability a bowler bowls 4 dot balls?
- **Why it matters:** Ideal for situations with a fixed number of trials and two outcomes, helping to predict success rates for specific actions.

Inferential Statistics

This is where we move from describing what happened to making predictions and drawing conclusions.

- **Hypothesis Testing:**
- **Detail:** A formal procedure to decide whether to accept or reject a claim (the null hypothesis, H0) about a population based on sample data. You formulate H0 and an alternative hypothesis (H1). You then collect data and calculate a test statistic, which gives you a p-value.
- **p-value:** The probability of observing data as extreme as (or more extreme than) what you got, *assuming the null hypothesis is true*.
- If $p<\alpha$ (a pre-determined significance level, usually 0.05), you reject H0. This suggests the observed effect is statistically significant and likely not due to random chance.
- If $p\geq\alpha$, you fail to reject H0. This means there isn't enough evidence to conclude a significant effect.
- **Application in Cricket:**
- **Comparing Performance:** Is Player A's average truly higher than Player B's, or is the difference just due to random variation? (H0: A's average = B's average; H1: A's average □= B's average).
- **Impact of a New Batting Technique:** Does a new technique significantly improve a batsman's strike rate? (H0: Strike rate unchanged; H1: Strike rate increased).
- **Pitch Effect:** Does playing on a particular pitch type significantly reduce a bowler's economy?
- **Why it matters:** Provides a rigorous framework for making data-driven decisions and avoiding false conclusions based on small sample sizes or random fluctuations.
- **Correlation Coefficient (r):**
- **Detail:** A numerical value between -1 and +1 that indicates the strength and direction of a *linear* relationship between two quantitative variables.
- r=1: Perfect positive linear relationship (as one increases, the other increases proportionally).
- r=−1: Perfect negative linear relationship (as one increases, the other decreases proportionally).

▪ r=0: No linear relationship.

○ **Application in Cricket:**

▪ **Runs and Balls Faced:** Strong positive correlation for batsmen.

▪ **Bowling Speed and Wickets:** Is there a correlation between higher bowling speeds and more wickets? (Not always a perfect linear relationship, but interesting to explore).

▪ **Dot Ball % and Economy Rate (bowlers):** Strong negative correlation (more dot balls usually means better economy).

○ **Why it matters:** Helps identify potential relationships between different performance metrics, which can inform coaching decisions or strategic insights. *Important note: Correlation does not imply causation!*

- **Regression Analysis (especially Linear Regression):**

○ **Detail:** A powerful statistical method used to model the relationship between a dependent variable (what you want to predict) and one or more independent variables (predictors). Linear regression models this relationship with a straight line.

○ **Equation (Simple Linear Regression):** $Y=\beta 0+\beta 1X+\epsilon$

▪ Y: Dependent variable (e.g., Runs Scored)

▪ X: Independent variable (e.g., Balls Faced)

▪ β0: Y-intercept (the value of Y when X is 0)

▪ β1: Slope (the change in Y for a one-unit change in X)

▪ ϵ: Error term (the unpredictable part)

○ **Application in Cricket:**

▪ **Predicting Runs:** Predict a batsman's runs based on the number of balls they faced, or the specific overs they played.

▪ **Impact of Boundary Hitting:** Model how boundary percentage impacts a team's total score.

▪ **Predicting Wickets:** Predicting wickets taken based on line, length, and speed metrics.

○ **Why it matters:** Enables prediction and helps understand the *magnitude* of the influence of one variable on another. It can help answer "how much" questions.

Advanced Concepts (often incorporating Machine Learning)

- **Optimization Algorithms:**

○ **Detail:** Mathematical procedures used to find the best possible solution from a set of alternatives, often by maximizing or minimizing some objective function subject to certain constraints.

○ **Application in Cricket:**

▪ **Optimal Batting Order:** Given player strengths and weaknesses against different bowling types, and match situations, what's the optimal batting order to maximize expected runs?

▪ **Field Placement Optimization:** Where should fielders be placed to minimize runs conceded and maximize wicket chances against a specific batsman on a specific pitch?

▪ **Player Workload Management:** Optimizing training schedules to maximize performance and minimize injury risk.

○ **Why it matters:** Moves beyond analysis to prescriptive solutions, helping teams make the best decisions for a given objective.

- **Time Series Analysis:**

○ **Detail:** A set of techniques for analyzing data points collected over time. It looks for trends, seasonality, cycles, and irregular components.

○ **Application in Cricket:**

▪ **Player Form Tracking:** Monitoring a player's batting average or economy rate over consecutive matches to identify dips or improvements in form.

▪ **Forecasting Performance:** Predicting a player's likely performance in upcoming matches based on their recent form and historical trends.

▪ **Identifying Fatigue:** Analyzing workload metrics (e.g., GPS data, bowling spells) over time to detect patterns that might indicate fatigue or injury risk.

○ **Why it matters:** Recognizes that performance isn't static and evolves over time. It helps understand dynamic changes and forecast future states.

By applying these mathematical terms and techniques, cricket analysts can move beyond simple statistics to build sophisticated models that provide deep insights into individual player performance, inform strategic decisions, and ultimately, enhance a team's chances of success.

Conclusion

In conclusion, cricket is more than just a sport; it's a way of life. As the "Gentleman's Game," cricket celebrates values of sportsmanship, companionship, and fair play that transcend the boundaries of the field. In this conclusion from this dessertation I have taken a keen view to different aspects of the games.this dessertation has helped me to learn the sport cricket with more seriousness and makes it more interesting

I have learned various formulas through which cricket is carried through and the whole games works.through this research I have came to known that cricket is not just a game it's a full subject to read.

Source of Support: Nil

Conflict of interest: Nil

Acknowledgement: None

References:

1. Agresti, A. (2015):"Foundations of Linear and Generalized Linear Models", Wiley. Bernardo, J. M., & Smith, A. F. M. (2000):"Bayesian Theory", Wiley.
2. Boyd, S., & Vandenberghe, L. (2004):"Convex Optimization", Cambridge University Press.
3. Durrett, R. (2019): "Probability: Theory and Examples (5th ed.)" , Cambridge University Press.
4. Lehmann, E. L., & Casella, G. (1998):"Theory of Point Estimation (2nd ed.)", Springer.

5. Lehmann, E. L., & Romano, J. P. (2005):“Testing Statistical Hypotheses (3rd ed.)”,Springer.
6. Norris, J. R. (1997):“Markov Chains”, Cambridge University Press.
7. Osborne, M. J. (2004):“An Introduction to Game Theory”, Oxford University Press.
8. Leamon, N., & Jones, B. (2019):“*Hitting Against the Spin: How Cricket Really Works”*Polaris Publishing.
9. Dalal, D., & Sharma, M. (2024):“Cricket Match Analytics and Prediction using Machine Learning”,*International Journal of Computer Applications, 186*(26).

How to cite this article: Alam A, Goyal AK, Arunam C. The cricketers performance analysis in the key of mathematical and statistical terms. Subharti J of Interdisciplinary Research, Dec. 2025; Vol. 7: Issue 3, 5-13