Central Limit Theorem in Complete Feedback Games

Speaker: Raghavendra Tripathi

University of Washington, Seattle

Department of Mathematics and Statistics, IIT Kanpur

11 September, 2024

Lady tasting tea experiment (1920)

Ronald Fisher

Milk first or
Tea first

Muriel Bristol

5. Statement of Experiment

A LADY declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. We will consider the problem of designing an experiment by means of which this assertion can be tested. For this purpose

The Design of Experiments: Chatpter II

Lady tasting tea experiment

Ronald Fisher

Milk first or
Tea first

Muriel Bristol

Lady tasting tea experiment

Ronald Fisher

Milk first or
Tea first

Muriel Bristol

$\sharp\{$Number of correct guesses$\} \sim$ Hypergeometric$(N = 8, K = 4, n = 4)$.

- $N = 8$ Population size
- $K = 4$ Number of success states
- $n = 4$ Number of draws

Helping Muriel

Without any supernatural ability and without any external help, Muriel is expected to make only 4 correct guesses.

Helping Muriel

Without any supernatural ability and without any external help, Muriel is expected to make only 4 correct guesses.

What if Muriel is provided help?

Helping Muriel

Without any supernatural ability and without any external help, Muriel is expected to make only 4 correct guesses.

What if Muriel is provided help? For instance, if Muriel is told whether her guess is correct or not after each guess, can she improve the number of correct guesses?

Helping Muriel

Without any supernatural ability and without any external help, Muriel is expected to make only 4 correct guesses.

What if Muriel is provided help? For instance, if Muriel is told whether her guess is correct or not after each guess, can she improve the number of correct guesses?

Optimal Strategy and Score

Knowing that x cups of Type T and y cups of type M are remaining, she should guess the type corresponding to $\max(T, M)$.

Helping Muriel

Without any supernatural ability and without any external help, Muriel is expected to make only 4 correct guesses.

What if Muriel is provided help? For instance, if Muriel is told whether her guess is correct or not after each guess, can she improve the number of correct guesses?

Optimal Strategy and Score

Knowing that x cups of Type T and y cups of type M are remaining, she should guess the type corresponding to $\max(T, M)$. With this strategy, she can make $373/70 = 5.3$ correct guesses (on average). **(Diaconis and Graham '81)**

Complete feedback games: Setup

- Consider a well-shuffled deck of cards of n-types: $1, 2, \ldots, n$.
- Number of cards of type i : m.
- Total number of cards: $n \times m$.

Complete feedback games: Setup

- Consider a well-shuffled deck of cards of n-types: $1, 2, \ldots, n$.
- Number of cards of type i : m.
- Total number of cards: $n \times m$.

Lady tasting tea experiment

$$n = 2, \quad m = 4\ .$$

Complete feedback games: Setup

- Consider a well-shuffled deck of cards of n-types: $1, 2, \ldots, n$.
- Number of cards of type i : m.
- Total number of cards: $n \times m$.

Lady tasting tea experiment

$$n = 2, \quad m = 4 \; .$$

Usual deck of cards

$$n = 4, \quad m = 13$$

or

$$n = 13, \quad m = 4$$

Complete feedback games: Setup

- Consider a well-shuffled deck of cards of n-types: $1, 2, \ldots, n$.
- Number of cards of type i : m.
- Total number of cards: $n \times m$.

Lady tasting tea experiment

$$n = 2, \quad m = 4 \ .$$

Usual deck of cards

$$n = 4, \quad m = 13$$

or

$$n = 13, \quad m = 4$$

Zener Cards

$$n = 5, \quad m = 5 \ .$$

Complete feedback games: Setup

- Consider a well-shuffled deck of cards of n-types: $1, 2, \ldots, n$.
- Number of cards of type i : m.
- Total number of cards: $n \times m$.

Rule of the game

- The player guesses the type of the topmost card in the deck.
- The player is shown the card and the card is removed from the deck.
- The game continues till the deck is exhausted.

Complete feedback games: Setup

- Consider a well-shuffled deck of cards of n-types: $1, 2, \ldots, n$.
- Number of cards of type i : m.
- Total number of cards: $n \times m$.

Rule of the game

- The player guesses the type of the topmost card in the deck.
- The player is shown the card and the card is removed from the deck.
- The game continues till the deck is exhausted.

Score $T_{m,n} =$ Total number of correct guesses at the end.

Complete feedback games: Setup

- Consider a well-shuffled deck of cards of n-types: $1, 2, \ldots, n$.
- Number of cards of type i : m.
- Total number of cards: $n \times m$.

Rule of the game

- The player guesses the type of the topmost card in the deck.
- The player is shown the card and the card is removed from the deck.
- The game continues till the deck is exhausted.

Score $T_{m,n} =$ Total number of correct guesses at the end.

Greedy guessing maximizes the mean number of correct guesses (Diaconis and Graham)

Complete feedback games: General Setup

- Consider a well-shuffled deck of cards of n-types: $1, 2, \ldots, n$.
- Number of cards of type i : m_i.
- Total number of cards: $\sum_{i=1}^{n} m_i$.

Rule of the game

- The player guesses the type of the topmost card in the deck.
- The player is shown the card and the card is removed from the deck.
- The game continues till the deck is exhausted.

Score $T_{m,n} =$ Total number of correct guesses at the end.

Greedy guessing maximizes the mean number of correct guesses (Diaconis and Graham' 81)

The simplest case: $m = 1$

- The probability of guessing the first card correctly: $\frac{1}{n}$.
- The probability of guessing the i-th card correctly: $\frac{1}{n-i+1}$.

The simplest case: $m = 1$

- The probability of guessing the first card correctly: $\frac{1}{n}$.
- The probability of guessing the i-th card correctly: $\frac{1}{n-i+1}$.

$$\mathbb{E}[T_{1,n}] = 1 + \frac{1}{2} + \ldots + \frac{1}{n} \sim \log(n)\ .$$

$$\mathrm{Var}[T_{1,n}] \sim \log(n)\ .$$

The simplest case: $m = 1$

- The probability of guessing the first card correctly: $\frac{1}{n}$.
- The probability of guessing the i-th card correctly: $\frac{1}{n-i+1}$.

$$\mathbb{E}[T_{1,n}] = 1 + \frac{1}{2} + \ldots + \frac{1}{n} \sim \log(n) \ .$$

$$\mathrm{Var}[T_{1,n}] \sim \log(n) \ .$$

Gaussian fluctuation

$$\frac{T_{1,n} - \log(n)}{\log(n)} \xrightarrow{n \to \infty} \mathcal{N}(0,1) \ .$$

Brief History-I

- $n = 1$ is trivial.

Brief History-I

- $n = 1$ is trivial.
- Blackwell and Hodge (1957): $n = 2$ and large $m_1 = m_2 = m$.
  - Two types of treatment assigned to $2m$ patients arriving one by one.
  - Hospital may rule out some subjects on the grounds of medical conditions.
  - Hospital may bias the experiment: Selection bias.

Brief History-I

- $n = 1$ is trivial.
- Blackwell and Hodge (1957): $n = 2$ and large $m_1 = m_2 = m$.
  - Two types of treatment assigned to $2m$ patients arriving one by one.
  - Hospital may rule out some subjects on the grounds of medical conditions.
  - Hospital may bias the experiment: Selection bias.
  - Optimal expected score:

$$\mathbb{E}[T_{2,m}] \sim m + \frac{1}{2}\sqrt{\pi m} - \frac{1}{2} - O(\frac{1}{n}) \ .$$

Brief History-I

- $n = 1$ is trivial.
- Blackwell and Hodge (1957): $n = 2$ and large $m_1 = m_2 = m$.
  - Two types of treatment assigned to $2m$ patients arriving one by one.
  - Hospital may rule out some subjects on the grounds of medical conditions.
  - Hospital may bias the experiment: Selection bias.
  - Optimal expected score:

$$\mathbb{E}[T_{2,m}] \sim m + \frac{1}{2}\sqrt{\pi m} - \frac{1}{2} - O(\frac{1}{n}) \ .$$

- Diaconis and Graham (1981):
  - Fixed n and large m:

$$\mathbb{E}[T_{n,m}] \sim m + \frac{\pi}{2} M_n \sqrt{m} + o_n(\sqrt{m}) \ ,$$

  where M_n is the expected value of n i.i.d. standard Gaussian.

Brief History-I

- $n = 1$ is trivial.
- Blackwell and Hodge (1957): $n = 2$ and large $m_1 = m_2 = m$.
  - Two types of treatment assigned to $2m$ patients arriving one by one.
  - Hospital may rule out some subjects on the grounds of medical conditions.
  - Hospital may bias the experiment: Selection bias.
  - Optimal expected score:

$$\mathbb{E}[T_{2,m}] \sim m + \frac{1}{2}\sqrt{\pi m} - \frac{1}{2} - O(\frac{1}{n}) \ .$$

- Diaconis and Graham (1981):
  - Fixed n and large m:

$$\mathbb{E}[T_{n,m}] \sim m + \frac{\pi}{2} M_n \sqrt{m} + o_n(\sqrt{m}) \ ,$$

    where M_n is the expected value of n i.i.d. standard Gaussian.
  - Similar aymptotics for fixed n and $\mathbf{m} = (m_1, \ldots, m_n)$.

Brief History-I

- Diaconis and Graham (1981):
  - When $n = 2$ and $m_1 = m_2 = m$, the total score $T_{2,m}$ satisfies Rayleigh limit law:

$$\mathbb{P}\left(\frac{T_{2,m} - m}{\sqrt{m/2}} \leq x\right) \to 1 - e^{\frac{-x^2}{2}} \qquad x \geq 0\ .$$

  - The above correct version is due to Michael Proschan '91 (Annals of Statistics)

Brief History-I

- Diaconis and Graham (1981):
  - When $n = 2$ and $m_1 = m_2 = m$, the total score $T_{2,m}$ satisfies Rayleigh limit law:

$$\mathbb{P}\left(\frac{T_{2,m} - m}{\sqrt{m/2}} \leq x\right) \to 1 - e^{\frac{-x^2}{2}} \qquad x \geq 0\ .$$

  - The above correct version is due to Michael Proschan '91 (Annals of Statistics)
  - The asymptotic for the expected number of correct guesses and the fluctuations are also known when $m_1 \neq m_2$ but $m_1/(m_1 + m_2) \to p$.

Brief History-I

- Diaconis and Graham (1981):
  - When $n = 2$ and $m_1 = m_2 = m$, the total score $T_{2,m}$ satisfies Rayleigh limit law:

$$\mathbb{P}\left(\frac{T_{2,m} - m}{\sqrt{m/2}} \leq x\right) \to 1 - e^{\frac{-x^2}{2}} \qquad x \geq 0\ .$$

  - The above correct version is due to Michael Proschan '91 (Annals of Statistics)
  - The asymptotic for the expected number of correct guesses and the fluctuations are also known when $m_1 \neq m_2$ but $m_1/(m_1 + m_2) \to p$.
  - This case behaves very differently. In particular,

$$\mathbb{P}\left(T_{2,(m_1,m_2)} - \max\{m_1, m_2\} = k\right) \to \gamma(1-\gamma)^k\ , \quad \text{where } \gamma = \frac{2|p-q|}{1+|p-q|}.$$

Brief History-I

- Diaconis and Graham (1981):
  - When $n = 2$ and $m_1 = m_2 = m$, the total score $T_{2,m}$ satisfies Rayleigh limit law:

$$\mathbb{P}\left(\frac{T_{2,m} - m}{\sqrt{m/2}} \leq x\right) \to 1 - e^{\frac{-x^2}{2}} \qquad x \geq 0\,.$$

  - The above correct version is due to Michael Proschan '91 (Annals of Statistics)
  - The asymptotic for the expected number of correct guesses and the fluctuations are also known when $m_1 \neq m_2$ but $m_1/(m_1 + m_2) \to p$.
  - This case behaves very differently. In particular,

$$\mathbb{P}\left(T_{2,(m_1,m_2)} - \max\{m_1, m_2\} = k\right) \to \gamma(1-\gamma)^k\,, \quad \text{where} \quad \gamma = \frac{2|p-q|}{1+|p-q|}.$$

- Kuba, Panholzer and Prodinger (2009), Kuba and Panholzer (2012), Kuba and Panholzer (2024)
  - More precise understanding of $T_{2,(m_1,m_2)}$.

Brief History-I

- Diaconis and Graham (1981):
  - When $n = 2$ and $m_1 = m_2 = m$, the total score $T_{2,m}$ satisfies Rayleigh limit law:

$$\mathbb{P}\left(\frac{T_{2,m} - m}{\sqrt{m/2}} \le x\right) \to 1 - e^{\frac{-x^2}{2}} \qquad x \ge 0\,.$$

  - The above correct version is due to Michael Proschan '91 (Annals of Statistics)
  - The asymptotic for the expected number of correct guesses and the fluctuations are also known when $m_1 \neq m_2$ but $m_1/(m_1 + m_2) \to p$.
  - This case behaves very differently. In particular,

$$\mathbb{P}\left(T_{2,(m_1,m_2)} - \max\{m_1, m_2\} = k\right) \to \gamma(1-\gamma)^k\,, \quad \text{where } \gamma = \frac{2|p-q|}{1+|p-q|}.$$

- Kuba, Panholzer and Prodinger (2009), Kuba and Panholzer (2012), Kuba and Panholzer (2024)
  - More precise understanding of $T_{2,(m_1,m_2)}$.
  - In particular, Rayleigh limit law holds as long as $m_1 - m_2 = o(m_1)$.

Brief History-I

- Diaconis and Graham (1981):
  - When $n = 2$ and $m_1 = m_2 = m$, the total score $T_{2,m}$ satisfies Rayleigh limit law:

$$\mathbb{P}\left(\frac{T_{2,m} - m}{\sqrt{m/2}} \leq x\right) \to 1 - e^{\frac{-x^2}{2}} \qquad x \geq 0\ .$$

  - The above correct version is due to Michael Proschan '91 (Annals of Statistics)
  - The asymptotic for the expected number of correct guesses and the fluctuations are also known when $m_1 \neq m_2$ but $m_1/(m_1 + m_2) \to p$.
  - This case behaves very differently. In particular,

$$\mathbb{P}\left(T_{2,(m_1,m_2)} - \max\{m_1, m_2\} = k\right) \to \gamma(1-\gamma)^k\ , \quad \text{where} \ \ \gamma = \frac{2|p-q|}{1+|p-q|}.$$

- Kuba, Panholzer and Prodinger (2009), Kuba and Panholzer (2012), Kuba and Panholzer (2024)
  - More precise understanding of $T_{2,(m_1,m_2)}$.
  - In particular, Rayleigh limit law holds as long as $m_1 - m_2 = o(m_1)$.
  - Phase transition for the limit law of $T_{2,(m_1,m_2)}$.

A Brief History-II

- Diaconis, Graham, He, Spiro (2022): For fixed m as $n \to \infty$

$$\mathbb{E}[T_{m,n}] \sim (1 + o(1))H_m \log(n) \ ,$$

where $H_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}$.

A Brief History-II

- Diaconis, Graham, He, Spiro (2022): For fixed m as $n \to \infty$

$$\mathbb{E}[T_{m,n}] \sim (1 + o(1)) H_m \log(n) \ ,$$

where $H_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}$.

- He and Ottolini (2022): As $n \to \infty$ and $\mathbf{m} = (m_1, \ldots, m_n)$

$$\mathbb{E}[T_{n,\mathbf{m}}] \sim H_{m^*} H_n + \sum_{j=1}^{m^*} \log(\gamma_j) + O\left(\log(n) \left(\frac{\log(n)}{n} \right)^{1/m^*} \right) \ ,$$

where $m^* = \max_{i=1}^{n} m_i$ and $\gamma_j = \left(\frac{1}{n} \sum_{i=1}^{n} \binom{m_i}{j} \right)^{1/j}$.

A Brief History-II

- Diaconis, Graham, He, Spiro (2022): For fixed m as $n \to \infty$

$$\mathbb{E}[T_{m,n}] \sim (1 + o(1)) H_m \log(n) \ ,$$

where $H_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}$.

- He and Ottolini (2022): As $n \to \infty$ and $m_i \equiv m$ is fixed

$$\mathbb{E}[T_{n,m}] \sim H_m H_n + \sum_{j=1}^{m} \frac{1}{j} \log \binom{m}{j} + O\left(\log n \left(\frac{\log n}{n} \right)^{1/m} \right) \ .$$

Brief History-II

- Ottolini and Steinerberger (2023):
  - When $(\log n)^{3+\epsilon} \leq c \cdot m$, then

$$\mathbb{E}(T_{m,n}) \sim m + \frac{\pi}{\sqrt{2}}\sqrt{m \log n} + o_{c,\epsilon}(\sqrt{m \log(n)}) \ .$$

Brief History-II

- Ottolini and Steinerberger (2023):
  - When $(\log n)^{3+\epsilon} \leq c \cdot m$, then

$$\mathbb{E}(T_{m,n}) \sim m + \frac{\pi}{\sqrt{2}}\sqrt{m \log n} + o_{c,\epsilon}(\sqrt{m \log(n)}) \ .$$

- Compare this to He and Ottolini (2022): When m is fixed and $n \to \infty$

$$\mathbb{E}[T_{n,m}] \sim H_m H_n + \sum_{j=1}^{m} \frac{1}{j} \log \binom{m}{j} + O\left(n^{-1/m}\right) \ .$$

Brief History-II

- Ottolini and Steinerberger (2023):
  - When $(\log n)^{3+\epsilon} \leq c \cdot m$, then

$$\mathbb{E}(T_{m,n}) \sim m + \frac{\pi}{\sqrt{2}}\sqrt{m \log n} + o_{c,\epsilon}(\sqrt{m \log(n)}) \ .$$

- Compare this to He and Ottolini (2022): When m is fixed and $n \to \infty$

$$\mathbb{E}[T_{n,m}] \sim H_m H_n + \sum_{j=1}^{m} \frac{1}{j} \log \binom{m}{j} + O\left(n^{-1/m}\right) \ .$$

- This suggests a phase transition. Likely at around $\log n \sim m$.

Brief History-II

- Proof Heuristic for Ottolini and Steinerberger

Brief History-II

- Proof Heuristic for Ottolini and Steinerberger
  - Let $X_i(t) \in \{0, 1, \ldots, m\}$ be the number of cards of type i when the remaining deck has size t.

Brief History-II

- Proof Heuristic for Ottolini and Steinerberger
  - Let $X_i(t) \in \{0, 1, \ldots, m\}$ be the number of cards of type i when the remaining deck has size t.

$$\mathbb{E}(T_{m,n}) = \sum_{t=1}^{mn} \frac{\mathbb{E}(\max_i X_i(t))}{t} \ .$$

Brief History-II

- Proof Heuristic for Ottolini and Steinerberger
  - Let $X_i(t) \in \{0, 1, \dots, m\}$ be the number of cards of type i when the remaining deck has size t.

$$\mathbb{E}(T_{m,n}) = \sum_{t=1}^{mn} \frac{\mathbb{E}(\max_i X_i(t))}{t} \ .$$

  - $\mathbb{E}(X_i(t)) \approx mp$ and $\mathrm{Var}(X_i(t)) \approx mp(1-p)$ where $p = t/nm$.

Brief History-II

- Proof Heuristic for Ottolini and Steinerberger
  - Let $X_i(t) \in \{0, 1, \ldots, m\}$ be the number of cards of type i when the remaining deck has size t.

$$\mathbb{E}(T_{m,n}) = \sum_{t=1}^{mn} \frac{\mathbb{E}(\max_i X_i(t))}{t} \ .$$

  - $\mathbb{E}(X_i(t)) \approx mp$ and $\mathrm{Var}(X_i(t)) \approx mp(1-p)$ where $p = t/nm$.
  - Pretend that X_is are independent and normally distributed with correct mean and variance. Then,

$$\begin{aligned}\mathbb{E}(T_{m,n}) &= \sum_{t=1}^{mn} \frac{\mathbb{E}(\max_i X_i(t))}{t} \\ &\approx m + \sqrt{2m \log n} \int_0^1 \sqrt{\frac{1-p}{p}} \, dp \ .\end{aligned}$$

- In the general regime the variance, and the fluctuations are not understood.

Our Result

Setup

- Deck of card: $\mathbf{m}^n = (m_1, \ldots, m_n)$.

Our Result

Setup

- Deck of card: $\mathbf{m}^n = (m_1, \ldots, m_n)$. Here n denotes the number of distinct types of cards and m_i denotes the multiplicity of the card of type i.

Our Result

Setup

- Deck of card: $\mathbf{m}^n = (m_1, \ldots, m_n)$. Here n denotes the number of distinct types of cards and m_i denotes the multiplicity of the card of type i.
- $\mathbf{m}^n_{\max} := \max_{i=1}^n m_i$ denotes the highest multiplicity in the card.

Our Result

Setup

- Deck of card: $\mathbf{m}^n = (m_1, \ldots, m_n)$. Here n denotes the number of distinct types of cards and m_i denotes the multiplicity of the card of type i.
- $\mathbf{m}^n_{\max} := \max_{i=1}^n m_i$ denotes the highest multiplicity in the card.
- ϵ_n is the fraction of types i that appear with mutiplicity $\mathbf{m}^n_{\max}$.

Our Result

Setup

- Deck of card: $\mathbf{m}^n = (m_1, \ldots, m_n)$. Here n denotes the number of distinct types of cards and m_i denotes the multiplicity of the card of type i.
- $\mathbf{m}^n_{\max} := \max_{i=1}^n m_i$ denotes the highest multiplicity in the card.
- ϵ_n is the fraction of types i that appear with mutiplicity $\mathbf{m}^n_{\max}$.

Theorem (Ottolini and T.'2023)

Assume that $m_i \leq m$ for some m and $\epsilon_n \geq \epsilon$ for some $\epsilon > 0$ independent of n. Then,

$$\mathbb{E}[T_{\mathbf{m}^n,n}] \sim \mathrm{Var}[T_{\mathbf{m}^n,n}] \sim \left(1 + \frac{1}{2} + \ldots + \frac{1}{\mathbf{m}^n_{\max}}\right) \log n \quad \textit{as} \quad n \to \infty\ .$$

Our Result

Setup

- Deck of card: $\mathbf{m}^n = (m_1, \ldots, m_n)$. Here n denotes the number of distinct types of cards and m_i denotes the multiplicity of the card of type i.
- $\mathbf{m}^n_{\max} := \max_{i=1}^n m_i$ denotes the highest multiplicity in the card.
- ϵ_n is the fraction of types i that appear with mutiplicity $\mathbf{m}^n_{\max}$.

Theorem (Ottolini and T.'2023)

Assume that $m_i \leq m$ for some m and $\epsilon_n \geq \epsilon$ for some $\epsilon > 0$ independent of n. Then,

$$\mathbb{E}[T_{\mathbf{m}^n,n}] \sim \mathrm{Var}[T_{\mathbf{m}^n,n}] \sim \left(1 + \frac{1}{2} + \ldots + \frac{1}{\mathbf{m}^n_{\max}}\right) \log n \quad \textit{as} \quad n \to \infty \, .$$

And, there exists $C(\epsilon, m) > 0$ such that

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\frac{T_{\mathbf{m}^n,n} - \mathbb{E}[T_{\mathbf{m}^n,n}]}{\sqrt{\mathrm{Var}[T_{\mathbf{m}^n,n}]}} \leq x \right) - \Phi(x) \right| \leq C(\epsilon, m) \frac{\log \log n}{\sqrt{\log n}}$$

Proof Idea:

Define the following random variables:

$T_j = \max\{t \in \{0, \ldots, |\mathbf{m}|\}$: Among last t cards, no card appears more than j times}

Proof Idea:

Define the following random variables:

$$T_j = \max\{t \in \{0, \dots, |\mathbf{m}|\} : \text{Among last } t \text{ cards, no card appears more than } j \text{ times}\}$$

Example deck with 3 types and $\mathbf{m} = (3, 3, 2)$

Listed from the last to the first:

$$(1, 2, 2, 1, 3, 1, 2, 3).$$

In this case, $T_1 = 2, T_2 = 5$

Proof Idea:

Define the following random variables:

$$T_j = \max\{t \in \{0, \dots, |\mathbf{m}|\} : \text{Among last } t \text{ cards, no card appears more than } j \text{ times}\}$$

$$\widetilde{W}_{j,t} = \text{No. of cards that appear exactly } j \text{ times before some card appears } j+1 \text{ times}$$

Example deck with 3 types and $\mathbf{m} = (3, 3, 2)$

Listed from the last to the first:

$$(1, 2, 2, 1, 3, 1, 2, 3).$$

In this case, $T_1 = 2, T_2 = 5$

Proof Idea:

Define the following random variables:

$$T_j = \max\{t \in \{0, \ldots, |\mathbf{m}|\} : \text{Among last } t \text{ cards, no card appears more than } j \text{ times}\}$$

$$\widetilde{W}_{j,t} = \text{No. of cards that appear exactly } j \text{ times before some card appears } j+1 \text{ times}$$

Example deck with 3 types and $\mathbf{m} = (3, 3, 2)$

Listed from the last to the first:

$$(1, 2, 2, 1, 3, 1, 2, 3).$$

In this case, $T_1 = 2, T_2 = 5$
$\widetilde{W}_1 = 2, \widetilde{W}_2 = 2, \widetilde{W}_3 = 2.$

Proof Idea contd...

Lemma (Key Lemma)

For any deck $\mathbf{m}$, the optimal score $S_{\mathbf{m}}$ can be written as

$$S_{\mathbf{m}} = \sum_{j=1}^{m} \sum_{s=1}^{\widetilde{W}_j} X_{j,s},$$

where the $X_{j,s}$ are conditionally independent – given the $\widetilde{W}_j$'s – Bernoulli random variables with $\mathbb{P}(X_{j,s} = 1) = \frac{1}{s}$.

Proof Idea contd...

Lemma (Key Lemma)

For any deck $\mathbf{m}$*, the optimal score* $S_{\mathbf{m}}$ *can be written as*

$$S_{\mathbf{m}} = \sum_{j=1}^{m} \sum_{s=1}^{\widetilde{W}_j} X_{j,s},$$

where the $X_{j,s}$ *are conditionally independent – given the* $\widetilde{W}_j$*'s – Bernoulli random variables with* $\mathbb{P}(X_{j,s} = 1) = \frac{1}{s}$.

Strategy

1. CLT for the conditional score.

Proof Idea contd...

Lemma (Key Lemma)

For any deck $\mathbf{m}$, the optimal score $S_{\mathbf{m}}$ can be written as

$$S_{\mathbf{m}} = \sum_{j=1}^{m} \sum_{s=1}^{\widetilde{W}_j} X_{j,s},$$

where the $X_{j,s}$ are conditionally independent – given the $\widetilde{W}_j$'s – Bernoulli random variables with $\mathbb{P}(X_{j,s} = 1) = \frac{1}{s}$.

Strategy

1. CLT for the conditional score.
   - Conditioned on $\widetilde{W}_j, j = 1, \ldots, m$, we prove CLT for the conditional score with suitable conditional mean $\mu'_{\mathbf{m}}$ and variance $\sigma'^2_{\mathbf{m}}$.

Proof Idea contd...

Lemma (Key Lemma)

For any deck $\mathbf{m}$*, the optimal score* $S_{\mathbf{m}}$ *can be written as*

$$S_{\mathbf{m}} = \sum_{j=1}^{m} \sum_{s=1}^{\widetilde{W}_j} X_{j,s},$$

where the $X_{j,s}$ *are conditionally independent – given the* $\widetilde{W}_j$*'s – Bernoulli random variables with* $\mathbb{P}(X_{j,s} = 1) = \frac{1}{s}$.

Strategy

1. CLT for the conditional score.
   - Conditioned on $\widetilde{W}_j, j = 1, \ldots, m$, we prove CLT for the conditional score with suitable conditional mean $\mu'_{\mathbf{m}}$ and variance $\sigma'^2_{\mathbf{m}}$.
2. Removing the conditioning.
   - This requires understanding the behavior of T_js and the dependence of $\widetilde{W}_j$s on T_j.

Some other variants

- Partial Feedback Games

Some other variants

- Partial Feedback Games
  - Diaconis and Graham (1981): The greedy strategy is not optimal.
  - Diaconis, Graham, He, Spiro (2022), Nie (2022): Expected score $m + \Theta(\sqrt{m})$ uniformly in n for $n >> m >> 1$.

Some other variants

- Partial Feedback Games
  - Diaconis and Graham (1981): The greedy strategy is not optimal.
  - Diaconis, Graham, He, Spiro (2022), Nie (2022): Expected score $m + \Theta(\sqrt{m})$ uniformly in n for $n >> m >> 1$.
  - Non-normal fluctuation when $m = 1$. (Diaconis and Graham' 81)
  - The variance and the fluctuations are currently unknown for $m > 1$.

Some other variants

- Partial Feedback Games
  - Diaconis and Graham (1981): The greedy strategy is not optimal.
  - Diaconis, Graham, He, Spiro (2022), Nie (2022): Expected score $m + \Theta(\sqrt{m})$ uniformly in n for $n >> m >> 1$.
  - Non-normal fluctuation when $m = 1$. (Diaconis and Graham' 81)
  - The variance and the fluctuations are currently unknown for $m > 1$.
- Near-Optimal Strategies in Partial Feedback games
  - Chung, Diaconis, Graham, Mallows (2021): Optimal Strategy
  - Diaconis, Graham and Holmes (2022)

Some other variants

- Partial Feedback Games
  - Diaconis and Graham (1981): The greedy strategy is not optimal.
  - Diaconis, Graham, He, Spiro (2022), Nie (2022): Expected score $m + \Theta(\sqrt{m})$ uniformly in n for $n >> m >> 1$.
  - Non-normal fluctuation when $m = 1$. (Diaconis and Graham' 81)
  - The variance and the fluctuations are currently unknown for $m > 1$.
- Near-Optimal Strategies in Partial Feedback games
  - Chung, Diaconis, Graham, Mallows (2021): Optimal Strategy
  - Diaconis, Graham and Holmes (2022)
- Problem of minimizing the total score
  - Diaconis and Graham (1981): Strategy
  - He and Ottolini (2022): Asymptotics for the mean
  - Fluctuations are not understood.

Some other variants

- Partial Feedback Games
  - Diaconis and Graham (1981): The greedy strategy is not optimal.
  - Diaconis, Graham, He, Spiro (2022), Nie (2022): Expected score $m + \Theta(\sqrt{m})$ uniformly in n for $n >> m >> 1$.
  - Non-normal fluctuation when $m = 1$. (Diaconis and Graham' 81)
  - The variance and the fluctuations are currently unknown for $m > 1$.
- Near-Optimal Strategies in Partial Feedback games
  - Chung, Diaconis, Graham, Mallows (2021): Optimal Strategy
  - Diaconis, Graham and Holmes (2022)
- Problem of minimizing the total score
  - Diaconis and Graham (1981): Strategy
  - He and Ottolini (2022): Asymptotics for the mean
  - Fluctuations are not understood.
- Decks that are not properly shuffled
  - Ciucu (1998): Dovetail shuffle
  - Liu (2021): Riffle-Shuffle with complete feedback
  - Kuba and Panholzer (2023): Limit law with one riffle-shuffle and no feedback

Thank you!