A Flexible Generalized Probability Core and Quantitative Strategy Analysis for Game Design

Angela Li
Applied Mathematics & Physics
Stony Brook University
Stony Brook, NY
angela.li.4@stonybrook.edu

Zhengnan Li
Katz School of Science and Health
Yeshiva University
New York, NY
zli6@mail.yu.edu

David Li
Katz School of Science and Health
Yeshiva University
New York, NY
david.li@yu.edu

***Abstract*—Level design is a critical component in creating engaging and balanced computer games, where difficulty and complexity must be carefully adjusted to enhance player experience. This paper proposes a flexible framework for strategy game level design using probability theory. By modifying the parameters of a generalized probability model, game designers can produce a wide range of levels that require unique strategies, avoiding repetition while maintaining challenge. A prototype demonstrating this approach, along with probability theory-based derivations and strategy analyses, is presented. This framework has potential applications in various game genres, contributing to dynamic and unpredictable gameplay experiences.**

***Index Terms*—Normal Distribution, Sampling, Central Limit Theorem, Probability Density Function (PDF), Cumulative Distribution Function (CDF)**

I. Introduction

Probability is a critical yet complex element in modern game design, particularly in strategy games where it enhances unpredictability and player engagement. By introducing uncertainty into gameplay, probability can create more dynamic and immersive experiences, preventing the game from becoming monotonous or overly predictable [1]. However, probability also presents a unique set of challenges, as it can lead to frustration when outcomes seem unfair or disproportionately influenced by luck. This paper addresses these issues by exploring the role of probability in game design and proposing a structured approach to balance its use effectively.

A. The Role of Probability in Game Design

In many games, probability governs the outcomes of key decisions, influencing factors such as combat success, loot drops, or character development. It plays an especially important role in strategy games, where it introduces an element of chance that keeps gameplay fresh and forces players to adapt to unpredictable outcomes [2], [3]. For example, in games like Star Vikings and XCOM, probability determines the success or failure of actions such as attacks or abilities. Players must not only plan based on their strategy but also account for the possibility of failure due to chance, which adds depth and variability to the gameplay experience.

B. The Benefits of Probability

One of the greatest advantages of incorporating probability into game design is its ability to keep players guessing. By introducing uncertainty, probability prevents the game from becoming repetitive and predictable. Even with a well-executed plan, a low-probability event can drastically change the outcome of a game, forcing players to think on their feet. This dynamic aspect can lead to heightened engagement, as players must continuously adjust their strategies. For instance, in Star Vikings, players can upgrade characters to increase the chance of triggering secondary effects. These random elements can lead to unexpected yet exciting results, keeping the game challenging and engaging [4].

C. The Challenges of Probability

Despite its benefits, probability can also frustrate players, particularly when the outcomes feel unfair or out of their control. Missing a critical shot with a high success rate or losing a battle due to bad luck can leave players feeling powerless, which can diminish their enjoyment. This issue is especially prevalent in games where success hinges heavily on chance, as seen in XCOM, where missing a 90% chance shot can have significant consequences. Players often interpret such outcomes as evidence that the game is ”rigged” against them, which can lead to disengagement and dissatisfaction [5]. The challenge for game designers is to balance probability in a way that maintains unpredictability without alienating players. [6], [7]

D. Communicating Probability to Players

One of the key challenges in implementing probability in games is ensuring that players understand how it works and how it affects their decisions. Probability is inherently abstract, and players often struggle to grasp its nuances. Games like XCOM address this by providing clear, transparent feedback on the likelihood of success for each action, allowing players to make informed decisions. In contrast, games like Darkest Dungeon and Renowned Explorers International Society have struggled with this aspect, as they fail to clearly communicate how various factors impact probability. When players cannot see how probability is affecting their choices, they may feel that their success or failure is arbitrary, further contributing to frustration.

E. The Need for Balance

Ultimately, the use of probability in game design is a balancing act [9]. While it can create tension and excitement, it can also lead to frustration if not implemented carefully. Good game design ensures that players have ways to mitigate the impact of chance or at least feel that their success is not entirely out of their hands. Designers must decide how much influence probability should have over a game's outcomes and ensure that the mechanics support both unpredictability and player agency [8].

F. Contributions of this Paper

This paper presents a generalized framework for using probability in game design, particularly in strategy games. By adjusting parameters within a probabilistic model, developers can create diverse levels with unique challenges that avoid repetition. This framework allows for both flexibility in design and the creation of levels that require distinct winning strategies. A prototype is provided to demonstrate the practicality of this approach, along with a detailed derivation of the probability model and an analysis of several possible winning strategies. This method provides a scalable solution that can be adapted to a wide range of game genres, ensuring that probability enhances gameplay rather than detracts from it.

II. The Probability Core for Game Design

A. The Set of Available Actions

In action role-playing games (RPGs), players exert direct control over a character's movements and combat actions, engaging in real-time battles by pressing one or more attack buttons to execute offensive maneuvers. These games typically feature hack-and-slash mechanics reminiscent of arcade-style combat, although some incorporate elements from fighting, brawling, or shooting genres. When shooting mechanics are integrated, the game is classified as a "role-playing shooter," a subgenre of action RPGs that blends traditional role-playing elements with shooter dynamics.

In a simplified scenario, each player has a set of available actions. The set of n available actions of player i can be represented as,

$$a_{(i,j)} \in \{a_{(i,1)}, a_{(i,2)}, a_{(i,3)}, ..., a_{(i,n)}\} \tag{1}$$

where

- i represents the ith player in this game. For example, in a two-player game, $i \in \{1, 2\}, \quad i \in \mathbb{Z}$.
- n represents the total n available actions of the ith player in this game.
- j represents the jth action that the ith player chooses. $1 \leq j \leq n, \quad j \in \mathbb{Z}$.

B. The Utility Function of Each Action

Each of the available actions has a utility function. The "utility" represents the damage this action will cause. The values of a utility function follow a normal distribution with the probability density function,

$$U_{(i,j)} = U(a_{(i,j)}) \tag{2}$$

$$p(U_{(i,j)}) = \frac{1}{\sqrt{2\pi\sigma^2_{(i,j)}}} \exp\left(-\frac{(U_{(i,j)} - \mu_{(i,j)})^2}{2\sigma^2_{(i,j)}}\right) \tag{3}$$

where

- $p(U_{(i,j)})$ is the probability density function of the utility function on the jth action by the ith player in this game, assuming the utility function $U_{i,j}$ follows the normal distribution.
- $\sigma_{(i,j)}$ represents the standard deviation of the utility of the action $a_{(i,j)}$.
- $\mu_{(i,j)}$ represents the mean of the utility of the action $a_{(i,j)}$.

III. Design Proposal 1: $\mu_{(i,j)}$ in Uniform Distribution

A. The Uncertainty in Game Design

Although the utility function has the theoretical model as shown in equation (3), the mean $\mu_(i,j)$ and the standard deviation $\sigma_{(i,j)}$ are unknown at the beginning of each game. The fun part of the game for the player is to estimate the mean $\mu_{(i,j)}$ and the standard deviation $\sigma_{(i,j)}$ by playing the game and figure out the best winning strategy as quick as possible.

For simplicity, in each mission, the game generates the mean $\mu_{(i,j)}$ by a uniform distribution between μ_{min} and μ_{max}.

$$\mu_{(i,j)} \sim \mathcal{U}(\mu_{\text{min}}, \mu_{\text{max}}) \tag{4}$$

where

- μ_{min} is the pre-defined lower bound of all μ values.
- μ_{max} is the pre-defined upper bound of all μ values.

Similarly, the game generates the $\sigma_{(i,j)}$ by another uniform distribution between σ_{min} and σ_{max}.

$$\sigma_{(i,j)} \sim \mathcal{U}(\sigma_{\text{min}}, \sigma_{\text{max}}) \tag{5}$$

where

- σ_{min} is the pre-defined lower bound of all σ values.
- σ_{max} is the pre-defined upper bound of all σ values.

Since the player doesn't have prior knowledge of the generated $\mu_{i,j}$ and $\sigma_{i,j}$, the player must play the game and estimate the values.

B. The Distribution of the Largest $\mu_{(i,j)}$

The player doesn't know which action has the best utility, but the player can estimate the theoretical largest $\mu_{(i,j)}$ and its distribution.

If the game uniformly chooses n random values as $\mu_{(i,j)}$ from the range $[\mu_{\text{min}}, \mu_{\text{max}}]$, the player can calculate the distribution of the largest $\mu_{(i,j)}$ value by using the cumulative distribution function (CDF) and probability density function (PDF) for order statistics, particularly for the largest value in a sample.

- CDF of a Uniform Distribution:
  The CDF of a uniform distribution $\mathcal{U}(\mu_{\text{min}}, \mu_{\text{max}})$ is given by:

$$P(\mu) = p(x \leq \mu) = \frac{\mu - \mu_{\text{min}}}{\mu_{\text{max}} - \mu_{\text{min}}} \tag{6}$$

  where $\mu_{\text{min}} \leq \mu \leq \mu_{\text{max}}$.
  This represents the probability that a random variable x chosen from this range is less than or equal to μ.
- CDF of the Largest $\mu_{(i,j)}$:
  Let $\mu_{(i,1)}, \mu_{(i,2)}, \mu_{(i,3)}, \ldots, \mu_{(i,n)}$ be independent and identically distributed (i.i.d.) random variables drawn from a uniform distribution $\mathcal{U}(\mu_{\text{min}}, \mu_{\text{max}})$. Denote the largest $\mu_{(i,j)}$ as:

$$\mu_n = \max(\mu_{(i,1)}, \mu_{(i,2)}, \mu_{(i,3)}, \ldots, \mu_{(i,n)}) \tag{7}$$

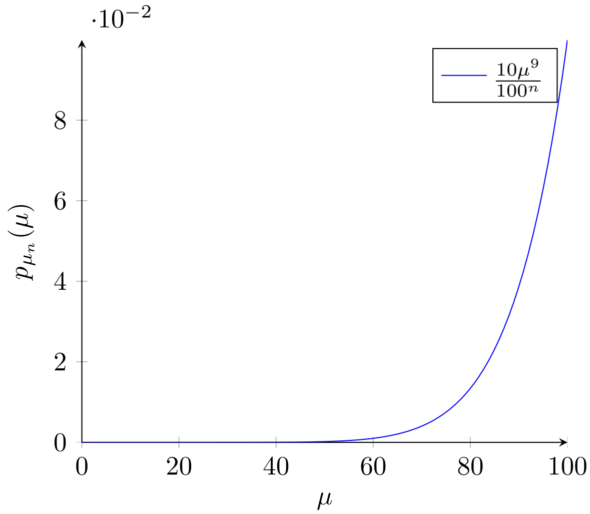
  where μ_n is the largest number in n random values of $\mu_{(i,j)}$.
  The CDF of the largest value μ_n is the probability that all n random values are less than or equal to μ_n. For independent and identically distributed (i.i.d.) random variables, the CDF of the largest $\mu_{(i,j)}$ is:

$$\begin{aligned} P_{\mu_n}(\mu) &= p(\mu_n \leq \mu) \\ &= (p(x \leq \mu))^n \\ &= \left(\frac{\mu - \mu_{\text{min}}}{\mu_{\text{max}} - \mu_{\text{min}}}\right)^n \end{aligned} \tag{8}$$

- PDF of the Largest $\mu_{(i,j)}$:
  To find the probability density function (PDF) of the largest $\mu_{(i,j)}$, one can differentiate the CDF with respect to μ to get the PDF of the largest $\mu_{(i,j)}$:

$$\begin{aligned} p_{\mu_n}(\mu) &= \frac{d}{d\mu}\left[\left(\frac{\mu - \mu_{\text{min}}}{\mu_{\text{max}} - \mu_{\text{min}}}\right)^n\right] \\ &= n\left(\frac{\mu - \mu_{\text{min}}}{\mu_{\text{max}} - \mu_{\text{min}}}\right)^{n-1} \cdot \frac{1}{\mu_{\text{max}} - \mu_{\text{min}}} \\ &= \frac{n(\mu - \mu_{\text{min}})^{n-1}}{(\mu_{\text{max}} - \mu_{\text{min}})^n} \end{aligned} \tag{9}$$

  This PDF shows that the distribution is highly skewed, with most of the probability mass concentrated near μ_{max}, since the largest $\mu_{(i,j)}$ in a sample of n values is more likely to be close to the upper bound of the range μ_{max}. The PDF increases rapidly as μ approaches μ_{max}, meaning the largest value μ_n in the sample is typically close to μ_{max}.
  The following plot shows the PDF curve when $n = 10$ and $\mu_{\text{max}} - \mu_{\text{min}} = 100$.



C. The Mean of the Largest $\mu_{(i,j)}$

We want to find $\mathbb{E}[\mu_n]$, the expectation of the maximum, as shown in (7). The mean of the largest $\mu_{(i,j)}$ can be calculated using integration:

$$\begin{aligned} \mathbb{E}[\mu_n] &= \int_{\mu_{\text{min}}}^{\mu_{\text{max}}} \mu p_{\mu_n}(\mu) d\mu \\ &= \int_{\mu_{\text{min}}}^{\mu_{\text{max}}} \mu \frac{n(\mu - \mu_{\text{min}})^{n-1}}{(\mu_{\text{max}} - \mu_{\text{min}})^n} d\mu \\ &= n\left(\frac{\mu_{\text{max}} - \mu_{\text{min}}}{n+1} + \frac{\mu_{\text{min}}}{n}\right) \\ &= \frac{n\mu_{\text{max}} + \mu_{\text{min}}}{n+1} \end{aligned} \tag{10}$$

The details of the derivation are included in the Appendix A of this paper. This integral would give the average value of the largest $\mu_{(i,j)}$ from a sample of n. The mode of the distribution is close to μ_{max} because the PDF is heavily skewed toward the maximum possible value in the range.

Using the formula (9), when $n = 10$, $\mu_{\text{max}} = 100$, and $\mu_{\text{min}} = 0$, the expected value (mean) of the largest $\mu_{(i,j)}$ in a sample of 10 uniformly randomly chosen numbers from the range [0, 100] is approximately 90.91, which is very close to $\mu_{\text{max}} = 100$.

D. The Variance of the Largest $\mu_{(i,j)}$

The distribution is highly concentrated near the maximum, with smaller variance as μ approaches μ_{max}. In practical terms, this means most samples will have a largest number close to μ_{max}, but there will be occasional outliers where the largest number is significantly lower.

The variance of the largest $\mu_{(i,j)}$ can be calculated using integration:

$$
\begin{aligned}
\mathbb{V}[\mu_n] &= \mathbb{E}[\mu_n^2] - \mathbb{E}[\mu_n]^2 \\
&= \int_{\mu_{\min}}^{\mu_{\max}} \mu^2 p_{\mu_n}(\mu) d\mu - \mathbb{E}[\mu_n]^2 \\
&= \int_{\mu_{\min}}^{\mu_{\max}} \mu^2 \frac{n(\mu - \mu_{\min})^{n-1}}{(\mu_{\max} - \mu_{\min})^n} d\mu - \mathbb{E}[\mu_n]^2 \\
&= n\left[\frac{(\mu_{\max} - \mu_{\min})^2}{n+2} + 2\mu_{\min}\frac{\mu_{\max} - \mu_{\min}}{n+1} + \mu_{\min}^2 \frac{1}{n}\right] \\
&\quad - \mathbb{E}[\mu_n]^2 \\
&= n\left[\frac{(\mu_{\max} - \mu_{\min})^2}{n+2} + 2\mu_{\min}\frac{\mu_{\max} - \mu_{\min}}{n+1} + \mu_{\min}^2 \frac{1}{n}\right] \\
&\quad - n^2\left(\frac{\mu_{\max} - \mu_{\min}}{n+1} + \frac{\mu_{\min}}{n}\right)^2 \\
&= \frac{n(\mu_{\max} - \mu_{\min})^2}{n+2} - \frac{n^2(\mu_{\max} - \mu_{\min})^2}{(n+1)^2}
\end{aligned} \tag{11}
$$

The details of the derivation are included in the Appendix B of this paper.

Using the formula (11), when $n = 10$, $\mu_{\max} = 100$, and $\mu_{\min} = 0$, the variance of the largest $\mu_{(i,j)}$ in a sample of 10 uniformly chosen random numbers from the range $[0, 100]$ is approximately 68.87, i.e., the standard deviation is 8.30.

The variance indicates that while most largest $\mu_{(i,j)}$ will be near $\mu_{\max}$, there is still some variability in the outcomes.

IV. Design Proposal 2: $\mu_{(i,j)}$ in Normal Distribution

The game can also generate the mean $\mu_{(i,j)}$ by a normal distribution with the mean μ_μ and the standard deviation σ_μ.

$$\mu_{(i,j)} \sim \mathcal{N}(\mu_\mu, \sigma_\mu^2) \tag{12}$$

If the game randomly chooses n values as $\mu_{(i,j)}$ from the distribution $\mathcal{N}(\mu_\mu, \sigma_\mu^2)$, the player can find the distribution of the largest $\mu_{(i,j)}$ value in the similar approach.

- Setup and Rescaling the Normal Distribution
  Let $\mu_{(i,1)}, \mu_{(i,2)}, \mu_{(i,3)}, \ldots, \mu_{(i,n)}$ be independent and identically distributed (i.i.d.) random variables drawn from a normal distribution $\mathcal{N}(\mu_\mu, \sigma_\mu^2)$. Denote the largest $\mu_{(i,j)}$ as:

  $$\mu_n = \max(\mu_{(i,1)}, \mu_{(i,2)}, \mu_{(i,3)}, \ldots, \mu_{(i,n)}) \tag{13}$$

  The goal is to find $\mathbb{E}[\mu_n]$, the expectation of the maximum. Instead of working directly with $\mathcal{N}(\mu_\mu, \sigma_\mu^2)$, it's convenient to rescale and standardize the normal variables.
  Let:

  $$Z_{(i,j)} = \frac{\mu_{(i,j)} - \mu_\mu}{\sigma_\mu} \tag{14}$$

  Now $Z_{(i,j)} \sim \mathcal{N}(0, 1)$, and the maximum of the standardized variables is:

  $$Z_n = \max(Z_{(i,1)}, Z_{(i,2)}, Z_{(i,3)}, \ldots, Z_{(i,n)}) \tag{15}$$

  Thus, we can write the maximum of the original normal variables as:

  $$\mu_n = \mu_\mu + \sigma_\mu Z_n \tag{16}$$

- The cumulative distribution function (CDF) of the maximum Z_n can be expressed in terms of the CDF of an individual standard normal variable Z.
  Let $P_Z(z)$ be the CDF of a standard normal variable:

  $$P_Z(z) = p(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \tag{17}$$

  The CDF of Z_n, the maximum of n i.i.d. standard normal variables, is:

  $$
  \begin{aligned}
  P_{Z_n}(z) &= P(Z_n \le z) \\
  &= P(Z_{(i,1)} \le z, Z_{(i,2)} \le z, \ldots, Z_{(i,n)} \le z) \\
  &= [P_Z(z)]^n
  \end{aligned} \tag{18}
  $$

  The probability density function (PDF) of Z_n is obtained by differentiating the CDF:

  $$p_{Z_n}(z) = \frac{d}{dz} P_{Z_n}(z) = n[P_Z(z)]^{n-1} p_Z(z) \tag{19}$$

  where $p_Z(z)$ is the PDF of a standard normal distribution:

  $$p_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{20}$$

- The Mean of the Largest $\mu_{(i,j)}$
  To find the expectation $\mathbb{E}[Z_n]$, we use the following integral:

  $$\mathbb{E}[Z_n] = \int_{-\infty}^{\infty} z \cdot p_{Z_n}(z) dz \tag{21}$$

  Substituting the expression for $p_{Z_n}(z)$:

  $$\mathbb{E}[Z_n] = \int_{-\infty}^{\infty} z \cdot n[P_Z(z)]^{n-1} \cdot p_Z(z) dz \tag{22}$$

  This integral doesn't have a simple closed-form solution, but it can be approximated using asymptotic methods for large n, then the distribution of Z_n converges to a Gumbel distribution after appropriate scaling. The approximation for the expectation of Z_n is asymptotically given by:

  $$\mathbb{E}[Z_n] \approx \sqrt{2 \ln n} - \frac{\ln(\ln n)}{2\sqrt{2 \ln n}} \tag{23}$$

  Transform back to original scale by the equation (16), the expectation of the largest $\mu_{(i,j)}$ is approximately:

  $$\mathbb{E}[\mu_n] \approx \mu_\mu + \sigma_\mu \cdot \left(\sqrt{2 \ln n} - \frac{\ln(\ln n)}{2\sqrt{2 \ln n}}\right) \tag{24}$$

- The Variance of the Largest $\mu_{(i,j)}$

With Gumbel distribution, the approximation for the expectation of Z_n^2 is given by:

$$\mathbb{E}[Z_n^2] \approx 2\ln n - \frac{2\ln(\ln n)}{\sqrt{2\ln n}} \tag{25}$$

For large n, the variance of Z_n from n standard normal variables is:

$$\begin{aligned}\mathbb{V}[Z_n] &= \mathbb{E}[Z_n^2] - \mathbb{E}[Z_n]^2 \\ &\approx \left[2\ln n - \frac{2\ln(\ln n)}{\sqrt{2\ln n}}\right] - \left[\sqrt{2\ln n} - \frac{\ln(\ln n)}{2\sqrt{2\ln n}}\right]^2 \\ &= \ln(\ln n) - \frac{2\ln(\ln n)}{\sqrt{2\ln n}} - \frac{\ln^2(\ln n)}{8\ln n}\end{aligned} \tag{26}$$

Transform back by the equation (16), the variance of the largest $\mu_{(i,j)}$ is,

$$\mathbb{V}[\mu_n] \approx \sigma_\mu^2 \cdot \left[\ln(\ln n) - \frac{2\ln(\ln n)}{\sqrt{2\ln n}} - \frac{\ln^2(\ln n)}{8\ln n}\right] \tag{27}$$

These approximations hold well for large n but are also reasonably good for smaller n in the game design and strategy analysis.

V. The Strategy Analysis

For each action the player selects, the resulting utility is a random sample from the utility function of that action, which follows a normal distribution (3). The ith player aims to quickly identify the best action from the set $\{a_{(i,1)}, a_{(i,2)}, a_{(i,3)}, ..., a_{(i,n)}\}$, given the time constraints of each mission, in order to maximize the likelihood of completing the mission before the timeout.

In order to identify the best action, the player can gain experience, i.e., collect samples, by randomly choosing actions. Since each action's outcome is from its utility function based on random sampling, the observed utility samples of each action can be used to estimate the mean $\mu_{(i,j)}$ and its standard deviation $\sigma_{(i,j)}$.

The estimated $\mu_{(i,j)}$ is the average value of the observed utility values from the action $a_{(i,j)}$,

$$\widehat{\mu_{(i,j)}} \cong \overline{U_{(i,j)}} = \frac{1}{N}\sum_{k=1}^{N} U_{(i,j)} \tag{28}$$

where N is the sample size of the action $a_{(i,j)}$.

The variance of the sample mean is,

$$\mathbb{V}[\widehat{\mu_{(i,j)}}] = \frac{\sigma_{(i,j)}^2}{N} \tag{29}$$

Since the player doesn't know the standard deviation $\sigma_{(i,j)}$ of the utility function $U_{(i,j)}$, the player can estimate it by the sample variance as,

$$\sigma_(i,j)^2 \cong \mathbb{V}[U_{(i,j)}] = \frac{1}{N-1}\sum_{k=1}^{N}(U_{(i,j)} - \overline{U_{(i,j)}})^2 \tag{30}$$

Then the variance of the sample mean is,

$$\begin{aligned}\mathbb{V}[\widehat{\mu_{(i,j)}}] &= \frac{\sigma_(i,j)^2}{N} \\ &\cong \frac{1}{N(N-1)}\sum_{1}^{N}(U_{(i,j)} - \overline{U_{(i,j)}})^2\end{aligned} \tag{31}$$

Based on the central limit theorem, when N is larger, $\mathbb{V}[\widehat{\mu_{(i,j)}}]$ is smaller, and then the estimation $\widehat{\mu_{(i,j)}}$ in (28) is more accurate.

Then there are several strategies.

A. The Greedy Strategy

The player wants to identify the optimal action.

- Theoretically, the best action must be the one with the highest mean $\mu_{(i,j)}$.
  Once the player collects plenty of samples, the estimation of $\mu_{(i,j)}$ will become more accurate. Then the player can choose the optimal action with the highest $\mu_{(i,j)}$.

$$a^*_{(i,j)} = \arg\max_{a_{(i,j)}} \widehat{\mu_{(i,j)}} \tag{32}$$

  The player will use the optimal action $a^*_{(i,j)}$ to complete the rest of the mission in the game.
- In reality, the player doesn't have unlimited amount of time to collect plenty of samples, so the estimation of $\mu_{(i,j)}$ won't be accurate.

$$\mu_{(i,j)} \sim \mathcal{N}(\widehat{\mu_{(i,j)}}, \mathbb{V}[\widehat{\mu_{(i,j)}}]) \tag{33}$$

  Because of the uncertainty, the player will never be sure which action is the best one. However, the player can estimate the probability of an action being the best action. Consider another action's estimation of $\mu_{(i,k)}$ with the PDF,

$$\mu_{(i,k)} \sim \mathcal{N}(\widehat{\mu_{(i,k)}}, \mathbb{V}[\widehat{\mu_{(i,k)}}]) \tag{34}$$

  Then the distribution of $\mu_{(i,j)} - \mu_{(i,k)}$ is,

$$(\mu_{(i,j)} - \mu_{(i,k)}) \sim \mathcal{N}(\widehat{\mu_{(i,j)}} - \widehat{\mu_{(i,k)}}, \mathbb{V}[\widehat{\mu_{(i,j)}}] + \mathbb{V}[\widehat{\mu_{(i,k)}}]) \tag{35}$$

  The probability of $\mu_{(i,j)} \geq \mu_{(i,k)}$ is given by,

$$P(\mu_{(i,j)} - \mu_{(i,k)} \geq 0) = \Phi\left(\frac{\widehat{\mu_{(i,j)}} - \widehat{\mu_{(i,k)}}}{\sqrt{\mathbb{V}[\widehat{\mu_{(i,j)}}] + \mathbb{V}[\widehat{\mu_{(i,k)}}]}}\right) \tag{36}$$

  where Φ is the CDF of the standard normal distribution. The player will set a probability threshold P_τ. In practice, this threshold could be 95% for greater certainty in action selection, though it requires more time for sampling. Alternatively, it could be set lower, such as 60%, if the player is willing to take on more risk for faster action selection.
  Therefore, the player will choose the action $a_{(i,j)}$ if and only if,

$$\Phi\left(\frac{\widehat{\mu_{(i,j)}}-\widehat{\mu_{(i,k)}}}{\sqrt{\mathbb{V}[\widehat{\mu_{(i,j)}}]+\mathbb{V}[\widehat{\mu_{(i,k)}}]}}\right)\geq P_\tau \quad \text{for any } k\in[1,n], k\neq j \tag{37}$$

B. The Non-greedy Strategy

The primary goal is mission completion, so identifying the optimal action is often unnecessary. When the mission is under a time constraint, searching for the optimal action may consume too much time, risking failure. Instead, the player can choose sub-optimal actions, as long as they successfully lead to completing the mission.

In practice, the player will keep collecting samples by randomly choosing actions, until the estimated mean of one action is greater than $\mathbb{E}[\mu_n]$. Then the player can choose this action to complete the rest of the mission.

Similar to the analysis of the greedy strategy, the estimated mean $\mu_{(i,j)}$ in (33) will never be 100% accurate, so the distribution of $\mu_{(i,j)}-\mathbb{E}[\mu_n]$ is,

$$(\mu_{(i,j)}-\mathbb{E}[\mu_n])\sim\mathcal{N}(\widehat{\mu_{(i,j)}}-\mathbb{E}[\mu_n],\mathbb{V}[\widehat{\mu_{(i,j)}}]+\mathbb{V}[\mu_n]) \tag{38}$$

With a user-defined probability threshold P_τ, the player will choose the action $a_{(i,j)}$ if this action is the first one that satisfies,

$$\Phi\left(\frac{\widehat{\mu_{(i,j)}}-\mathbb{E}[\mu_n]}{\sqrt{\mathbb{V}[\widehat{\mu_{(i,j)}}]+\mathbb{V}[\mu_n]}}\right)\geq P_\tau \tag{39}$$

It indicates that the action $a_{(i,j)}$ is very likely to have a high enough value of estimated $\mu_{(i,j)}$ greater than $\mathbb{E}[\mu_n]$ with a probability P_τ.

As soon as the player finds the first action that satisfies the inequality (39), the player can stop the random sampling and use the same action for the rest of the mission.

C. The Mixed Strategy

The non-greedy strategy doesn't guarantee to identify a sub-optimal action because there is a chance that none of the actions has a mean $\mu_{(i,j)}\geq\mathbb{E}[\mu_n]$.

In this case, the player will automatically switch the strategy from non-greedy to greedy. The algorithm is shown below,

```
while taking random actions do
    if Non-greedy Strategy finds an action then
        take the sub-optimal action for the rest of the mission
    else if Greedy Strategy finds an action then
        take the optimal action for the rest of the mission
    end if
end while
```

VI. Cheating, Anti-Cheating and Unpredictability

A. Cheating

Cheating in computer games refers to players exploiting weaknesses, bugs, or using third-party tools to gain unfair advantages over others, often disrupting game balance and fairness.

Using the proposed probability core for game design, one can develop the tools to collect the real-time action utility data and implement these strategies to identify the best action, while human players won't be able to read and store the real-time data, and can't compute the CDF in real-time, so they will have disadvantages.

B. Anti-cheating

A simple anti-cheating technique for this probability core is to randomly refresh the (4) and (12), or refresh them in each mission of the game. This unpredictability ensures that long-term strategies can be interrupted by unforeseen events, keeping the game fresh and challenging even after repeated playthroughs.

The proposed probability core with the anti-cheating technique drives decision-making and uncertainty, but also introduces risk assessment into tactical decisions, ensuring that players must constantly weigh the odds of success before committing to an action. This type of probabilistic design improves player engagement by providing a balance between skill and luck [10].

C. Balancing Probability and Strategy

When games rely on random generation to create unique playthroughs, players must adapt their strategies based on uncertain and evolving game states.

A significant challenge in designing probability-based strategy games is achieving the right balance between randomness and player control [11], [12]. If the influence of probability is too great, it may lead to frustration, as outcomes seem arbitrary. Conversely, if probability plays too small a role, the game may become predictable and repetitive.

Game designers can adjust the parameters of this probability core to gain the balance between probability and strategy.

VII. Conclusions and Discussions

This paper presents a generalized probability-based framework for game design, particularly in strategy games. By leveraging probability theory, the framework allows for flexible level generation and strategy development, leading to varied gameplay experiences that challenge players with unpredictability. The model enables the creation of diverse game levels that avoid repetition and support dynamic player decision-making.

The use of normal and uniform distributions in estimating action utilities introduces uncertainty, which enhances engagement by requiring players to balance risk and strategy under time constraints. Furthermore, the introduction of both greedy and non-greedy strategies provides options for different playstyles, depending on the player's risk tolerance and mission requirements.

A key challenge highlighted in this work is balancing randomness and player control. Excessive reliance on probability can lead to frustration, while too little can result in predictable and less engaging gameplay. The framework offers a flexible solution, enabling developers to adjust the role of probability according to their design needs.

In terms of anti-cheating measures, the paper outlines techniques such as random refreshing of distributions to prevent players from exploiting real-time data. This approach ensures fairness and preserves the unpredictability essential to maintaining challenging gameplay.

Future research could explore refining the balance between randomness and player control, as well as expanding the applicability of this framework to more complex multiplayer environments where strategy and probability intersect at a deeper level [13], [14]. Additionally, integrating AI-driven elements into this model could further enhance the game's adaptability and provide more personalized gameplay experiences for players.

REFERENCES

[1] Drivet, Alessio. (2022). Probability and Game. 10.4018/978-1-6684-7589-8.ch046.

[2] E. Adams, Fundamentals of game design. Pearson Education, 2013

[3] J. Schell, The Art of Game Design: A book of lenses. CRC Press, 2008.

[4] Small, M., & Tse, C. K. (2012). Predicting the outcome of roulette. Chaos (Woodbury, N.Y.), 22(3), 033150. doi:10.1063/1.4753920

[5] M. Kwiatkowska, G. Norman, and D. Parker, PRISM 4.0: Verification of probabilistic real-time systems. Proc. 23rd International Conference on Computer Aided Verification (CAV'11), ser. LNCS, G. Gopalakrishnan and S. Qadeer, Eds., vol. 6806. Springer, 2011, pp. 585–591.

[6] K. Sen, M. Viswanathan, and G. Agha, Statistical model checking of black-box probabilistic systems. Computer Aided Verification. Springer, 2004, pp. 202–215.

[7] E. M. Clarke, Jr., O. Grumberg, and D. A. Peled, Model Checking. Cambridge, MA, USA: MIT Press, 1999.

[8] Tversky, A., & Kahneman, D. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263–291. doi:10.2307/1914185

[9] Oliveira Júnior, Ailton Paulo & Datori Barbosa, Nilceia. (2024). THE DIGITAL GAME "PROBABILITY IN ACTION": MOTIVATION, USER EXPERIENCE AND LEARNING. REAMEC - Rede Amazônica de Educação em Ciências e Matemática. 12. e24048. 10.26571/reamec.v12.17397.

[10] Campbell, P. J. (2007). Games of chance with multiple objectives. Metrika, 66(3), 305–313. doi:10.100700184-006-0112-5

[11] Xu, Anqi. (2024). Research on the Ways to Win in a Simple Poker Game. Highlights in Science, Engineering and Technology. 107. 150-156. 10.54097/hkat8b41.

[12] Kehagias, Ath & Gkyzis, Georgios & Karakoulakis, A. & Kyprianidis, A.. (2024). A Game Theoretic Analysis of the Three-Gambler Ruin Game. 10.48550/arXiv.2406.07878.

[13] Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. Management Science, 42(12), 1676–1690. doi:10.1287/mnsc.42.12.1676

[14] Harmer, G. P., & Abbott, D. (1999). Losing strategies can win by Parrondo's paradox. Nature, 402(6764),864–864. doi:10.1038/47220

APPENDIX A

$$\begin{aligned}&\int_{\mu_{\min}}^{\mu_{\max}} \mu \cdot \frac{n(\mu-\mu_{\min})^{n-1}}{(\mu_{\max}-\mu_{\min})^n}\,d\mu\\&= \frac{n}{(\mu_{\max}-\mu_{\min})^n}\int_{\mu_{\min}}^{\mu_{\max}} \mu(\mu-\mu_{\min})^{n-1}\,d\mu\end{aligned} \tag{40}$$

- Introduce a change of variable to simplify the expression. Define:

$$u = \mu - \mu_{\min}, \quad \text{so} \quad du = d\mu$$

This implies that when $\mu = \mu_{\min}$, $u = 0$, and when $\mu = \mu_{\max}$, $u = \mu_{\max} - \mu_{\min}$

Under this substitution, the integral becomes:

$$\begin{aligned}&\frac{n}{(\mu_{\max}-\mu_{\min})^n}\int_0^{\mu_{\max}-\mu_{\min}} (u+\mu_{\min})u^{n-1}\,du\\&= \frac{n}{(\mu_{\max}-\mu_{\min})^n}\Big[\int_0^{\mu_{\max}-\mu_{\min}} u^n\,du +\\&\qquad \mu_{\min}\int_0^{\mu_{\max}-\mu_{\min}} u^{n-1}\,du\Big]\end{aligned} \tag{41}$$

- The integrals $\int_0^{\mu_{\max}-\mu_{\min}} u^n\,du$ and $\int_0^{\mu_{\max}-\mu_{\min}} u^{n-1}\,du$ can be computed using the power rule for integration:

$$\int u^k\,du = \frac{u^{k+1}}{k+1} \tag{42}$$

Therefore:

$$\begin{aligned}\int_0^{\mu_{\max}-\mu_{\min}} u^n\,du &= \frac{(\mu_{\max}-\mu_{\min})^{n+1}}{n+1}\\ \int_0^{\mu_{\max}-\mu_{\min}} u^{n-1}\,du &= \frac{(\mu_{\max}-\mu_{\min})^{n}}{n}\end{aligned} \tag{43}$$

- Substituting these results back into the expression:

$$\begin{aligned}&\frac{n}{(\mu_{\max}-\mu_{\min})^n}\cdot\left[\frac{(\mu_{\max}-\mu_{\min})^{n+1}}{n+1} + \mu_{\min}\frac{(\mu_{\max}-\mu_{\min})^{n}}{n}\right]\\&= \frac{n}{(\mu_{\max}-\mu_{\min})^n}\cdot(\mu_{\max}-\mu_{\min})^n\left[\frac{(\mu_{\max}-\mu_{\min})}{n+1}+\frac{\mu_{\min}}{n}\right]\\&= n\left[\frac{(\mu_{\max}-\mu_{\min})}{n+1}+\frac{\mu_{\min}}{n}\right]\end{aligned} \tag{44}$$

APPENDIX B

$$\begin{aligned}&\int_{\mu_{\min}}^{\mu_{\max}} \mu^2 \cdot \frac{n(\mu-\mu_{\min})^{n-1}}{(\mu_{\max}-\mu_{\min})^n}\,d\mu\\&= \frac{n}{(\mu_{\max}-\mu_{\min})^n}\int_{\mu_{\min}}^{\mu_{\max}} \mu^2(\mu-\mu_{\min})^{n-1}\,d\mu\end{aligned} \tag{45}$$

- To simplify the integral, let's make the substitution $u = \mu - \mu_{\min}$. Then:

$du = d\mu$

When $\mu = \mu_{\text{min}}, u = 0$

When $\mu = \mu_{\text{max}}, u = \mu_{\text{max}} - \mu_{\text{min}}$

Also, $\mu = u + \mu_{\text{min}}$. So the integral becomes:

$$\frac{n}{(\mu_{\text{max}} - \mu_{\text{min}})^n} \int_0^{\mu_{\max} - \mu_{\min}} (u + \mu_{\text{min}})^2 u^{n-1} \, du$$

$$\begin{aligned}
&= \frac{n}{(\mu_{\text{max}} - \mu_{\text{min}})^n} \int_0^{\mu_{\max} - \mu_{\min}} \left(u^2 + 2u\mu_{\text{min}} + \mu_{\text{min}}^2\right) u^{n-1} du \\
&= \frac{n}{(\mu_{\max} - \mu_{\min})^n} \Big[\int_0^{\mu_{\max} - \mu_{\min}} u^{n+1} \, du + \\
&\qquad 2\mu_{\min} \int_0^{\mu_{\max} - \mu_{\min}} u^{n} \, du + \\
&\qquad \mu_{\min}^2 \int_0^{\mu_{\max} - \mu_{\min}} u^{n-1} \, du \Big]
\end{aligned} \tag{46}$$

- We can now integrate each term using (42):

$$\begin{aligned}
\int_0^{\mu_{\text{max}} - \mu_{\text{min}}} u^{n+1} du &= \frac{(\mu_{\text{max}} - \mu_{\text{min}})^{n+2}}{n+2} \\
\int_0^{\mu_{\text{max}} - \mu_{\text{min}}} u^{n} du &= \frac{(\mu_{\text{max}} - \mu_{\text{min}})^{n+1}}{n+1} \\
\int_0^{\mu_{\text{max}} - \mu_{\text{min}}} u^{n-1} du &= \frac{(\mu_{\text{max}} - \mu_{\text{min}})^{n}}{n}
\end{aligned} \tag{47}$$

- Substituting the results of the integrals back into the original expression:

$$\begin{aligned}
&\frac{n}{(\mu_{\text{max}} - \mu_{\text{min}})^n} \Big[\frac{(\mu_{\text{max}} - \mu_{\text{min}})^{n+2}}{n+2} + \\
&\qquad 2\mu_{\text{min}} \frac{(\mu_{\text{max}} - \mu_{\text{min}})^{n+1}}{n+1} + \\
&\qquad \mu_{\text{min}}^2 \frac{(\mu_{\text{max}} - \mu_{\text{min}})^{n}}{n} \Big] \\
&= \frac{n}{(\mu_{\text{max}} - \mu_{\text{min}})^n} \cdot (\mu_{\text{max}} - \mu_{\text{min}})^n \Big[\frac{(\mu_{\text{max}} - \mu_{\text{min}})^{2}}{n+2} + \\
&\qquad 2\mu_{\text{min}} \frac{\mu_{\text{max}} - \mu_{\text{min}}}{n+1} + \\
&\qquad \mu_{\text{min}}^2 \frac{1}{n} \Big] \\
&= n \left[\frac{(\mu_{\text{max}} - \mu_{\text{min}})^{2}}{n+2} + 2\mu_{\text{min}} \frac{\mu_{\text{max}} - \mu_{\text{min}}}{n+1} + \mu_{\text{min}}^2 \frac{1}{n} \right]
\end{aligned} \tag{48}$$