Random Numbers and David Lynch

Sean Franco, www.seanfranco.com

May 18, 2025

Abstract

This inquiry evaluates the randomness in David Lynch's 853 "Today's Number Is..." (TNI) series from August 2020 to December 2022. An API collected the transcript, and extracted the TNI number and date. This information was verified (with "Seth M-T"'s YouTube Channel) sequence of the TNI series. To evaluate the degree of foul and fair play involved in the TNI series statistical analyses were performed. Statistical Analyses performed were a chi-squared goodness of fit test, a difference in days analysis and a Monte Carlo portfolio analysis using simulated data, multiple correspondence analysis, and mahalanobis distances. The distribution of TNI numbers does not significantly deviate from random, simulated data from the multinomial distribution. Excluding other evaluation methods, and an outlier DiD gap of 84, the analyses did not identify foul play that would influence the TNI random selections.

1 Introduction

It was sad when the painter, and film director David Lynch passed away in January 2025. In response to this news like other fans, I re-watched some of his movies, interviews and such. It was a pleasure to revisit his content, and his short-form, videos on his YouTube channel. David Lynch's daily YouTube content was particularity enjoyable during the onset, and duration of the Coronavirus 19 pandemic. Some of his memorable video series during this period were his "Weather Reports", "What Is David Working on Today?", and his "Today's Number Is..." series.

David Lynch's "Today's Number Is..", hence referred as (TNI) videos were appealing to watch. The videos involved David Lynch announcing the date, followed by showing the viewer a close view of a partially transparent jar as in Figure 1. This jar is filled with ten ping pong balls that are each labeled with a natural number from one to ten. The jar had black tape on the bottom to facilitate a random number selection, and a transparent top section which David overturns in each video to prove that it was filled with ten ping pong balls labeled with numbers. The first entry in this series occurred on August 16^{th} , 2020 titled: "JAR FINAL", and the last entry in this series was on December 16^{th} , 2022 and titled: "TODAY'S NUMBER IS... 12/16/22". All video entries except JAR FINAL video use the latter naming scheme.



Figure 1: The first TNI video. David Lynch showcases the first number (8) in the series along with the partially obscure jar from which the numbers are selected for the entire series.

The videos are fairly formulaic where David would show the numbers in the visible section of the jar, then "swirl the numbers", followed by "pick a number" directions. The videos resolve with the proclamation of "today's number is..." followed by that particular day's number. Over the 2.33 years of this series, David maintains a systematic pattern of speech, word choices, and form such that each video is fairly consistent and systematic in sentences, and the form in their transcript.

Another particular enjoyable aspect of David Lynch's YouTube uploads were the various contributions and fan engagement in the comments sections. On many of the TNI videos some repeated channels commented about their invested interest in the TNI selections. Comments even referenced doubt and speculations of a particular number's frequency and overall concern of foul play. This inquiry investigates the degree of which David Lynch's TNI series involved fair play of random number selections.



Figure 2: David Lynch's 853 TNI numbers in blue, and 853 simulated natural numbers in orange. Both TNI numbers and the simulated random numbers are from the multinomial distribution.



Figure 3: David Lynch's TNI total counts grouped by years. 2020 only incurred 138 days, whereas 2021 had 365 days, and 2022 had 350 days.

2 Methods

The easiest, and unrealistic method to determine if this series involved fair number selections is to interview the people involved in this production and to inspect the physical materials such as the balls, and jar for each video. This inspection offers insight if there is any other factors involved during the filming and editing process. Of course this is not possible, nor is it feasible to retroactively inspect the physical materials and raw footage. So an alternative approach utilizing statistical methods was performed.

The YouTube API was used on David Lynch's YouTube channel which saved all TNI videos in a playlist [2]. The API retrieve each entries metadata for all 852 TNI entries and the bonus starting video (JAR FINAL) as in Figure 1 for a total of 853 TNI videos. Data from each video included the video title, date of publication, and the entire video transcript. Most TNI videos involved David speaking systematically, and in a consistent pattern. These attributes aided the extraction of a given day's number from the transcripts.

Out of 853 largely consecutive daily videos, only 39 entries needed a manual entry. Upon review of these 39 entries, the main reason why they needed manual review is due to YouTube's transcription feature not accurately capturing David's words. This could be due to David changing microphones, or rooms where he would typically film the TNI videos. This was noted visually. However, David was largely consistent with his video structure and terminology. These entries, along with the first, hidden entry for the TNI were manually reviewed for their TNI and inputted for a total of 853 entries in the data. Verification of the transcript dates and TNI numbers were compared to another YouTube channel: "Seth M-T" who dedicated themselves to providing summary statistics of TNI numbers. Both the dates and TNI numbers were verified as correct compared to "Seth M-T"'s efforts [4]. Once data was collected, cleaned, and verified as an accurate TNI number sequence, then statistical analysis was performed.

3 Discussion

The guiding question of this inquiry is: how fair were David Lynch's TNI numbers? In most videos transparent actions were performed such as starting each TNI video by showcasing the 10

numbered balls in the jar, followed by a blind "swirling the numbers" before selecting a number for that given day see figure 1. We can asses confidence in the randomness of Mr.Lynch's TNI selections beyond the scope of his actions. The audience trust in the TNI series fair play rests on repeated selections of the 10 unique, natural number balls from 1 to 10. In statistics these equal expected probability balls are categorical data types from the multinomial distribution.



Figure 4: David Lynch's 853 TNI probabilities are in blue bars, and the probabilities of the entire (365,000) simulated portfolio are in orange bars. The TNI numbers experience random error in their probabilities. The simulated portfolio converges towards expected probabilities of 0.1 from the multinomial distribution due to the law of large numbers.

A Pearson's Chi-Squared goodness of fit test was performed on the TNI numbers as observed in the blue bars in Figure 2. This statistical test offers insight on whether any given frequency of the natural number category differ significantly from expected values. The null hypothesis was not rejected at the 5% level, so variation of each number category is likely attributed to random error. This suggests that the occurrence of the 10 given number categories did not significantly differ from expected values at the 5% significant level. Plainly spoken, the variation in the TNI categories is more likely due to randomness than altered TNI occurrences. To further illustrate randomness, simulated, random natural numbers from 1 to 10 were generated 853 times to match the length of the TNI data and were plotted in orange bars in Figure 2 to show the variety of randomness within the multinomial distribution.



Figure 5: The distribution of the difference in days for the 10 natural numbers. The 10 counts for difference-in-day 0 account for the first occurrence of the 10 natural numbers.

The difference in days (DiD) for all numbers were plotted in Figure 4. DiD gaps of zero account for the first occurrences of each 10 numbers. As expected, the distribution is right-skewed where smaller DiD gaps are more common than larger DiD gaps. A negative binomial distribution was fitted in Figure 5 on the DiD gaps since they are overdispersed count data. An outlier of 84 DiD occurred in 2021 for the number 9. The probability of a DiD gap of 84 days was calculated to be 0.002% based on the negative binomial distribution (see the orange line in Figure 5). This could be interpreted as a random outlier, or as foul play in the TNI selections. Additional analysis is warranted to determine if David Lynch influenced the TNI selections in other ways.

A Monte Carlo analysis was performed to compare the 2021 sample of David Lynch's TNI results (that has the 84 DiD gap) with simulated data. Simulated arrays of 365 observations of natural numbers from 1 to 10 were generated 1,000 times from the multinomial distribution. Essentially one years worth of simulated TNI data was generated 1,000 times. The Law of Large Numbers indicates that larger observations minimize randomness as they converge to expected values [3, 1]. The simulated portfolio of 1,000 random natural numbers in orange approach their expected probabilities of 0.1 from the multinomial distribution as observed in Figure 4 compared to the TNI numbers. Leveraging the Law of Large Numbers, a sample of TNI results for the year





Figure 6: 1,000 simulated portfolio points and the 2021 TNI subset were plotted in the first 2 MCA components. The first three standard deviations (from inner to outer) were plotted in black ovals from the mahalanobis distances. Outlier points are identified in purple for 5 dimensions of MCA components that are beyond the third standard deviation based on 2 degrees of freedom Chi-Squared test of mahalanobis distances

2021 was compared to the simulated portfolio to determine the fit of the 2021 TNI numbers in reference to the simulated portfolio.

The 2021 TNI subset was appended to the simulated data of a shape (1001, 365). The data was standardized, and a multiple correspondence analysis was performed to reduce the categorical variables to 5 MCA components. Five MCA components were chosen as their total explained variance was 99%. Due to visualization limitations the first 2 MCA components were plotted (in 2 dimensions) in Figure 6 where each coordinate point is a representation of the original 1,001 iterations in green points. The 2021 TNI sample is marked in a black star in Figure 6, and is not an outlier across the 5 MCA components. Mahalanobis distances were calculated for the 5 MCA components. The first three standard deviation ellipses of the first 2 MCA components were plotted in the inner (1σ) , middle (2σ) and the outer (3σ) black rings.

Across the 5 MCA components, most of the data resides within the first and second standard

deviations respectively, including David Lynch's 2021 TNI sample. Outlier coordinates of all 5 MCA components are identified in purple points and occur beyond the third standard deviation. The 2021 TNI sample is located near the first standard deviation which suggests that this set of numbers is similar to the rest of the simulated portfolio. Whereas if this sample occurred at a larger standard deviation compared to the simulated portfolio, then an assumption of randomness would be in question. Additionally the 2021 TNI sample is not identified as an outlier across the 5 MCA components based on its mahalanobis distance in a degree 5, Chi-Squared test which suggests that the 2021 TNI data is within expectations of randomness. This supports the conclusion that the 84 DiD gap is likely due to random chance, rather than foul play.

4 Conclusion

This inquiry evaluated the degree of fair and foul play of David Lynch's "Today's Number Is.." (TNI) series. An API retrieved the transcript of the TNI series to retrieve the particular TNI number and date. These entries were verified with the data provided by other David Lynch TNI fan channels (thank you to "Seth M-T"'s YouTube channel) and statistical analysis was performed.

The null hypothesis of the goodness of fit chi-squared test was not rejected which suggests no identifiable category on different proportions. This suggests that the series overall did not have any particular category occur more often than any other category. A difference in days (DiD) analysis was performed and a negative binomial distribution was fit to identify probabilities of particular TNI numbers repeating. An outlier of a 84 DiD gap was identified for the year 2021 so additional analysis was performed to evaluate if foul play occurred for the 2021 TNI sample. A Monte Carlo portfolio analysis was performed to benchmark the 2021 TNI sample to 1,000 simulated data. A Multiple Correspondence Analysis reduced the portfolio to 5 dimensions and mahalanobis distances were calculated on the MCA portfolios. Across all 5 MCA components, the mahalanobis distance of the 2021 TNI sample was not identified as an outlier based on a Chi-Squared test with 5 degrees of freedom.

Across each analysis no statistically significant evidence was found to suggest foul play or to prompt further evaluation into the distribution of TNI numbers. Excluding other evaluation methods, no foul play was identified that would influence TNI expected probabilities.

References

- Geraghty, M. A. and College, D. A. (2022). Inferential Statistics and Probability A Holistic Approach. Statistics LibreTexts.
- [2] Lynch, D. (2025). David lynch theater youtube channel. Accessed: 2025-03-05.
- [3] Pishro-Nik, H. (2014). Introduction to probability, statistics, and random processes. Chapter 7.1.1: Law of Large Numbers.
- [4] T., S. M. (2025). Seth m. t. youtube channel. Accessed: 2025-03-05.