Analyzing Rarity, Pricing, and Market Behavior in Pokémon Cards Using Probability Distributions

By: Dominic Zelinsky

Abstract

This study aims to apply statistical methods to analyze a dataset containing 25,598 Pokémon trading cards, examining the card rarity, listing frequency, and pricing. This was done by using discrete and continuous probability distribution; Modeling card selection using binomial and geometric distribution; Listing rates using Poisson and uniform; and pricing distributions using beta and joint. Also utilized Excel in order to produce valuable figures to analyzing and connecting the data to real world scenarios.

Table of Figures

Introduction

The Pokémon Franchise has been a staple in pop culture for close to 30 years now. Taking the gaming world by storm, it broke through multiple markets. The first games spawned spinoff games, a television series, and the topic of this article is the Pokémon Trading Card Game. This study analyzes a dataset containing 25,598 Pokémon cards, columns including the Pokémon themselves, the Card Type, Generation, Card Number, and the pricing in £ (pound sterling). The mean being £5.96, Standard Deviation £21.60, a median £0.89, a maximum of £899.99. With such a large dataset, it is perfect to support such a robust statistical analysis.

The objective is to apply these probability and distribution models learned from the Mathematical Statistics with Applications (Wackerly et. 2008) to understand the card selection, listing frequency, and pricing dynamics. Using questions from chapters 3(Probability), 4(Discrete Distributions, and 5(Continuous Distributions) I created 6 questions to conduct this analysis. These models would be helpful to collectors trying to predict acquiring a rare card, assisting sellers in estimating list rates, and providing investors the ability to assess the price distributions.

Methodology

Chapter 3:

a. Problem Based off 3.122(pg. 136)
   - A shop receives listing requests for "Reverse Holo" Pokémon cards, according to ta Poisson distribution with an average of 7 per hour. During a given hour, what are the probabilities that:
      a) No more than three listing requests arrive?
      b) At least two listing requests arrive?
      c) Exactly five listing requests arrive?
   - **Method:** This question utilizes Poisson distribution, $\lambda = 7$. The probabilities calculated used PMF and cumulative probabilities.

b. Problem Based off (3.168(pg. 147)
   - A collector identifies Pokémon cards as "Reverse Holo" or "Standard" in a guessing game of 100 cards. Guessing correct P(correct) = 0.5. Let Y denote the number of correct guesses.
      a) Find $E(Y)$
      b) Find the standard deviation of Y.
      c) Calculate $\mu \pm 2\sigma \; and \; \mu \pm 3\sigma$
      d) If 50 correct answers are a passing score, is a passing score likely? Explain.

   - **Method:** This problem utilizes Binomial distribution with $n = 100, p = 0.5$. It also uses the Expected value, variance, and normal approximation to compute the probabilities and intervals.

Chapter 4:

c. Problem Based off 4.2(Pg. 166)
   - A collector is searching for a "Charizard Reverse Holo" Pokémon. A seller is having a special event where the collector can pick from 5 flipped over cards. Only one of the cards is the Charizard, and the other four cards are different cards. The cards are randomly shuffled, and the collector selects and checks one card at a time without replacing it. He is trying to find the Charizard, but does not know which of the 5

positions that it is in. Let Y denote the number of trials in which the Charizard card is found.

a. Find the probability function for Y.
b. Give the corresponding distribution function.
c. What is $P(Y < 3)$? $P(Y \leq 3)$? $P(Y = 3)$?
d. If Y is a continuous random variable, we argued that for all -∞ < $a < \infty$, $P(Y = a) = 0$. Do any of your answers in part “c” contradict this claim? Why?

- **Method:** This problem utilizes discrete Uniform Distribution due to the finite number of trials without replacing a card. The Probabilities are calculated using PMF and CMF and a discussion of discrete vs continuous variables.

a. Problem based off 4.45(pg.176)

- Upon looking at the prices of Pokémon cards, a collector discovers that the normalized price of the cards is uniformly distributed between 0 and 1, with values mapped to a range between £20 and £25. Let Y Denote the normalized price.

  a) Find the probability that the price is below £22.
  b) Find the probability that the price exceeds £24.

- **Method:** This problem utilizes Uniform distribution over [20,25]. Probabilities were calculated using the cumulative distribution function (CDF).

Chapter 5:

b. Problem based off 5.1(pg. 232)

- The Pokémon card listings are randomly assigned to one or more of three sellers (A, B, C). Let Y1 denote the number of listings assigned to seller A and Y2 the number assigned to seller B. Each seller can receive 0,1, or 2 listings.

  a) Find the joint probability function for Y1 and Y2
  b) Find F (1,0)

- **Method:** This problem utilizes Continuous density function, normalized to find (k). The probabilities were computed via integration and conditional probability formulas.

c. Problem based off 5.45(pg. 251)

- The normalized price of Pokémon cards (Price £ / 899.99) has density $f(y) = ky(1-y), 0 \leq y \leq 1$. Let Y denote the normalized price.
  a) Find k
  b) Find P$(.4 \ \leq Y \leq 1)$
  c) Find P$(.4 \leq Y < 1)$
  d) Find P$(Y \leq .4 \mid Y \leq .8)$
  e) Find P$(Y < .4 \mid Y < \ .8)$

- **Method:** This problem utilizes continuous density function normalized to find k.

Results

The following are the solutions and results from the problems created.

Problem based off 3.121(Poisson)

Parameters: Poisson ($\lambda = 7$), number of "Reverse Holo" requests per hour

- $\boldsymbol{P(Y \leq 3)}$:

$$P(Y \leq 3) = \sum_{k=0}^{3} \frac{e^{-7} * 7^k}{k!},\ e^{-7} \approx 0.000912$$
$$P(Y = 0) = \frac{0.000912}{0!} = 0.000912$$
$$P(Y = 1) = \frac{0.000912 * 7^1}{1!} = 0.006384$$
$$P(Y = 2) = \frac{0.000912 * 7^2}{2!} = 0.02234$$
$$P(Y = 3) = \frac{0.000912 * 7^3}{3!} = 0.052136$$
$$P(Y \leq 3) = 0.000912 + 0.006384 + 0.02234 + 0.052136 \approx 0.081776$$
$$P(Y \leq 3) \approx 8.18\%$$

- $\boldsymbol{P(Y \geq 2)}$:

$$P(Y \geq 2) = 1 - (P(Y = 0) + P(Y = 1) = 1 - (0.000912 + 0.006384)$$
$$P(Y \geq 2) \approx 1 - 0.007296 \approx 0.992704$$
$$P(Y \geq 2) = 99.27\%$$

- $\boldsymbol{P(Y = 5)}$:

$$P(Y = 5) = \frac{e^{-7} * 7^5}{5!}$$
$$P(Y = 5) = \frac{0.000912 * 7^5}{5!} \approx 0.1277332$$

$$P(Y = 5) \approx 12.77\%$$

Problem based off 3.168(Binomial)

Parameters: $n = 100(\# \; of \; trials), \; p = 0.5(probability \; of \; correct \; guess),$

- *Find $E(Y)$*

$$E(Y) = \mu$$
$$\mu = n * p = 100 * 0.5 = 50$$
$$E(Y) = 50$$

- *Find the standard deviation of Y.*

$$\sigma = \sqrt{n * p * (1 - p)}$$
$$\sigma = \sqrt{100 * 0.5 * 0.5} = \sqrt{25}$$
$$\sigma = 5$$

- *Calculate $\mu \; \pm 2\sigma \; and \; \mu \; \pm 3\sigma$*

$$\mu \pm 2\sigma = 50 \pm 2 * 5 = [40{,}60],$$
$$\mu \pm 3\sigma = 50 \; \pm 3 * 5 = [35{,}65]$$

- *If 50 correct answers are a passing score, is it likely to be a passing score? Explain.*

$$P(Y \geq 50) \approx Z = \frac{50 - \mu}{\sigma} = \frac{50 - 50}{5} = 0$$

Continuity Correction:

$$Z = \frac{49.5 - 50}{5} = -0.1 = 0.5398(From \; the \; Standard \; Normal \; Table)$$
$$P(Y \geq 50) \approx 53.98\%$$

Problem based off 4.2(Discrete Uniform)

Parameters: $n = 5(total \; cards),$

$$P(success) = \frac{1}{5}(Probability \; of \; selecting \; the \; Charizard \; per \; trial)$$

- *Find the probability function for Y.*

$$P(Y = 1) = \frac{1}{5} = 0.2$$
$$P(Y = 2) = \frac{4}{5} * \frac{1}{4} = 0.2$$
$$P(Y = 3) = \frac{4}{5} * \frac{3}{4} * \frac{1}{3} = 0.2$$
$$P(Y = 4) = \frac{4}{5} * \frac{3}{4} * \frac{2}{3} * \frac{1}{2} = 0.2$$

$$P(Y=5)=\frac{4}{5}*\frac{3}{4}*\frac{2}{3}*\frac{1}{2}*1=0.2$$

PMF: $P(Y=k)=\frac{1}{5}=0.2, k=1,2,3,4,5$

- *Give the corresponding distribution function.*

$$F(y)=P(Y\leq y)=\sum_{k=1}^{|y|}0.2$$

$$F(y)=\begin{cases}0, & y<1\\ 0.2, & 1\leq y<2\\ 0.2, & 2\leq y<3\\ 0.2, & 3\leq y<4\\ 0.2, & 4\leq y<5\\ 1, & y\geq 5\end{cases}$$

- *What is* $P(Y<3)$? $P(Y\leq 3)$? $P(Y=3)$?

$$P(Y<3)=P(Y=1)+P(Y=2)=0.2+0.2=0.4$$
$$P(Y<3)=40\%$$
$$P(Y\leq 3)=P(Y=1)+P(Y=2)+P(Y=3)=0.2+0.2+0.2=0.6$$
$$P(Y\leq 3)=60\%$$
$$P(Y=3)=0.2=20\%$$

- *If Y is a continuous random variable, we argued that for all* $-\infty<a<\infty$, $P(Y=a)=0$. *Do any of your answers in part "c" contradict this claim? Why?*

  $P(Y=3)=0.2\neq 0$ meaning it is discreate and there is no contradiction with continuous $P(Y=a)=0$

Problem based off 4.45(Uniform)

Parameters: Y – Uniform [0,1], $f(y)=1,\ Z=20+5Y, so\ Z\sim Uniform[20,25]$

$f(z)=\frac{1}{25-20}=\frac{1}{5}$

- *Find the probability that the price is below £22.*

$$Z<22=20+5Y<22=Y<\frac{22-20}{5}=0.4$$
$$P(Y<0.4)=\int_0^{0.4}1dy=0.4$$
$$P(Y<0.4)=40\%$$

- *Find the probability that the price exceeds £24.*

$$Z > 24 = 20 + 5Y = Y > \frac{24-20}{5} = 0.8$$
$$P(Y > 0.8) = \int_{0.8}^{1} 1dy = 1 - 0.8 = 0.2$$
$$P(Y > 0.8) = 20\%$$

Problem based off 5.1(Joint Probability)

Parameters: $Y1, Y2 = 0,1,2,$

Totals Outcomes = $3 * 3 = 9.$

Equally likely probability = $\frac{1}{9}$

- *Find the joint probability function for Y1 and Y2*

Y1/Y2	0	1	2
0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

- *Find F (1,0):*

$$F(1,0) = P(Y1 \leq 1, Y2 \leq 0) = P(Y1 = 0, Y2 = 0) + P(Y1 = 1, Y2 = 0)$$
$$P(Y1 = 0, Y2 = 0) = \tfrac{1}{9}, P(Y1 = 1, Y2 = 0) = \tfrac{2}{9}$$
$$F(1,0) = \frac{1}{9} + \frac{2}{9} = \frac{3}{9}$$
$$F(1,0) = \frac{1}{3}$$

Problem based off 5.45(Beta-like)

- *Find k*

$$\int_0^1 ky(1-y)dy = 1$$
$$\int_0^1 y - y^2 dy = \left(\frac{1^2}{2} - \frac{1^3}{3}\right) - \left(\frac{1^2}{2} - \frac{1^3}{3}\right) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$
$$k * \frac{1}{6} = 1$$
$$k = 6$$
$$f(y) = 6y(1-y)$$

- *Find* $P(.4\ \leq Y \leq 1)$

$$\mathrm{P}(.4\ \leq Y \leq 1) = \int_{0.4}^{1} 6(y - y^2)\, dy$$

$$\int_{0.4}^{1} 6(y - y^2) dy = (3 * 1^2 - 2 * 1^3) - (3 * 0.4^2 - 2 * 0.4^3) = 0.648$$

$$\mathrm{P}(.4\ \leq Y \leq 1) = 64.8\%$$

- *Find* $P(.4 \leq Y < 1)$

Since Y is continuous: $P(.4 \leq Y < 1) = P(.4\ \leq Y \leq 1)$

$$P(.4 \leq Y < 1) = 64.8\%$$

- *Find* $P(Y \leq .4 \mid Y \leq .8)$

$$P(Y \leq .4 \mid \mathrm{Y} \leq .8) = \frac{P(Y \leq 0.4)}{P(Y \leq 0.8)}$$

$$P(Y \leq 0.4) = \int_0^{0.4} 6(y - y^2) dy$$

$$\int_0^{0.4} 6(y - y^2) dy = (3 * 0.4^2 - 2 * 0.4^3) - (3 * 0^2 - 2 * 0^3) = 0.352$$

$$P(Y \leq 0.8) = \int_0^{0.8} 6(y - y^2) dy$$

$$\int_0^{0.8} 6(y - y^2) dy = (3 * 0.8^2 - 2 * 0.8^3) - (3 * 0^2 - 2 * 0^3) = 0.896$$

$$P(Y \leq .4 \mid \mathrm{Y} \leq .8) = \frac{0.352}{0.896} \approx 0.392587$$

$$P(Y \leq .4 \mid \mathrm{Y} \leq .8) = 39.29\%$$

- *Find* $P(Y < .4 \mid Y <\ .8)$

Since Y is continuous: $P(Y < 0.4 \mid \mathrm{Y} < 0.8) = P(Y \leq 0.4 \mid \mathrm{Y} \leq 0.8)$

$$P(Y < .4 \mid \mathrm{Y} < .8) = 39.29\%$$

Figures

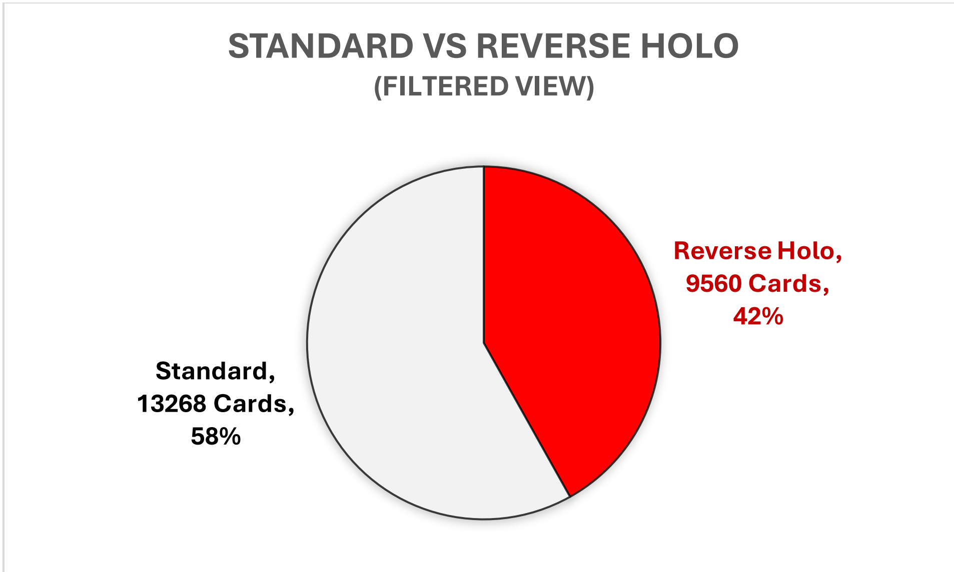

Figure 1: This chart compares the distribution of "Standard" and "Reverse Holo" cards (22,828 out of 25,598 total cards).

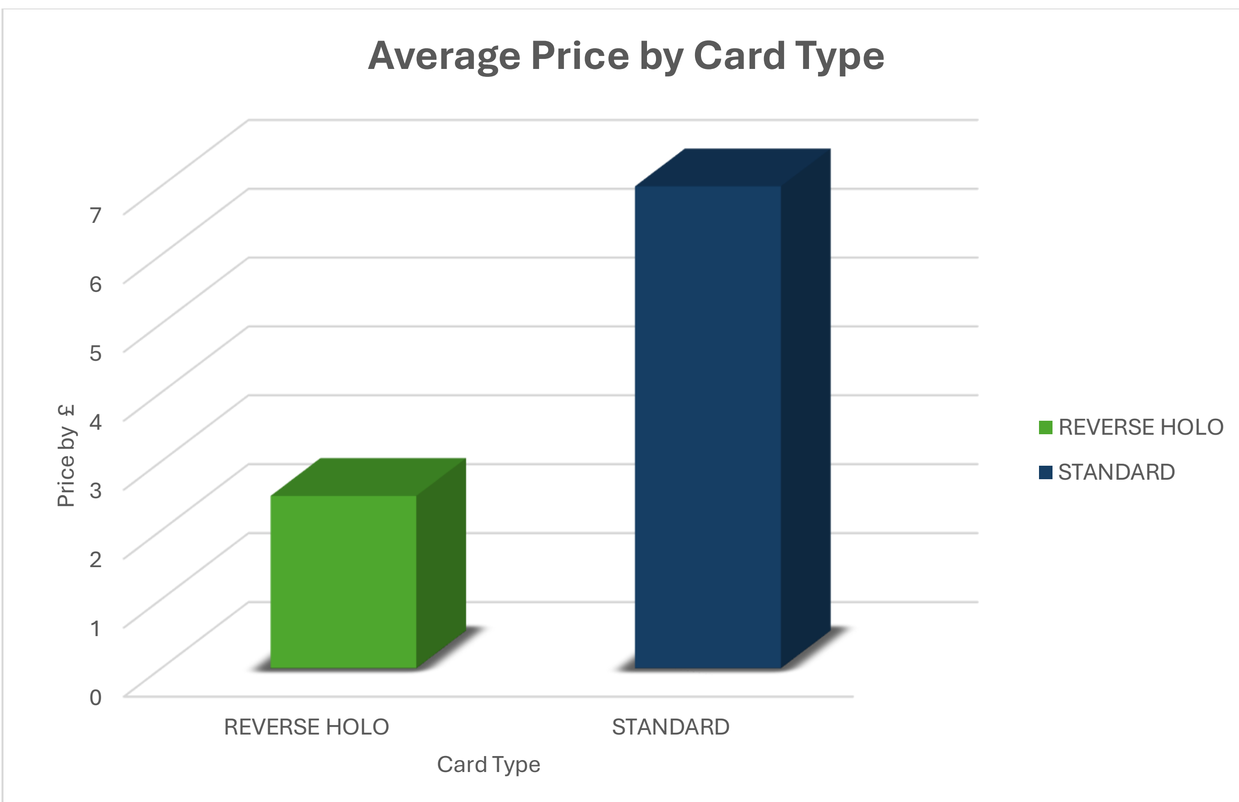

Figure 2: This displays the average price of "Standard" and "Reverse Holo" cards. Showing that standard cards have a significantly higher average price.

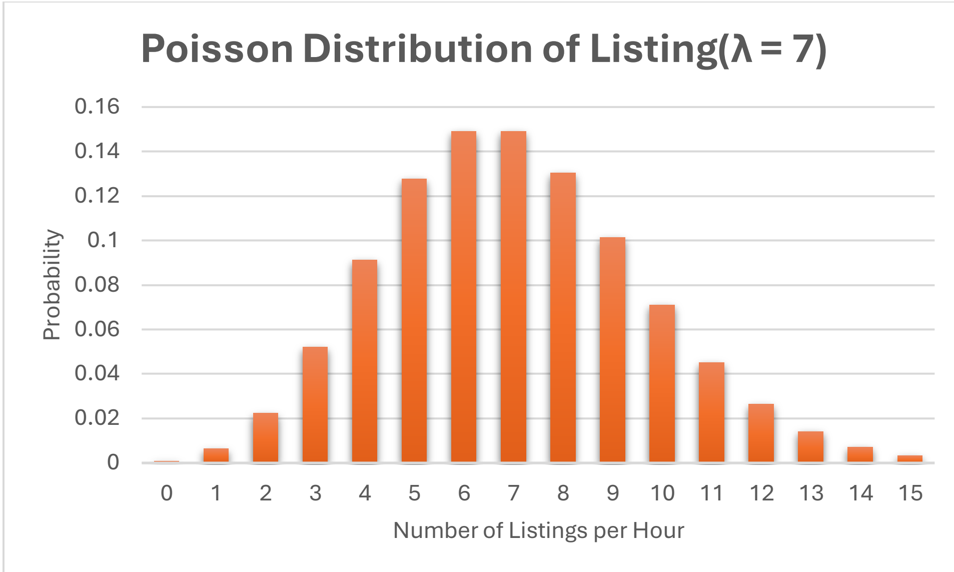

Figure 3: This is a Poisson distribution graph, which models the number of "Reverse Holo" card listing per hour. With an assumed rate of $\lambda = 7$.

Discussion

The results that were produced by the problems created allow us to gain a better understanding of the Pokémon card market. The problems provide information for collectors, sellers, and investors. Below, the results of each problem are discussed in detail connecting the statistical outcomes and figures to practical applications and market dynamics.

Poisson Problem (3.122):

- There is an 8.18% probability of receiving fewer than 3 "Reverse Holo" listing requests per hour, suggesting that low activity is rare. However, the 99.27% probability for 2 or more listings per hour indicates a consistent demand for the cards. When looking at part C we can conclude that the 12.77% probability of getting exactly 5 requests per hour highlights moderate activity. This information allows sellers to prepare for more steady listings, adjusting their inventory for hourly fluctuations. This also allows collectors to find an opportunity to find one during more low-activity hours. A flaw of this model is that it works off constant rates and does not include a promotion or event causing a spike in requests. When observing Figure 3, we see that it supports this interpretation as it visualizes the expected frequency of listings under a Poisson distribution with an assumed rate of $\lambda = 7$. This clearly shows that moderate listings are more likely, and the extremes are uncommon.

Binomial Problem (3.168):

- When evaluating the results of the binomial problem, the outcomes align with the expectations. The correct identifications hovered around 50, indicating a consistent success probability of 0.5, or 50%. This suggests that guessing would allow for an even chance of success, displaying the real-world difficulties or lack thereof, that collectors or graders may face when trying to distinguish between a “Standard” and “Reverse Holo” card. As shown in Figure 1, these two types of cards make up a substantial portion of the dataset (22,828 out of 25,598, ≈ 89%). However, the intervals between [40, 60] and [35, 65] correct guesses do highlight their variation and error that is present despite ideal conditions. A flaw of this model is that it may simplify the actual grading and identifying experience, where external factors like individual skill, experience, or condition of the card may affect the true probability.

Discrete Uniform (4.2):

- This question was asking for the probability of Y. In this case Y denoted the Probability of find a “Charizard Reverse Holo” among five cards, this yielded a probability of 20% for each of the 5 trials. It also asked for the probability of less than 3 attempts, 3 or less attempts, or exactly 3 attempts. For pulling the card in less than 3 attempts the probability was 40%, 60% for 3 or less, and 20% for exactly 3. These results allow sellers to market their cards like the Charizard as a premium card. The discrete nature of the probability of it taking exactly 3 trials equaling 20% rather than 0 shows that there is no contradiction to continuous variable properties. An issue with this model though is that it assumes that if a trial like this were to be conducted at a sellers table, that there would be an equal selection probability of pulling the prize card.

Uniform (4.45):

- The results of the uniform problem indicate that most cards are accessibly priced. This is supported by a 40% probability of normalized prices being below £22 and a chance of exceeding £24 with a probability of 20%. This reflects a broader price distribution, with a skew towards lower to mid-range prices. This allows buyers to target deals that are below the midrange, while giving the sellers the ability to focus on higher volume sales of the lower-priced cards, maximizing their revenue in the end. While this distribution model does provide a baseline for pricing expectations, the uniform assumption oversimplifies the price distribution. Due to the different rarities and popularity amongst the community it could introduce natural price clustering. Figure 2 shows an unexpected result that could highlight the flaws of this model. Despite the “Reverse Holo” card type being considered rarer than the “Standard” card type, surprisingly the “Standard” has a higher average price. This suggests that there are factors beyond rarity that could affect the value of these cards, further proving the limited use of this type of distribution.

Joint Probability (5.1):

❖ The Joint PMF shows balanced probabilities ($\frac{2}{9} for\ Y1 = 1, Y2 = 1$) for listing assignments across all sellers, with F(1,0) = $\frac{1}{3}$ indicating a moderate chance of limited assignments to seller A and none to B. This shows that there is a competitive market where listings are distributed fairly among sellers, preventing dominance by any one seller. This also benefits collectors as it implies a more diverse source of buying options as no single seller can monopolize the market. This assumes equal assignment probabilities which may not be reliable when considering buyer reputation, online platform visibility, and other external factors.

Beta-Like (5.45):

❖ The results of the Beta-like problem based off question 5.45 from the textbook indicate that there is a concentration of cards in the mid-to-high price range. This is due to the probability of normalized prices being 64.8%, which exceeds 0.4. This is equivalent to about £360 relative to £ the maximum price of £899.99. When looking at the conditional probability of the lower-priced cards it leaves you 39.29% (from part “d” $P(Y \leq .4 \mid Y \leq\ .8)$)/ This shows that many are still below the mid-range pricing. This kind of distribution informs sellers and allows them to develop pricing strategies that encourage competitive pricing above the 0.4 threshold to align with market trends. This also shows collectors what cards to target below the threshold to save money. Once again, this type of distribution shows the skewed nature of the pricing but does not reflect external factors like market hype.

Conclusion

Throughout this study we applied various discrete and continuous distributions to analyze the dataset containing 25,598 Pokémon Trading Cards. This provided valuable insights into pricing dynamics, card acquisition, and listing frequency. By going through 6 statistical problems based off problems from *Mathematical Systems with Applications (*Wackerly et al.,2008), this analysis was able to model key market behaviors, offering guidance for those in the Pokémon TCG community (e.g. Collectors, sellers, and investors).

The Poisson question (3,122) was able to reveal the consistent demand for “Reverse Holo” cards with a listing of two or more having a probability of 99.27%. This information enabled sellers to optimize their inventories and collectors to identify low-activity periods to target a specific card. The binomial question (3.168) was able to display the challenges of authentication. With authenticators having a passing probability of 53.98%, it showcased the need to improve training for grading accuracy, especially due to high-value cards reaching prices of up to £899.99. The discrete uniform question (4.2) showed that in a seller’s special event of acquiring a “Charizard Reverse Holo” only had a probability of 20%. The uniform question based off (4.45) highlighted more accessible pricing within £20- £25, which contrasted with the skewed distribution which had a median of £0.89. This assisted collectors in making cost-effective purchases. The joint probability model (5.1) demonstrated the competitive seller market, which ensured a diverse and spread-out market. The beta-like model (5.45) revealed that there is a mid-to-high price concentration of 64.8% of cards being above £360, which helped sellers shape competitive pricing strategies.

While these findings can assist those within the market, these informed decisions must keep in mind the assumptions the models used. Those assumptions being constant rates, equal probabilities, and simplified distributions. Most prices, rarities, and inventory can drastically vary with time due to new releases, special promotions, community bias, and sellers’ reputation. Use models that can adapt and take into consideration how complex the Pokémon Trading Card market would be useful if a future study is to be done.

Work Cited

1. Lacey, Jack. *Pokémon Trading Cards Dataset*. Kaggle, 2024. https://www.kaggle.com/datasets/jacklacey/pokemon-trading-cards
2. Wackerly, Dennis D., William Mendenhall III, and Richard L. Scheaffer. *Mathematical Statistics with Applications*, 7th ed., Cengage Learning, 2011.