Optimal match length in knock-out tournaments

Yuqiong Wang
yuqw@umich.edu

Joint work with Henri Doerks and Erik Ekström

Department of Mathematics
University of Michigan

Optimal Stopping and Related Fields Seminar, Columbia University

February 6, 2026

Outline

The goal of a tournament is to single out the best player.

Some sports play deterministic rules. e.g.,

- **Soccer:** 90 minutes regulation + 30 minutes extra time if needed. Historically some sequential rule (golden goal).
- **Basketball:** Each game is always 48 minutes.
- **American football:** Each game is always 60 minutes.

And some sports play sequential rules. e.g.,

- **Table tennis:** Game ends when someone wins 11 points with margin. Win four out of 7 sets (similar: tennis, badminton, volleyball).
- **Boxing:** stop when essentially obvious evidence appears.
- **Quidditch:** game ends immediately when the Snitch is caught.

Motivation: knock-out tournament design

In these games sometimes we think later stage games are more important:

- **Table tennis/badminton.** Early rounds are often best of 3 sets, while semifinals/finals are 5 or 7 sets.
- **Boxing.** Title fights are scheduled for 12 rounds, while non titled 4-10 rounds

Does this make intuitive sense:? The final determines the champion, while earlier rounds are numerous and each is less "informative" for the final winner.

Statistical viewpoint: longer matches reduce randomness and make the stronger player more likely to advance.

In this talk we study how to *optimally allocate match lengths* in a 2^n-player knock-out tournament. We fix a target tournament accuracy

$$\mathbb{P}(\text{best player wins the tournament}) \geq 1 - \eta.$$

Two design questions:

1. **Sequential vs. fixed length:** How much can we reduce the *average time per game* by using sequential stopping rules instead of fixed-length matches?
2. **Round allocation:** Should later stage games be longer? If yes, what is the *optimal round-dependent schedule*?

Outline

Relative efficiency of sequential vs non-sequential tests

To answer the first question, we focus on the classical testing problem for a Brownian motion with unknown drift.

For each match we compare two testing methods:

- **Sequential testing:** where the match ends when evidence is strong enough.
- **Fixed-length testing:** a pre-determined sample size (fixed match length), chosen to achieve the same error level.

Relative efficiency of sequential vs non-sequential tests

- The behavior of the players are often associated with some uncertainty.
- When a match is played, we can observe its "behavior" as a continuous process.
- We model the match behavior of Player 1 and 2 as a Brownian motion with drift:

$$X_t = \theta t + W_t,$$

where the unknown drift is unobservable:

$$\theta = \begin{cases} \frac{1}{2}, & \text{if Player 1 is better than Player 2,} \\ -\frac{1}{2}, & \text{if Player 2 is better than Player 1,} \end{cases}$$

- We thus test against the hypotheses H_+ and H_-, where

$$H_+ : \theta = \frac{1}{2}, \quad H_- : \theta = -\frac{1}{2}.$$

Non-sequential tests (fixed sample size)

- In a **fixed-size** sample test, one observes the process X over a time-interval $[0, T]$, where T is a constant chosen **before the sample is collected.**
- In a **sequential** test, choose a stopping time τ once sufficient evidence is obtained.
- Naturally, the average sample size in a sequential test ($\mathbb{E}[\tau]$) is smaller than the fixed sample size T.
- We study the average relative reduction $\frac{T-\mathbb{E}[\tau]}{T}$ as a function of the power of the test.
- Denote $1-\eta \in (\frac{1}{2}, 1)$ the power of a test, so that $\eta \in (0, \frac{1}{2})$ is the maximal probability with which the wrong hypothesis is accepted.
- We require that

$$\mathbb{P}_1(d=1) \geq 1-\eta \qquad \& \qquad \mathbb{P}_{-1}(d=-1) \geq 1-\eta.$$

Non-sequential tests (fixed sample size)

- By symmetry, an optimal decision at T is given by

$$d = \begin{cases} \frac{1}{2} & \text{if } X_T \geq 0 \\ -\frac{1}{2} & \text{if } X_T < 0. \end{cases}$$

- Therefore, we choose $T \geq T_\eta$, where T_η is defined by

$$\mathbb{P}_1(X_{T_\eta} \leq 0) = \mathbb{P}_{-1}(X_{T_\eta} \geq 0) = \Phi\left(-\sqrt{T_\eta}/2\right) = \eta,$$

where Φ and φ are the CDF and pdf of $N(0,1)$.

- Consequently,

$$T_\eta = 4\left(\Phi^{-1}(\eta)\right)^2.$$

Sequential tests

- In sequential experiments, the process is observed until an information level is achieved.
- Define the posterior probability process

$$\Pi_t := \mathbb{P}(\theta = \frac{1}{2} | \mathcal{F}_t)$$

- The process is observed until $\tau := \inf\{t \geq 0 : \Pi_t \notin (b, 1-b)\}$ for some $b \in (0, \frac{1}{2})$.
- Declare $\theta = \frac{1}{2}$ if $\Pi_\tau = 1 - b$ and $\theta = -\frac{1}{2}$ otherwise.
- We can show that

$$\mathbb{P}(\theta = d) = 1 - b.$$

- Therefore, to have the same power of the test, we let $b = 1 - \eta$.

Sequential tests

The expected value of observation time $\mathbb{E}[\tau]$ is characterized through finding the unique solution of the boundary value problem

$$\begin{cases} \frac{1}{2}\pi^2(1-\pi)^2 u_{\pi\pi} + 1 = 0, & \pi \in (\eta, 1-\eta), \\ u(\eta) = u(1-\eta) = 0, \end{cases}$$

with $\mathbb{E}[\tau] = u(\frac{1}{2}) = -2\psi(\eta)$, where $\psi(x) := (1-2x)\ln\frac{x}{1-x}$.

Remark 1

By Wald, there is one test that simultaneously minimizes $\mathbb{E}_{-1}[\tau]$ and $\mathbb{E}_1[\tau]$ over all tests with power $1-\eta$. The stopping time in that test is given by

$$\tau_\eta := \inf\left\{t \geq 0 : X_t \notin \left(\log\frac{\eta}{1-\eta}, \log\frac{1-\eta}{\eta}\right)\right\}.$$

The average sample size then satisfies

$$\mathbb{E}_{-1}\left[\tau_\eta\right] = \mathbb{E}_1\left[\tau_\eta\right] = -2\psi(\eta).$$

- We introduce the function $f : (0, \frac{1}{2}) \to [0,1]$ given by

$$f(\eta) := \frac{\mathbb{E}_1[\tau_\eta]}{T_\eta} = \frac{-\psi(\eta)}{2\left(\Phi^{-1}(\eta)\right)^2}. \tag{1}$$

- $f(\eta) < 1$, and $1 - f(\eta)$ represents the average sample size reduction by using sequential tests.
- We provide precise bounds on $f(\eta)$.

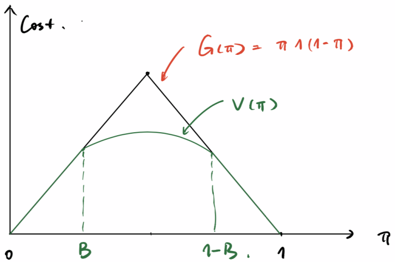
Remark 1. Equivalency to Shiryaev's testing problem

Recall the Bayesian formulation with constant cost per unit time:

$$V = \inf_{\tau, d} \{\mathbb{P}(d = 0, \Theta = 1) + \mathbb{P}(d = 1, \Theta = 0) + c\mathbb{E}[\tau]\}. \tag{2}$$

Problem (2) can be written as

$$V(\pi) = \inf_{\tau} \mathbb{E}_{\pi}[c\tau + \Pi_\tau \wedge (1 - \Pi_\tau)]$$



- The formulation with η is equivalent to the classical testing problem with c via a Lagrange formulation.
- In other words, the optimized Lagrange multiplier corresponds to $\frac{1}{c}$.

Remark 2. dependence on signal-to-noise ratio

- If the observation process is given by $X_t = \theta\mu t + W_t$, then the signal-to-noise ratio is μ.
- By Brownian scaling, the process

$$\tilde{X}_t := \mu X_{t/\mu^2} = \mu\left(\theta\frac{t}{\mu^2} + W_{t/\mu^2}\right) = \theta t + \tilde{W}_t,$$

  where $\tilde{W}_t := \mu W_{t/\mu^2}$ is a standard Brownian motion.
- Then $T_{\mu,\eta}$ needed to achieve a certain precision $1-\eta$ when the signal-to-noise ratio is μ satisfies

$$T_{\mu,\eta} = \frac{1}{\mu^2}T_{1,\eta} = \frac{1}{\mu^2}T_\eta.$$

- Similarly, $\mathbb{E}[\tau_{\mu,\eta}] = \frac{1}{\mu^2}E[\tau_{1,\eta}] = \frac{1}{\mu^2}E[\tau_\eta]$.
- Both scale inversely with μ^2, but the average reduction ratio $1-f(\eta)$ is independent of μ. It suffices to consider $\mu = \frac{1}{2}$.

The average reduction ratio is increasing in the error η.

Theorem 2

The function f is increasing on $(0, \frac{1}{2})$, with $f(0+) = \frac{1}{4}$ and $f(\frac{1}{2}-) = \frac{2}{\pi}$.

Remark. The minimal reduction is 36% and the maximal reduction is 75%. $f(0+) = \frac{1}{4}$ is well known [cf. Aivazjan('59), Shiyaev('78)], and numerically [cf. Eisenberg('91)] that convergence to 75% is rather slow, and that typical reductions range between 50% and 60%.

The monotonicity is perhaps obvious but new.

Table: Relative efficiency of SPRT.

Power $1-\eta$	$f(\eta)$	Average reduction
0.80	0.5871	41.29%
0.90	0.5351	46.49%
0.95	0.4897	51.03%
0.99	0.4160	58.40%
0.999	0.3609	63.91%

Proposition 3

We have $f(\eta) = \frac{1}{4} - \frac{\ln(-\ln \eta)}{8 \ln \eta} + o(\frac{\ln(-\ln \eta)}{\ln \eta})$ *as* $\eta \to 0$.

Remark. f approaches its limit at $\eta = \frac{1}{2}$ quadratically.

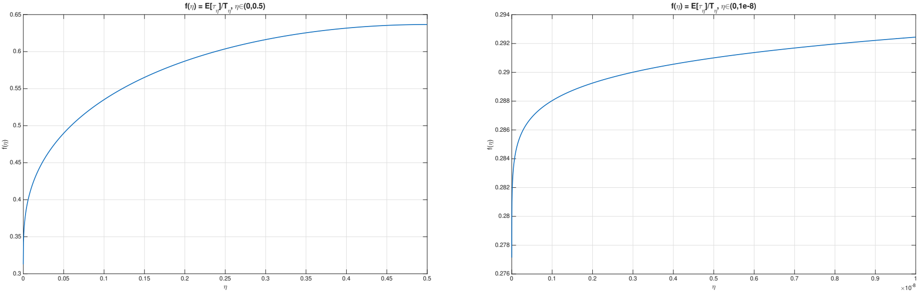


Figure: The function $f(\eta)$ for $\eta \in (0, \frac{1}{2})$ (left), and $f(\eta)$ for $\eta \in (0, 10^{-8})$ (right).

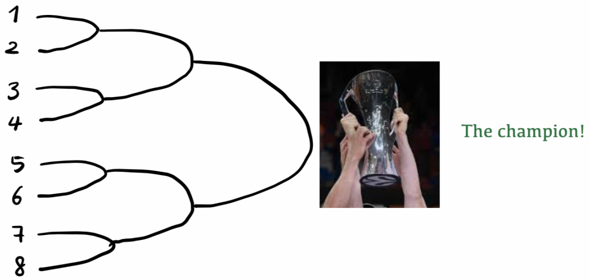
To reach a $70\%, 72\%, 74\%$ efficiency reduction, we need error probabilities η to be $2 \times 10^{-7}, 2 \times 10^{-12}, 2 \times 10^{-40}$, respectively.

Outline

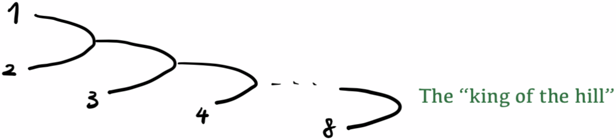
Problem formulation

There are many different forms of tournaments. Of the knockout types:

- Football, table tennis,. . .



- Children game, sorting:



Problem formulation

Consider a game of 2^n players (of football type). Want to minimize the observation cost (time length), subject to some probability error.

Assumptions:

- There is a distinct rank.
- Each match is a Brownian motion

$$X_t^{ij} = \Theta^{ij} t + W_t^{ij},$$

where

$$\Theta^{ij} = \begin{cases} \frac{1}{2}, \text{if } i \text{ is better than } j, \\ -\frac{1}{2}, \text{if } j \text{ is better than } i, \end{cases}$$

- Initially, we have a uniform prior distribution on $n!$ configurations. e.g., with $n = 2$:

$$\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 4 \\ 3 \end{pmatrix}, \dots, \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} \right\}, \quad \text{(24 configurations)}$$

- Each match in stage k is played until a pre-determined time T_k is reached.
- The winner is Player i if $X^{ij}_{T_k} \geq 0$, and Player j otherwise.
- We assume every game is **equally costly**.
- Fix $\eta \in (0, 2^{-n})$, we consider

$$\inf_{\{T_k\}_{k=1}^n} \sum_{k=1}^{n} 2^{n-k} T_k$$

subject to the constraint

$$\mathbb{P}(\text{the best player wins the tournament}) \geq 1 - \eta.$$

- Since a better player wins a match of length T with probability $\Phi(\frac{\sqrt{T}}{2})$, this is equivalent to

$$\inf_{\{T_k\}_{k=1}^n} \sum_{k=1}^{n} 2^{n-k} T_k \quad \text{subject to} \quad \prod_{k=1}^{n} \Phi(\frac{\sqrt{T_k}}{2}) \geq 1 - \eta. \tag{3}$$

Matches of fixed lengths

- Define the Lagrangian

$$\mathcal{L}(\{T_k\}_{k=1}^n, \lambda) := \sum_{k=1}^n 2^{n-k} T_k - \lambda \left(\sum_{k=1}^n a(T_k) - \log(1-\eta) \right)$$

with $a(T) := \log(\Phi(\frac{\sqrt{T}}{2}))$.

- a is increasing and strictly concave.
- For the kth round, take the first order condition: $\frac{\partial \mathcal{L}}{\partial T_k} = 2^{n-k} - a'(T_k) = 0$.

Theorem 4

Let $\eta \in (0, 2^{-n})$ be given. Then there is a unique solution $(T_1, ..., T_n) = (T_1^*, ..., T_n^*)$ with $T_k^* > 0$ for $k = 1, ..., n$ to the system

$$\begin{cases} \frac{a'(T_1)}{2^{n-1}} = \frac{a'(T_2)}{2^{n-2}} = ... = \frac{a'(T_n)}{2} \\ \sum_{k=1}^n a(T_k^*) = \log(1-\eta), \end{cases}$$

and $(T_1^*, ..., T_n^*)$ solves the optimization problem (3). Moreover, $T_1^* < T_2^* < \cdots < T_n^*$.

Sequential matches

- Each match in stage k is played until an information level is achieved.
- Define the posterior probability process

$$\Pi_t^{ij} := \mathbb{P}(\theta_{ij} = \frac{1}{2} | \mathcal{F}_t^{ij})$$

- Important assumption: symmetry.
- The games in the same round uses the same strategy. i.e., the match is thus played until $\tau_k^{ij} := \tau^{ij}(b_k) := \inf\{t \geq 0 : \Pi_t^{ij} \notin (b_k, 1 - b_k)\}$,
- Declare the winner to be Player i if $\Pi_{\tau_k}^{ij} = 1 - b_k$ and Player j if $\Pi_{\tau_k}^{ij} = b_k$.
- Denote the expected value of observation time τ_k

$$\mathcal{T}(b_k) := \mathbb{E}[\tau_k] = -2\psi(b_k).$$

- Consequently, the constrained sequential tournament design problem becomes

$$\inf_{\mathbf{b} \in (0, \frac{1}{2})^n} \sum_{k=1}^{n} 2^{n-k} \mathcal{T}(b_k) \quad \text{subject to} \quad \prod_{k=1}^{n} (1 - b_k) \geq 1 - \eta. \tag{4}$$

Sequential matches

- Similar to the fixed-time case, define

$$\mathcal{L}(\mathbf{b}, \lambda) := \sum_{k=1}^{n} 2^{n-k} \mathcal{T}(b_k) - \lambda \left(\sum_{k=1}^{n} \log(1 - b_k) - \log(1 - \eta) \right).$$

- The first order condition gives

$$\frac{\partial \mathcal{L}}{\partial b_k} = 2^{n-k} \mathcal{T}'(b_k) + \lambda \frac{1}{1 - b_k} = 0.$$

- $(1 - b)\mathcal{T}'(b)$ is strictly negative and increasing.

Theorem 5

Let $\eta \in (0, 2^{-n})$ be given. Then there is a unique solution $(b_1, ..., b_n) = (b_1^*, ..., b_n^*)$ with $b_k^* \in (0, 1/2)$ for $k = 1, ..., n$ to the system

$$\begin{cases} 2^{n-k}(1 - b_k)\mathcal{T}'(b_k) = 2^{n-l}(1 - b_l)\mathcal{T}'(b_l), \\ \sum_{k=1}^{n} \log(1 - b_k) = \log(1 - \eta). \end{cases}$$

Moreover, $(b_1^*, ..., b_n^*)$ solves the optimization problem (4), and $b_1^* \geq b_2^* \geq \cdots \geq b_n^*$.

Main takeaways

1. **Should always play sequential.**
2. **If we have to play fixed-length, let's observe more at later rounds.**

$$T_1^* \leq T_2^* \leq \cdots \leq T_n^*.$$

3. **When we play sequential, should also observe more at later rounds.**

$$b_1^* \geq b_2^* \geq \cdots \geq b_n^*, \qquad \mathbb{E}[\tau_1^*] \leq \mathbb{E}[\tau_2^*] \leq \cdots \leq \mathbb{E}[\tau_n^*].$$

Remark. Cost of each game affects the monotonicity.

Thank you!