Bayesian Inference of Honkai: Star Rail's Gacha Mechanics Using the Metalog Distribution

Haejin Cho
hcho435@gatech.edu

1 INTRODUCTION

This project investigates the gacha roll mechanics of *Honkai: Star Rail*, a turn-based strategy RPG developed by HoYoverse. Specifically, it examines whether the observed outcomes align with the game's officially stated probabilities and explores how the underlying probability distribution can be modeled using real user data.

1.1 Gacha Mechanics in *Honkai: Star Rail*

In *Honkai: Star Rail*, characters are categorized into three rarity levels: 3-star, 4-star, and 5-star. Among these, 5-star characters are the most desired, and every six or twelve weeks, a new 5-star character is introduced through a limited-time event called an "Event Warp." During these events, players engage in a gacha system — derived from the Japanese term "ガチャゲム"— where the outcome of each roll is determined by probability (Wikipedia, 2024). In general, players are most likely to obtain 3-star characters, occasionally 4-star characters, and only rarely a 5-star character. This project focuses specifically on modeling the probability of obtaining (or "pulling") a 5-star character.

1.2 Pity System

In the game, the developers explain the probability of obtaining a 5-star character as follows:

1. Warp Rate
   (a) The chances of obtaining 5-star characters and 4-star characters are as follows:
      i. During the event, the base chance of obtaining 5-star characters is 0.600%, and the consolidated chance including the guarantee is 1.600%.
      ii. Within 90 Warps, at least one 5-star character is guaranteed.
2. Boosted Rate
   (a) When you obtain a 5-star character during the Warp, there is a 50% chance

Figure 1—The featured 5-star character showcased in *Honkai: Star Rail*'s 3.2 version banner. Image source: ONE Esports

that it will be the featured 5-star character.

(b) There is also a 50% chance it will be any of the other obtainable 5-star characters of the current Warp phase, with equal chance among them.

(c) If the 5-star character you obtain is not the featured one, then the next 5-star character you obtain is guaranteed to be the featured character.

At first sight this can be confusing, so here are a few key points to understand the system:

1. **Pity System:** A "pity" refers to the number of pulls it takes to obtain a 5-star character. While the developer states that the base probability of pulling a 5-star character is 0.6%, this rate doesn't remain constant across all pulls. To prevent players from being indefinitely unlucky, the game introduces a *ceiling* (i.e. "hard pity")—a guaranteed 5-star character by the 90th pull. This means that the chance of pulling a 5-star on your 80th pull is significantly higher than on your first. Community data shows that hitting the *ceiling* is rare, suggesting the presence of a "soft pity" system—where the probability of a 5-star starts increasing **significantly** before the 90th pull (u/BanMidOnly, 2020). This ramp-up leads to an overall *consolidated* 5-star rate of 1.6%. However, the exact point where this probability begins to ramp up is officially unknown,

and one of our goals is to estimate this transition point using the real pull data.

2. **Featured Character Not Guaranteed:** Hitting a *ceiling* guarantees a 5-star character, but not necessarily the ***featured*** one—there's only a 50% chance. This means you have reached a "pity", and if you didn't obtained the featured one, then you lost "50/50". However, if you continue pulling, you eventually hit another pity—where the featured 5-star character is guaranteed, which is typically by the 180th pull (MisterMenPlays, 2023). In this project, we are focusing on the probability of any 5-star character—we won't differentiate between the featured 5-star characters and the regular 5-star characters.

2 RELATED WORK

In the HoYoverse player community, several analyses have been conducted to uncover the true mechanics behind the gacha system. Although these investigations primarily focus on *Genshin Impact*, the findings are consistent across other HoYoverse games like *Honkai: Star Rail* since they share the same gacha mechanism.

The analysis in u/Zacros, 2020 explores gacha behavior using data collected from high-spending users (i.e. *whales*). In total, 8,991 pulls were analyzed, obtaining 151 5-star characters. This results in an observed five-star character rate of approximately 1.68%, which is slightly above the advertised rate of 1.6%. Key findings include:

- **Soft Pity Trend**: A sharp increase in the probability of obtaining a 5-star character is observed beginning around the 75th pull, suggesting the presence of a "soft pity" mechanism.
- **Pull Distribution**: Most 5-star characters were acquired between the 75th and 80th pulls. 5-star characters can still be obtained beyond this range, but it is very rare.

Another post by u/BanMidOnly, 2020 attempts to reverse-engineer the soft pity system through simulation. It assumes a base 5-star rate of 0.6% until the 75th pull, and a significantly higher rate for pulls afterward, so that it reaches the overall rate of 1.6%. The probability distribution from this simulation is shown in Figure 2 . In this graph, the probability of obtaining a 5-star character peaks between the 76th and 80th pulls, then drops sharply through to the 90th pull—forming

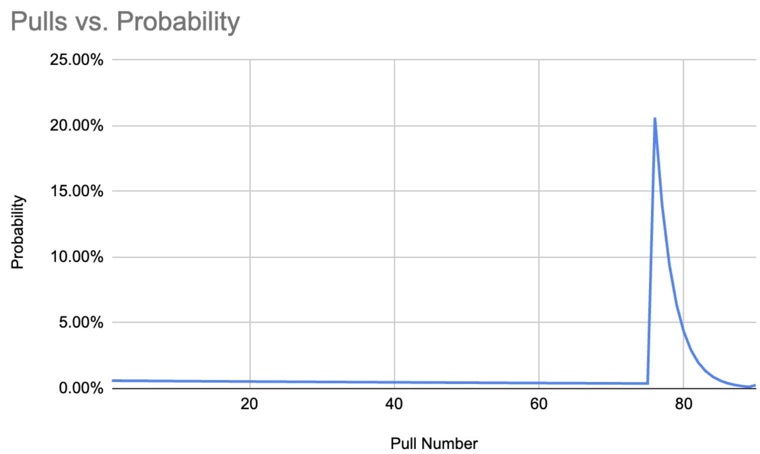

Figure 2—Simulated probability of obtaining 5-star character, reverse-engineered with the empirical assumptions. (Data source: u/BanMidOnly, 2020)

a left-skewed, unimodal distribution with a sharp peak. However, this analysis is based on the empirical assumptions and does not guarantee the true shape of the underlying probability distribution across pulls.

3 DATASET

3.1 Data Collecting Process

Since HoYoverse does not officially release user data, the only way to collect data is through directly access to individual user's pull histories. For this study, I scraped my own pull history in `warp_data_user1.csv` and that of another user whom I have access in `warp_data_user2.csv`, covering the past six months (earlier data was unavailable). The pull history data was automatically imported from the tool provided on the website Star Rail Station (starrailstation, 2025). Additionally, inspired by u/Zacros, 2020, I manually collected a third set of data consisting of 580 pulls from a Korean gaming Youtuber, Sameway, and saved the logs in `warp_data_user3_whale.csv` (Sameway, 2024).

3.2 Data Description

The files `warp_data_user1.csv` and `warp_data_user2.csv` each contain 7 columns:

- uid: A unique identifier for each individual pull.
- id: A unique ID assigned to each obtained item or character.

- rarity: The rarity level of the pulled item or character (3, 4, or 5).
- time: The timestamp of the pull in UTC.
- banner: The ID of the banner associated with the pull.
- type: The type of banner—1 (Standard), 11 (Character), or 12 (Weapon)
- manual: Indicates whether the data was manually imported.

`warp_data_user3_whale.csv` has 8 columns:

- uid: Same as above.
- pull_number: The absolute pull count across the banner session.
- rarity: same as above.
- type: same as above; this dataset contains only type 11 (Character).
- did_win_featured: Indicates whether the featured character was obtained.
- eidolon: The number of copies of the featured character acquired—E0 for 1 copy, E1 for 2 copies, and so on.
- pity_counter: The number of pulls it took to obtain a 5-star character.
- pity_counter_featured: The number of pulls it took to obtain the featured 5-star character.

The results from these three files are summarized in Table 1

File	Total Pulls	5-star	4-star	3-star	5-star Rate
`warp_data_user1.csv`	190	3	25	162	0.015789
`warp_data_user2.csv`	368	6	47	315	0.016304
`warp_data_user3_whale.csv`	580	10	82	488	0.017241

Table 1—Summary of pull results from user datasets

Figure 3 shows the 5-star pull results from user 3. Since these pulls are not common, the shape of the distribution is not immediately apparent, but there is a small cluster around the 75th pull. Figure 4 shows the cumulative probability of getting a 5-star, which starts to go up noticeably around the 75th pull.

4 BAYESIAN DATA ANALYSIS

4.1 Choosing the prior distribution for p

To model the likelihood of gacha pulls, several distributions can be considered, depending on how we define a "success":

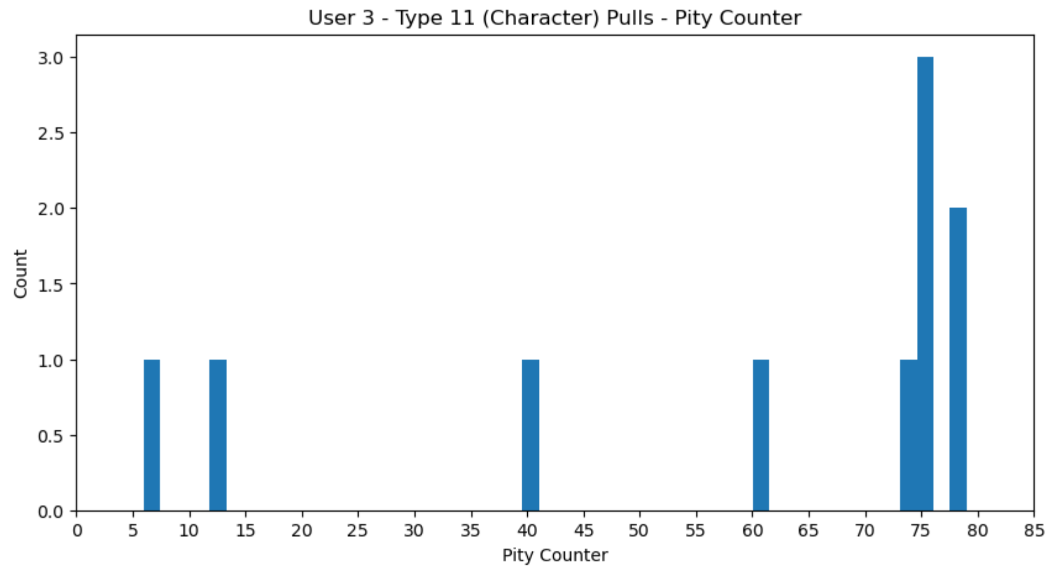

Figure 3—Distribution of 5-star pull counts from User 3, visualized as a histogram.

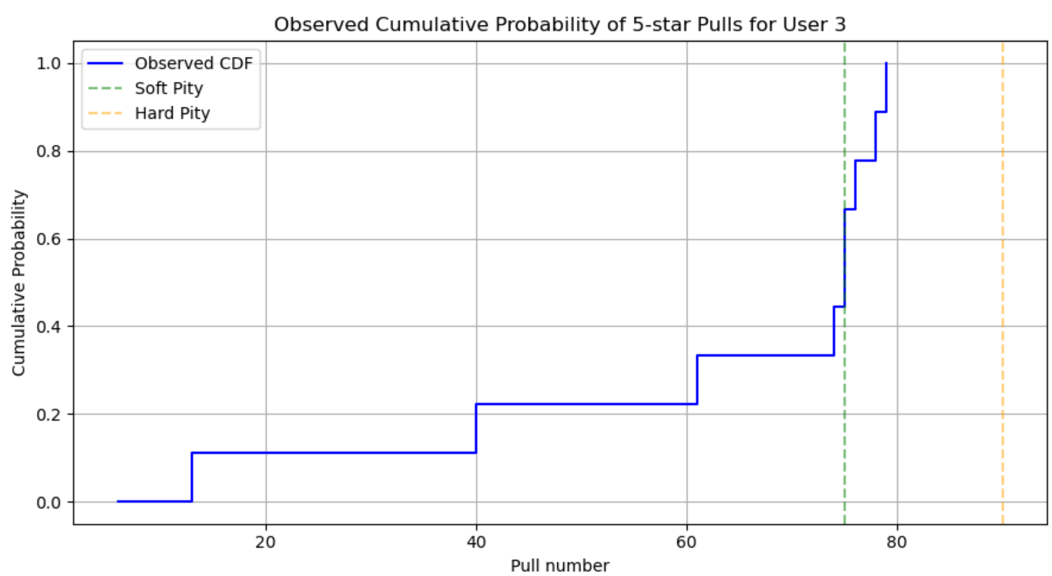

Figure 4—Cumulative probability of 5-star pulls for user3, with vertical dotted lines indicating soft and hard pity thresholds.

- A **Bernoulli distribution** can model the binary outcome of whether a pull results in a 5-star (win or not).
- A **Geometric distribution** can model the number of pulls until the first 5-star is obtained.
- A **Student's t-distribution** (or a Normal-like distribution) can be used to estimate the central tendency of successful pulls.

However, a key challenge in this analysis is determining an appropriate prior for the probability parameter `p`. Empirical observations from others suggest that the probability of obtaining a 5-star is very small until a certain point ("soft pity"), then increases sharply, and then drops significantly before reaching the "hard pity". Because the prior choice can significantly influence the posterior, especially in cases with limited or noisy data, it is important to choose a prior that reflects this behavior accurately.

Keelin, 2016 introduces three categories of probability distributions:

- **Type I**: Classical distributions derived from theoretical probability models (e.g., Normal, Exponential). Their legitimacy comes from established probabilistic foundations.
- **Type II**: Empirically motivated distributions that fit specific classes of data but lack a theoretical model.
- **Type III**: Flexible, general-purpose distributions designed to model almost any dataset, without relying on a predefined shape.

Given that the shape of the observed gacha pull data does not match any standard distribution, a Type III distribution can be a good choice for modeling the prior of `p`. In particular, I adopted the **metalog distribution** proposed by Keelin, 2016, which is a quantile-parameterized distribution that offers high flexibility and fits the data without the strong assumptions for the shape of the distribution. This allows the data to "speak for itself" while accommodating the complex, non-linear patterns observed in gacha mechanics. In this study, I forked the open-source `pymetalog` repository and modified it to be compatible with the latest version of `PyMC`, allowing for integration into a Bayesian modeling framework.

4.2 Model Setup

4.2.1 *Metalog Prior Distribution*

The metalog distribution provides a highly flexible framework for fitting continuous probability distributions directly from data, without requiring a specific parametric form. The core idea is to transform observed data into quantile space, then estimate the coefficients of a metalog basis expansion to approximate the cumulative distribution function.

The estimation process follows this flow (Keelin, 2016):

1. Start with raw data.
2. Assign probability values (e.g., empirical CDF).
3. Construct a quantile parameterization using basis functions defined by the metalog formulation.
4. Solve for the coefficients (denoted as $\mathbf{a} = (\alpha_1, \alpha_2, ..., \alpha_n)$) using Ordinary Least Squares.

To estimate the coefficient vector **a** in the OLS approach, the model solves this equation:

$$\mathbf{a} = [Y_n^T Y_n]^{-1} Y^T \mathbf{x}$$

where Y is the design matrix constructed from metalog basis functions of the input probabilities, **x** is the transformed data vector.

The equations for these basis functions and quantile expansion are detailed in the original paper Keelin, 2016 and on the official metalog website Keelin, 2016.

To use the metalog distribution, two key modeling decisions must be made:

- **Number of terms:** This determines the flexibility of the fitted distribution. More terms allow for more complex shapes, but the number of terms must not exceed the number of observations.
- **Boundedness:** Since gacha pulls are naturally constrained between 0 and 90, we apply a *upper/lower bounded* metalog to reflect the known support of the data.

Once the data is fitted, the model estimates a set of coefficients for each terms up to the user-specified maximum. In this case, I fitted the data with up to four terms, and the resulting coefficients are summarized in Table 2 below.

a_2	a_3	a_4
0.638790	1.271196	1.271196
0.847675	0.847675	0.481315
0.000000	-1.369050	-1.369050
0.000000	0.000000	2.111086

Table 2—Metalog distribution coefficients for up to four terms.

The shape of the fitted metalog probability distribution function for each term is shown below. Based on Figure 5, the 2-term model looks too simple and doesn't

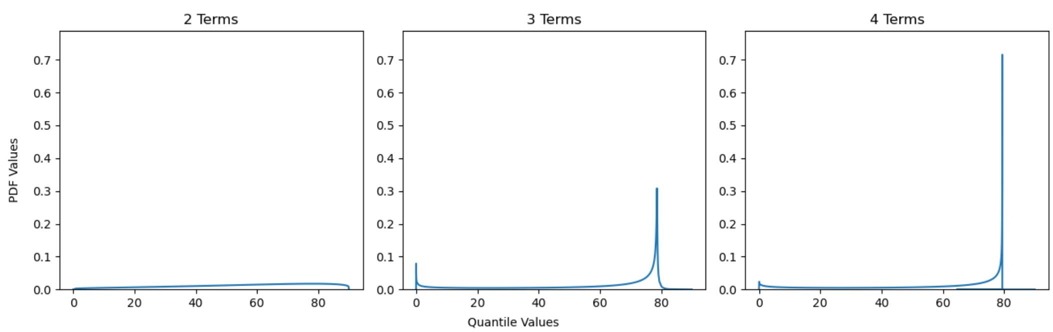

Figure 5—Visual representation of the fitted metalog probability density function.

capture the spike near the end. The 4-term model puts too much weight on pulls after 75, which might be too extreme. The 3-term model seems like a good middle ground—it captures the shape well without overfitting. The 3-term probability distribution function is again visualized in Figure 6 along with q0.1, q0.5, and q0.9 marks. Overall, the model provides a good fit, though the shape appears slightly off at the lower end of the pull range. Despite this issue, the distribution captures the overall trend well compared to the other two, so I am going to proceed with it.

4.2.2 *Metalog-Bernoulli*

First, I used the metalog prior combined with a Bernoulli likelihood. The data was transformed to include a binary indicator for whether each pull resulted in a 5-star character. After sampling 10,000 draws with 1,000 tuning steps, the posterior estimates are summarized as follows:

The mean probability is close to 1.6%, but the credible interval appears relatively

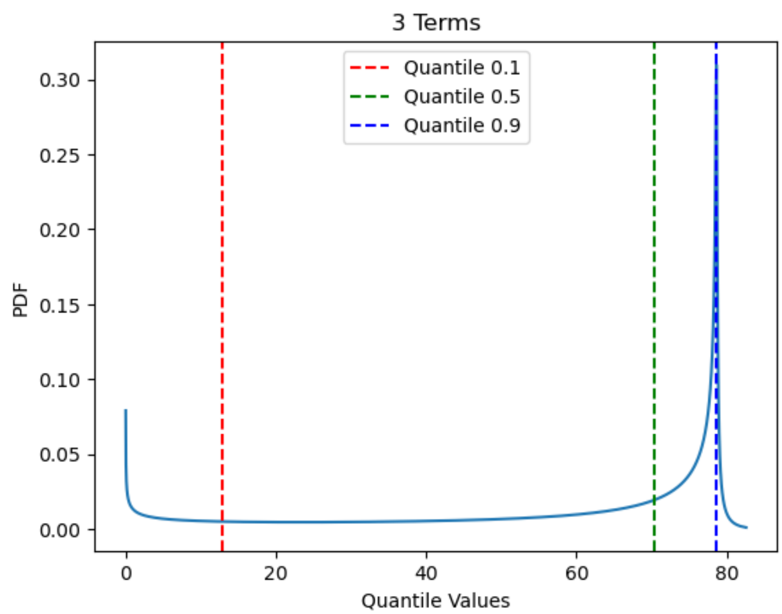

Figure 6—The visualization of the 3-term parameterized metalog probability distribution function.

Parameter	Mean	SD	HDI 2.5%	HDI 97.5%
θ	-4.113	0.324	-4.743	-3.490
p	0.017	0.005	0.007	0.028

Table 3—Posterior summary of θ and p parameters.

wide.

4.2.3 *Metalog-Geometric*

Next, I used the metalog prior with a Geometric likelihood. For this approach, I filtered the data to only include the pity counts when the 5-star characters were obtained. The estimated mean is quite different from the mean from user 3's data, and the credible interval for p seems quite wide.

4.2.4 *Metalog-Student's T*

Lastly, I used a Student's T distribution as a likelihood combined with the metalog prior. Although the estimated probability is lower than HoYoverse's advertised rate of 1.6%, the expected number of pulls (μ) seems to be the most reasonable.

Parameter	Mean	SD	HDI 2.5%	HDI 97.5%
θ	-4.067	0.322	-4.700	-3.429
p	0.018	0.005	0.008	0.029
μ	62.617	21.433	29.525	105.186

Table 4—Posterior summary of θ, p, and expected pull count $\mu = 1/p$.

Parameter	Mean	SD	HDI 2.5%	HDI 97.5%
θ	-4.304	0.049	-4.380	-4.205
τ	0.080	0.110	0.000	0.279
σ	6.460	5.144	1.034	15.980
p	0.013	0.001	0.012	0.015
μ	74.581	3.362	67.528	80.301

Table 5—Posterior summary for the metalog prior with the Student's T likelihood.

4.3 Hierarchical Modeling

4.3.1 *Model Setup*

Given that we have data from three distinct users, I explored a hierarchical modeling approach to capture user-specific variation. Since the number of 5-star pulls are all different across users, we expect both the posterior estimates and their uncertainty to be different. In this model, we estimate a separate parameter θ_{user} for each user, from which we derive user-specific values of p and μ.

A diagram of the model structure is shown in Figure 7.

4.3.2 *Observations*

User	μ	μ_HDI 2.5%	μ_HDI 97.5%	p	p_HDI 2.5%	p_HDI 97.5%
user1	69.30	26.48	92.79	0.0189	0.0095	0.0340
user2	74.92	64.14	84.37	0.0139	0.0115	0.0151
user3	74.00	65.73	80.77	0.0135	0.0123	0.0150

Table 6—Posterior summaries of expected pulls (μ) and probability (p) for each user from the hierarchy model.

Table 6 summarizes the expected posterior number of pulls (μ) and the corre-

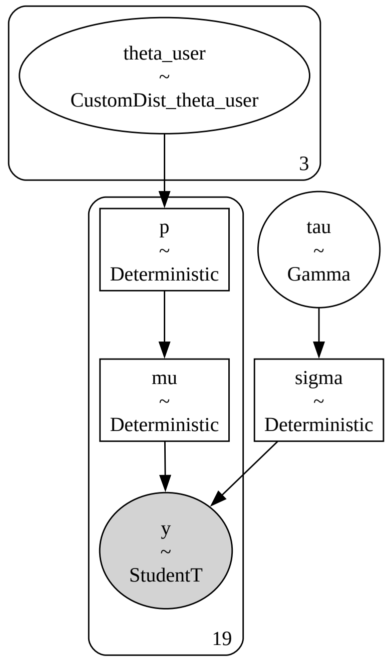

Figure 7—A graphviz digraph of the hierarchy model.

sponding success probability (p) for each user. User 1 appears to have a slightly higher probability of obtaining a 5-star character ($p = 0.0189$) compared to users 2 and 3. However, user 1 also has the wide credible interval, which suggests high uncertainty, making it difficult to confidently conclude that user 1 is objectively luckier. In contrast, user 3 has a more tightly estimated probability of $p = 0.0135$. Despite the probability is lower than HoYoverse's advertised rate of 1.6%, the expected number of pulls for user 2 and 3 seems to be aligning well with the known gacha mechanics, suggesting that the model captures the overall trend effectively.

Table 7 summarizes the posterior estimates for p and μ using a non-informative prior in the hierarchical model. Compared to the results with the metalog prior in Table 6, the estimated values for user 1 (showns as $p[0]$ and $\mu[0]$) are the most different. With the non-informative prior, user 1's estimated probability p is substantially higher ($p = 0.083$) and more uncertain, with a wide credible interval ranging from 0.010 to 0.647. This is likely due to small sample size of that user.

These results show a key benefit of the metalog prior: it incorporates known infor-

Parameter	Mean	SD	HDI 2.5%	HDI 97.5%
p[0]	0.083	0.196	0.010	0.647
p[1]	0.021	0.061	0.011	0.020
p[2]	0.014	0.001	0.012	0.016
μ[0]	56.714	28.034	0.193	84.375
μ[1]	72.202	12.808	44.518	84.854
μ[2]	72.501	5.384	61.132	80.310

Table 7—Posterior summary statistics for p and μ parameters with the non-informative prior distribution.

mation derived from the data, helping to stabilize inference when observations are limited. Therefore, when prior information is available—such as the known 1.6% average 5-star rate—it should be leveraged to improve model robustness and interpretability.

5 CONCLUSION

This project explored the potential of using quantile-parameterized distributions—specifically the metalog distribution—as a prior for estimating the probability structure of gacha mechanics. Compared to non-informative priors, the metalog prior provided more robust inference, particularly in cases with limited data. However, limitations remain: the sample size is still relatively small, some sampled values exceed the known ceiling of 90 pulls, and the model couldn't properly estimate the advertised 1.6% rate. Future work may address these issuse by collecting more data through user communities or exploring alternative prior and likelihood distributions.

6 REFERENCES

[1] Keelin, Thomas W (2016a). *The Metalog Distributions*. Accessed: April 19, 2025. URL: http://www.metalogdistributions.com/equations.html.

[2] Keelin, Thomas W (2016b). "The metalog distributions". In: *Decision Analysis* 13.4, pp. 243–277.

[3] MisterMenPlays (2023). *Pity system explained: 5-star guaranteed in Honkai: Star Rail*. Accessed: 2025-04-19. URL: https://www.mistermenplays.com/honkaistarrail/news/pity-system-explained-5-star-guaranteed.

[4] Sameway (2024). *샘(웨)의 인생 첫 풀돌 스타레일 "반디" 뽑기*. Accessed: April 19, 2025. URL: https://www.youtube.com/watch?v=furERjNnujI&t=3s.

[5] starrailstation (2025). *Star Rail Station — Warp Record Viewer*. Accessed: April 19, 2025. URL: https://starrailstation.com/en/warp#global.

[6] u/BanMidOnly (2020). *The 5% Rate is Not Uniform (0.6%) – There is a Soft Pity System in Place*. Posted on Reddit. Accessed: 2025-04-19. URL: https://www.reddit.com/r/Genshin_Impact/comments/jo9d9d/the_5_rate_is_not_uniform_06_there_is_a_soft_pity/.

[7] u/Zacros (2020). *Whale Watching Logs 2: The Blue Whale*. Posted on Reddit. Accessed: 2025-04-19. URL: https://www.reddit.com/r/Genshin_Impact/comments/jod9o4/whale_watching_logs_2_the_blue_whale/.

[8] Wikipedia (2024). *Gacha game — Wikipedia, The Free Encyclopedia*. Accessed: 2025-04-19. URL: https://en.wikipedia.org/wiki/Gacha_game.