

Estimating the Market Efficiency of Prediction Markets

author

Contents

1 Why this question	2
2 Background	2
3 Model	2
4 Data Observations	2
5 Deriving the MLE of θ	4
6 Conclusion	5
7 Conclusions and Future Work	6

1 Why this question

Prediction markets are platforms where participants can trade contracts based on the outcome of future events.^[1] Prediction markets are said to be more accurate than expert opinions, as it magnifies the "wisdom of the crowd" to a larger scale. ^[1]

I chose this question as I was curious about the effectiveness and accuracy of analysing the prices of the contracts in events that we would like to know the outcome of.

2 Background

The prediction market platform we will retrieve data from in this report is Kalshi. Kalshi contracts are based on whether real-world events occur, with YES/NO options.

The sum of the price of the YES and NO contracts sum to \$ 1.00. If the outcome of the event aligns with the contract the user purchases, the user will receive \$ 1.00.^[2]

The price of each contract is determined by employing a dynamic pricing model that directly correlates with the market's perceived probability of a specific event occurring.^[3]

3 Model

The outcome of a contract is either it happens or it doesn't, so we can model this as a Bernoulli distribution.

Let y_1, \dots, y_n be a sample of outcomes related to Kalshi contracts. We can model these as realisations of a random variable Y_i for $i = 1, \dots, n$.

$$Y_i \sim \{\text{Bernoulli}(p_i) : p_i \in [0, 1]\}$$

Where $Y_i = 1$ is the event where the market resolves to YES and $Y_i = 0$ is the event where the market resolves to NO, and p_i is the probability the event occurs.

We will take the market price of a contract at each time period as x_i cents. Since $x_i \in [0, 1]$, It can be said that if the market were perfectly efficient, $x_i = p_i$. Therefore, we will model each p_i as such:

$$p_i = x_i^\theta$$

$\theta < 1$	Price too low
$\theta = 1$	Perfectly efficient
$\theta > 1$	Price too high

As $\theta < 1 \Rightarrow x_i < p_i$ and $\theta > 1 \Rightarrow x_i > p_i$

4 Data Observations

The markets we will collect data from are respectively:

- 1.How many Grammy awards will Sabrina Carpenter win?^[4]
- 2.How many Grammy awards will Lady Gaga win?^[5]
- 3.How many Grammy awards will Kendrick Lamar win?^[6]
- 4.How many Grammy awards will Bad Bunny win?^[7]

These markets are all for the 2025 Grammy awards, and were chosen due to their similarity, so it can be assumed that the distribution their contract prices and event probability will be similar.

We will collect the data from 12 November 2025 to 25 January 2026, using the 'minute' frequency

available on Kalshi. The actual date of the Grammy Awards was 1 February 2026. The date 25 January 2026 was chosen so that there would be less effects of volatility during the award show affecting the prices. Each of these markets have 7 different contracts, which respectively pay out based on the number of awards each performer wins, from exactly 0 to 9.

Performer	Number of Grammy Awards won
Sabrina Carpenter	0
Lady Gaga	2
Kendrick Lamar	5
Bad Bunny	3

x_i = Price of the contract at the given time

$$y_i = \begin{cases} 1 & \text{if the contract paid out} \\ 0 & \text{if the contract did not pay out} \end{cases}$$

5 Deriving the MLE of θ

As our model of Y_i is a Bernoulli distribution, the model has probability function:

$$P(Y_i = 1|x_i; \theta) = p_i = x_i^\theta$$

The pmf of the distribution is:

$$f(y_i; \theta) = (x_i^\theta)^{y_i} (1 - x_i^\theta)^{1-y_i}$$

Hence the likelihood function of θ is:

$$L(\theta) = \prod_{i=1}^n f(y_i; \theta) = \prod_{i=1}^n (x_i^\theta)^{y_i} (1 - x_i^\theta)^{1-y_i}$$

With log-likelihood function of:

$$\log(L(\theta)) = \sum_{i=1}^n \log[(x_i^\theta)^{y_i} (1 - x_i^\theta)^{1-y_i}]$$

Using the scipy function `scipy.optimize.minimize`, we can find the minimum of $-\log(L(\theta))$ and hence find the θ with the maximum $L(\theta)$

Market	MLE of θ
Sabrina Carpenter	1.0804
Lady Gaga	1.2949
Kendrick Lamar	1.1229
Bad Bunny	1.1976

The data shown when calculating the MLE of θ seems to show that the price x_i of the contracts are consistently overpriced. Which means, the probability of the event actually happening was lower than what the price would indicate on the Kalshi market. This indicates a systematic market inefficiency, possibly due to

6 Conclusion

We did some experiments ...

7 Conclusions and Future Work

From our experiments we can conclude that ...

References

- [1] Katelyn Peters. Prediction markets explained: Types, uses, and real-world examples. <https://www.investopedia.com/terms/p/prediction-market.asp>, 2026. Accessed: 2026-02-01.
- [2] Kalshi. Working with event contracts. <https://news.kalshi.com/p/what-are-event-contracts>, 2026. Accessed: 2026-02-01.
- [3] Kalshi. How are prices determined? <https://help.kalshi.com/markets/markets-101/how-are-prices-determined>, 2026. Accessed: 2026-02-02.
- [4] Kalshi. How many grammy awards will sabrina carpenter win? <https://kalshi.com/markets/kxgrammycountsab/award/kxgrammycountsab-26>, 2026. Accessed: 2026-02-02.
- [5] Kalshi. How many grammy awards will lady gaga win? <https://kalshi.com/markets/kxgrammycountlad/award/kxgrammycountlad-26>, 2026. Accessed: 2026-02-02.
- [6] Kalshi. How many grammy awards will kendrick lamar win? <https://kalshi.com/markets/kxgrammycountken/award/kxgrammycountken-26>, 2026. Accessed: 2026-02-02.
- [7] Kalshi. How many grammy awards will bad bunny win? <https://kalshi.com/markets/kxgrammycountbad/award/kxgrammycountbad-26>, 2026. Accessed: 2026-02-02.