**MODULE: 01**

# **“Probability Distribution”**

**Course: ALY6050 Intro to Enterprise Analytics**

**Instructor: Professor Adam Jones**

**Winter 2025 – February 26**

**College of Professional Studies**

**A black background with a black square

Description automatically generated with medium confidence**

Content

1. Introduction
2. Problem Statement
3. Overview
4. Parameter
5. Functions to Calculate Probabilities
6. Simulate Probabilities
7. Calculate Risk Metrics
8. Chi-squared Goodness of fit
9. Visualizations
10. Comparison of Risk Metrics
11. Conclusion

**Introduction to Assignment:**

Overview and Rationale

This assignment applies probability theory to a sports betting scenario involving a best-of-three series between two Major League Baseball teams, the Boston Red Sox and the New York Yankees. The objective is to analyze a betting strategy by calculating the probability of the Red Sox winning the series and the corresponding net winnings. This analysis aims to provide insights into how probability theory can be leveraged to make informed betting decisions and evaluate the expected financial outcome of a given strategy.

Betting in sports inherently involves uncertainty and risk, which makes it an ideal scenario for applying probability theory. Understanding the likelihood of different outcomes allows for better decision-making and risk assessment. Additionally, simulating the series multiple times enables us to observe the distribution of possible outcomes, estimate the expected net winnings, and verify the theoretical calculations through statistical tests.

**Problem Statement:**

The Boston Red Sox and New York Yankees are scheduled to play a best-of-three series, where the first team to win two games emerges as the winner. The series can follow two different sequences:

1. Game 1 in Boston, Game 2 in New York, and Game 3 (if necessary) in Boston.
2. Game 1 in New York, Game 2 in Boston, and Game 3 (if necessary) in New York.

The probabilities for each team winning a game are as follows:

* The probability of the Red Sox winning at home (Boston) is 0.6.
* The probability of the Yankees winning at home (New York) is 0.57.

The betting scenario involves the following stakes:

* If the Red Sox win a game, the bettor wins $500.
* If the Red Sox lose a game, the bettor loses $520.

The outcomes of the games are assumed to be independent of each other.

**Objective of the Assignment:**

The assignment is divided into three parts:

1. **Part 1:** Analyze the series with the sequence: Boston, New York, and Boston (if necessary).
2. **Part 2:** Repeat the analysis with the sequence: New York, Boston, and New York (if necessary).
3. **Part 3:** Extend the analysis to a best-of-five series with alternating games, starting in Boston.

The analysis aims to:

* Calculate the probability of the Red Sox winning the series.
* Construct the theoretical probability distribution for net winnings.
* Estimate the expected net winnings using simulations.
* Perform statistical tests to compare simulated and theoretical distributions.
* Evaluate the profitability of the betting strategy.

**Tasks to Solve**

For each part, the following tasks will be completed:

1. **Calculate the Probability of Winning the Series:**
   * Identify all the possible ways the Red Sox can win the series.
   * Calculate the probability of each scenario.
   * Sum up the probabilities to get the total probability of winning the series.
2. **Construct the Theoretical Probability Distribution:**
   * Identify all possible net winnings.
   * Calculate the probability associated with each net winning.
   * Compute the **Expected Value (E(X))** and **Standard Deviation (SD(X))** of the net winnings.
3. **Simulation of Series Outcomes:**
   * Simulate **10,000 series** using Python to generate random outcomes for each game.
   * Calculate the net winnings for each series.
   * Estimate the expected net winnings using a **95% confidence interval**.
4. **Frequency Distribution and Chi-Squared Goodness of Fit Test:**
   * Construct a frequency distribution for the simulated net winnings.
   * Perform a **Chi-squared goodness of fit test** to compare the simulated distribution with the theoretical distribution.
5. **Evaluation of Betting Strategy:**
   * Compare the expected value from the theoretical distribution with the mean of the simulated outcomes.
   * Analyze whether the confidence interval contains the theoretical expected value.
   * Assess the risk and profitability of the betting strategy.

**Approach and Methodology**

The analysis has been conducted using **Python** in **Jupyter Notebook** with the following approach:

1. **Theoretical Calculations:**
   * Utilize probability theory to calculate the likelihood of each scenario.
   * Calculate the expected value and standard deviation of the net winnings.
   * Use these calculations to construct the theoretical probability distribution.
2. **Simulation:**
   * Leverage Python's numpy and random modules to simulate 10,000 series.
   * Generate random outcomes for each game according to the given probabilities.
   * Calculate the net winnings for each series and compile the results.
3. **Statistical Analysis:**
   * Estimate the expected net winnings from the simulations.
   * Construct a **95% confidence interval** to assess the precision of the estimate.
   * Perform a **Chi-squared goodness of fit test** using scipy.stats to compare the simulated distribution with the theoretical distribution.
4. **Visualization:**
   * Use matplotlib to create histograms for the distribution of net winnings.
   * Visualize the confidence interval and compare simulated vs. theoretical distributions.

**Tasks**

1. **Import Libraries**
   * 1. **NumPy (np)**: Used for numerical computing and array operations.
     2. **Matplotlib (plt)**: Used for creating visualizations and plots.
     3. **SciPy Stats (stats)**: Provides statistical functions and distributions.
     4. Ensures reproducibility by generating the same sequence of random numbers each time the code is run.
2. **Parameters**
   1. **p\_B\_win = 0.6**: The probability of the Boston Red Sox winning a game at their home stadium is 60%.
   2. **p\_NY\_win = 0.57**: The probability of the New York Yankees winning a game at their home stadium is 57%.
   3. **win\_amount = 500**: The amount gained for a winning bet.
   4. **lose\_amount = 520**: The amount lost for a losing bet. This is slightly higher than the win amount, likely reflecting a betting fee or vigorish.
   5. **n\_simulations = 10000**: The number of simulations to run. A high number ensures reliable statistical results.
3. **Function To Calculate Probabilities for each game** 
   1. Purpose: Calculates the win probability for each game in a series, based on the game location.
   2. Parameters:

game\_order: A list representing the sequence of games, where "B" indicates a game in Boston, and "NY" indicates a game in New York.

* 1. Logic:

If the game is in Boston ("B"), the probability of winning is p\_B\_win (0.6).

If the game is in New York ("NY"), the probability of losing is calculated as 1 - p\_NY\_win (0.43), since the Yankees' home win probability is 0.57.

* 1. Output: Returns a list of win probabilities corresponding to the sequence of games.

1. **Function to Simulate Series:**

Purpose:

This function simulates a playoff series between the Boston Red Sox and the New York Yankees multiple times to calculate the financial outcomes of betting on the Red Sox.

Key Details:

* game\_order: A list specifying the sequence of games, e.g., ["B", "NY", "B", "NY", "B"], where "B" is a game in Boston, and "NY" is in New York.
* wins\_needed: The number of games the Red Sox need to win to secure the series.

How It Works:

1. Simulate Games:
   * For each game in the series, the outcome ('W' for win or 'L' for loss) is simulated using the respective home team's win probability.
   * If the game is in Boston ('B'), the probability of a Red Sox win is p\_B\_win.
   * If the game is in New York ('NY'), the probability of a Red Sox win is 1 - p\_NY\_win.
2. Determine Series Outcome:
   * Count the number of games won by the Red Sox.
   * If the Red Sox wins are greater than or equal to wins\_needed, the series is considered won, and the profit is calculated as wins\_needed \* win\_amount.
   * Otherwise, the loss is calculated as -wins\_needed \* lose\_amount.
3. Repeat and Collect Results:
   * The series is simulated n\_simulations times to generate a distribution of possible outcomes.
   * The function returns an array of these outcomes, enabling further analysis of the financial risk and reward.
4. **Simulate All Three Parts**

The simulations of the three series scenarios show a mix of profits and losses. In Part 1 and Part 2 (3-game series), most outcomes are losses (-1040), indicating a higher risk, but some simulations yield a profit (1000) when the Red Sox win the required number of games. In Part 3 (a 5-game series), the potential for higher profits (1500) exists, but so does the risk of larger losses (-1560), as a Red Sox win of 3 games is needed. Overall, these results highlight the volatility and risk of betting in sports, with the potential for both significant rewards and losses depending on the series structure.

A screenshot of a computer

AI-generated content may be incorrect.

1. **Calculate Risk Metrics for Part 1, Part 2, Part 3**

**A screenshot of a computer screen

AI-generated content may be incorrect.**

* **Part 1** offers a **positive expected value** ($124.84 profit), but there is still a substantial **risk** (42.90% chance of loss). The **high coefficient of variation (8.09)** indicates a high risk relative to the potential reward.
* **Part 2** results in a **negative expected value** (-$61.41 loss), with a **higher probability of loss (52.03%)** and an even greater risk relative to the expected outcome (CV of 16.60). This suggests betting on this scenario is likely more volatile and less favorable.
* **Part 3** provides a **positive expected value** ($147.17 profit), but with even **higher risk** due to the longer series and greater variability (CV of 10.33). There's still a notable chance of losing money, although the potential rewards are higher.

**To conclude, metrics:**

* **Part 1** is the least risky with the potential for profit, while **Part 2** is the least favorable with a negative expected value.
* **Part 3** offers the greatest potential reward but comes with a higher risk due to the longer series length.

Betting on any of these series presents **substantial risk**, but the longer series (Part 3) could yield better returns if the Red Sox perform well.

1. **Chi-squared Goodness of fit test**

The **Chi-squared goodness of fit test** is used to assess whether observed frequencies (in this case, outcomes from simulations) differ significantly from expected frequencies. The key components of this test are:

**Chi-squared Statistic**: This is a measure of how much the observed data deviates from the expected data. The higher the value, the more the observed data differs from the expected distribution.

**p-value**: This tells us the probability of observing the data, or something more extreme, if the null hypothesis were true. A smaller p-value indicates stronger evidence against the null hypothesis (which typically states that the observed and expected distributions are the same).

A screenshot of a computer

AI-generated content may be incorrect.

**Part 1: B, NY, B (3-game series with 2 home games in Boston)**

**Chi-squared Statistic**: **158.2564**

This is a relatively **large statistic**, indicating that there is a **significant difference** betweenthe observed and expected frequencies of outcomes.

**p-value**: **0.0000**

The **p-value of 0.0000** suggests that the difference between observed and expected outcomes is **extremely significant**. This means that the observed outcomes are **highly unlikely** to have come from the expected distribution under the null hypothesis. Therefore, we reject the null hypothesis and conclude that the simulated outcomes **do not follow the expected distribution**.

**Part 2: NY, B, NY (3-game series with 2 away games in New York)**

**Chi-squared Statistic**: **10.1124**

This statistic is smaller than in Part 1, but it still indicates a **noticeable difference** between observed and expected outcomes.

**p-value**: **0.0015**

The **p-value of 0.0015** is still quite small, suggesting a **significant deviation** between the observed and expected frequencies. While not as extreme as Part 1, this still provides strong evidence against the null hypothesis, meaning the observed outcomes **do not perfectly match** the expected outcomes.

**Part 3: Best of Five (5-game series with alternating home and away games)**

**Chi-squared Statistic**: **144.0000**

The **Chi-squared statistics of 144** is also very high, indicating a **large deviation** between the observed and expected distributions of outcomes, like Part 1.

**p-value**: **0.0000** (and **3.5529642241553834e-33**)

The **value of 0.0000** (or the extremely small value, 3.5529642241553834e-33) further reinforces the idea that the observed outcomes **differ significantly** from what was expected. The result is **highly statistically significant**, meaning there is very strong evidence that the simulated outcomes do not follow the expected pattern.

**Conclusion:**

For all three parts, the **Chi-squared statistic**s are large, indicating that there is a **substantial deviation** between the observed and expected outcomes.

The **p-values** for all three parts are **very small** (close to 0), providing **strong evidence** against the null hypothesis. This suggests that the outcomes of the simulations do not conform to the expected distribution. Therefore, we conclude that there is a significant difference between the observed outcomes and the expected outcomes for each of the three series.

These results imply that the betting outcomes do not follow the expected distribution, and factors such as team performance, home/away advantage, or other dynamics in the series may be influencing the results more than initially anticipated.

1. **Visualizations:**

**Part: 1**

**Histogram of Net Winnings**

* Purpose: This histogram visualizes the frequency distribution of net winnings, with the x-axis representing the range of net winnings in dollars and the y-axis representing the frequency (or count) of data points that fall into each range.
* Insights:
  + There is a high concentration of data points at both extremes: -1000 and 1000. This suggests that a significant portion of the net winnings data falls within the two extreme ranges (negative and positive).
  + The absence of a noticeable spread between the two extremes (with the bars tightly clustered) implies that there may be two distinct groups of outcomes: one group consistently losing and the other consistently winning, with little variability in between.
  + The presence of two large peaks at -1000 and 1000 indicates that the majority of the data points are either in a large loss (-1000) or a significant gain (1000).

A graph with numbers and lines

AI-generated content may be incorrect.

A graph with a blue line

AI-generated content may be incorrect.

* Impact: This histogram helps to quickly visualize the overall distribution of net winnings, showing that a majority of individuals (or cases) fall into either a large loss or a significant gain. This could suggest a high-risk scenario, where there are few middle-ground results

**Box Plot of Net Winnings**

* Purpose: The box plot visually represents the spread and summary statistics of the net winnings data. The central box highlights the interquartile range (IQR), showing where the middle 50% of the data lies, while the whiskers extend to the minimum and maximum values (excluding outliers). The orange lines represent the median and quartiles.
* Insights:
  + The box is quite narrow, indicating a lack of variation in the data between the middle 50% of net winnings. This suggests that, among most participants, the net winnings are clustered within a relatively small range.
  + The whiskers are extended on either side of the box, suggesting the presence of extreme values, but these do not appear to significantly alter the central tendency of the data (as the median remains centered).

A screen shot of a graph

AI-generated content may be incorrect.

* + The fact that the box plot shows a narrow IQR around the center suggests that the central group of winnings is consistent, with few outliers.
* Impact: The box plot provides a clearer understanding of the data's spread, particularly focusing on the central tendency and the variability. It shows that while there are extreme values, most participants are clustered around the median, indicating consistency within the bulk of the data.

**Part: 2**

**Histogram of Net Winnings (Part 2 - NY, B, NY):**

* + The histogram shows a bimodal distribution, with two prominent peaks. One peak is around -1000 and the other around 750.
  + The left peak, around -1000, suggests a significant number of occurrences where the net winnings were negative (losses), potentially nearing or equal to -1000.
  + The right peak around 750 suggests a significant number of occurrences where the net winnings were positive, likely near the 750 range.
  + This histogram reflects two distinct groups or outcomes for the net winnings, one at a loss and another with moderate positive winnings. It could indicate a system with many low losses and a moderate number of positive outcomes.

A red vertical column with numbers

AI-generated content may be incorrect.

**A graph with a blue bar

AI-generated content may be incorrect.**

**Box Plot of Net Winnings (Part 2 - NY, B, NY):**

* + The box plot presents a summary of the distribution of net winnings. The central box shows that most of the data (the interquartile range) is between -1000 and 750, suggesting many observations lie in this range.
  + The plot shows a relatively small spread, with most data points concentrated around the lower and upper bounds. The median value seems to lie closer to 0, indicating that the distribution might have a tendency towards a break-even or neutral outcome.
  + There are no clear outliers, as the whiskers extend evenly from the box, reflecting that the majority of data points are contained within this interquartile range**.**

**A graph with a blue line

AI-generated content may be incorrect.**

**Part 3**

Histogram of Net Winnings (Part 3 - Best of Five)

Insights:

* The histogram shows the frequency distribution of net winnings across specific ranges (e.g., -1500,−1500,−1000, -500,500,0, 500,500,1000, $1500).
* The x-axis represents net winnings ($), and the y-axis represents frequency (how often net winnings fall into each range).
* The histogram provides a visual representation of how net winnings are distributed, allowing you to identify patterns or trends.

A green bar graph with numbers

AI-generated content may be incorrect.

A graph with a number of points

AI-generated content may be incorrect.

What the Histogram Tells Us:

1. Most Common Net Winnings:
   * The tallest bars indicate the most frequent ranges of net winnings. For example, if the bar at 0 is the tallest, it means that most games resulted in net winnings close to 0 *is the tallest*, *it means that most games resulted in net winnings close to* 0.
2. Spread of Net Winnings:
   * The histogram shows whether net winnings are concentrated in certain ranges or spread out evenly. For instance, if most bars are clustered around $0, it suggests that net winnings are relatively consistent and close to breaking even.
3. Outliers or Unusual Data:
   * If there are very short or tall bars at the extremes (e.g., -1500or1500*or*1500), it may indicate outliers—games with unusually high losses or gains.
4. Skewness of Data:
   * If the histogram is left-skewed (longer tail on the left), it means there are more games with losses. If it is right-skewed (longer tail on the right), it means there are more games with gains.

**Box Plot of Net Winnings (Part 3 - Best of Five)**

Insights:

* The box plot summarizes the distribution of net winnings, showing the median, quartiles, and potential outliers.
* The range of net winnings is from -1500to1500*to*1500, with specific points marked (e.g., -1000,−1000,−500, 0,0,500, $1000).
* The box plot provides a concise summary of the data's central tendency, variability, and skewness.

What the Box Plot Tells Us:

1. Median Net Winnings:
   * The line inside the box represents the median net winnings. If the median is close to 0,it suggests that half of the games resulted in net winnings above 0,*it suggests that half of the games resulted in net winnings above* 0 and half below.
2. Interquartile Range (IQR):
   * The box represents the middle 50% of the data (between the 25th and 75th percentiles). A wider box indicates greater variability in net winnings.
3. Outliers:
   * Points outside the whiskers of the box plot represent outliers—games with unusually high or low net winnings compared to the rest of the data.
4. Skewness:
   * If the box plot is asymmetric (e.g., the median is closer to the bottom or top of the box), it indicates skewness in the data. For example, if the median is closer to the bottom, the data is right-skewed (more games with gains).

**Combined Interpretation:**

* Histogram: The histogram shows that net winnings are distributed across a range, with the highest frequency likely around 0(breaking even). There may be some outliers at the extremes(−0(*breaking even*).*There may be some outliers at the extremes*(−1500 and $1500), but most net winnings are concentrated in the middle ranges.
* Box Plot: The box plot confirms that the median net winnings are close to $0, with a relatively symmetric distribution. The interquartile range (IQR) shows moderate variability, and there may be a few outliers at the extremes**.**
* The data suggests that in the **Best of Five** games, net winnings are generally close to breaking even ($0), with moderate variability.
* Most games result in net winnings within a reasonable range (e.g., -500to500*to*500), but there are occasional outliers with higher losses or gains.
* The symmetric distribution (as seen in the box plot) and the concentration of net winnings around $0 (as seen in the histogram) indicate that the games are relatively balanced, with no strong skew toward losses or gains.

**Comparison of Risk Metrics**

**Insights:**

* This graph compares three risk metrics across Part 1, Part 2, and Part 3:
  1. Mean ($): The average net winnings.
  2. CV (Risk per $1 Profit): The coefficient of variation, which measures risk relative to the mean (higher CV = higher risk per dollar of profit).
  3. Prob. of Loss: The probability of incurring a loss.
* The y-axis represents the value of each metric, while the x-axis lists the parts (Part 1, Part 2, Part 3) and possibly a Series (e.g., a combined or overall metric).

**What the Graph Tells Us:**

1. **Mean ($):**
   * The mean net winnings for each part indicate the average profitability. For example, if Part 1 has a higher mean than Part 2, it suggests that Part 1 is more profitable on average.
2. **CV (Risk per $1 Profit):**

**A graph with red and blue squares

AI-generated content may be incorrect.**

* + The coefficient of variation (CV) measures the risk relative to the mean. A higher CV indicates greater risk per dollar of profit.
  + For example, if Part 3 has a higher CV than Part 1, it means that Part 3 is riskier relative to its average profitability.

1. **Prob. of Loss:**
   * The probability of loss shows how likely it is to incur a loss in each part. For example, if Part 2 has a higher probability of loss than Part 1, it means that Part 2 is riskier in terms of potential losses.

**Trends:**

* If the means are increasing across parts (e.g., Part 1 < Part 2 < Part 3), it suggests that profitability is improving.
* If the CV is decreasing across parts, it indicates that risk per dollar of profit is decreasing.
* If the probability of loss decreases across parts, it suggests that the likelihood of incurring losses is reducing.

**Comparison of Expected Net Winnings and 95% Confidence Intervals**

**Insights:**

* This graph compares the expected net winnings (mean) and their 95% confidence intervals across Part 1, Part 2, and Part 3.
* The y-axis represents the expected net winnings ($), while the x-axis lists the parts (Part 1, Part 2, Part 3).

**What the Graph Tells Us:**

**A graph with blue squares

AI-generated content may be incorrect.**

1. **Expected Net Winnings:**
   * The expected net winnings for each part indicate the average profitability. For example, if Part 3 has a higher expected net winnings than Part 1, it suggests that Part 3 is more profitable on average**.**
2. **95% Confidence Intervals:**
   * The confidence intervals show the range within which the true mean net winnings are likely to fall, with 95% confidence.
   * A wider confidence interval indicates greater uncertainty or variability in the data.
   * For example, if Part 2 has a wider confidence interval than Part 1, it means that the net winnings in Part 2 are less predictable.

**Trends:**

* If the expected net winnings are increasing across parts (e.g., Part 1 < Part 2 < Part 3), it suggests that profitability is improving.
* If the confidence intervals are narrowing across parts, it indicates that the data is becoming more consistent and predictable.

**Combined Interpretation:**

* Risk Metrics: The comparison of risk metrics shows how profitability, risk per dollar of profit, and the probability of loss vary across parts. For example, Part 3 might have higher profitability but also higher risk compared to Part 1.
* Expected Net Winnings and Confidence Intervals: The comparison of expected net winnings and confidence intervals provides insights into the average profitability and the certainty of those estimates. For example, Part 2 might have lower expected net winnings but also lower variability, making it a safer option.
* **Part 1**: Likely has moderate profitability, moderate risk, and a moderate probability of loss.
* **Part 2**: May have lower profitability but also lower risk and a lower probability of loss, making it a safer option.
* **Part 3**: Could have higher profitability but also higher risk and a higher probability of loss, making it a riskier option.

**Final Conclusion:**

**Part 1:**

* Mean Net Winnings: $118.52
  + On average, you gained $118.52 per game in Part 1.
  + This indicates that Part 1 was profitable overall.
* Standard Deviation: $1010.60
  + The standard deviation measures the variability or risk in net winnings. A high standard deviation ($1010.60) suggests that the net winnings varied significantly from game to game.
  + This indicates high risk in Part 1, as outcomes were unpredictable.
* 95% Confidence Interval: (98.71,98.71,138.32)
  + The true mean net winnings for Part 1 are likely to fall between 98.71∗∗and∗∗98.71∗∗*and*∗∗138.32 with 95% confidence.
  + This narrow interval suggests that the estimate of the mean net winnings is relatively precise.

**Part 2:**

* **Mean Net Winnings**: -$72.84
  + On average, you lost **$72.84** per game in Part 2.
  + This indicates that Part 2 was **unprofitable** overall.
* **Standard Deviation**: $1018.68
  + Similar to Part 1, the standard deviation is high ($1018.68), indicating significant variability in net winnings.
  + This suggests **high risk** in Part 2 as well.
* **95% Confidence Interval**: (-92.80,−92.80,−52.87)
  + The true mean net winnings for Part 2 are likely to fall between **-92.80∗∗and∗∗−92.80∗∗*and*∗∗−52.87** with 95% confidence.
  + This interval confirms that Part 2 was consistently unprofitable, as the entire range is negative.

**Part 3:**

* **Mean Net Winnings**: $130.65
  + On average, you gained **$130.65** per game in Part 3.
  + This indicates that Part 3 was **profitable** overall and had the **highest average net winnings** among the three parts.
* **Standard Deviation**: $1521.62
  + The standard deviation is even higher in Part 3 ($1521.62), indicating **very high variability** in net winnings.
  + This suggests that Part 3 was the **riskiest** of the three parts, with outcomes being highly unpredictable.
* **95% Confidence Interval**: (100.83,100.83,160.47)
  + The true mean net winnings for Part 3 are likely to fall between **100.83∗∗and∗∗100.83∗∗*and*∗∗160.47** with 95% confidence.
  + This interval is wider than Part 1's, reflecting the higher variability in Part 3.
* **Part 1** was **moderately profitable** but came with **high risk**. It is a reasonable option for those willing to accept some variability in outcomes.
* **Part 2** was **unprofitable** and also **high-risk**. It is the least favorable option among the three.
* **Part 3** was the **most profitable** but also the **riskiest**. It is suitable for those willing to take on higher risk for the chance of higher returns.

**Recommendations:**

* If you prefer **stability and moderate profitability**, **Part 1** is the best choice.
* If you are willing to take **higher risk for higher returns**, **Part 3** is the best option.
* **Part 2** should be avoided due to its consistent losses and high risk.