Learning from Pairwise Preferences

Models, Estimation, Algorithms, and Recovery Guarantees

Suryanarayana Sankagiri

Wadhwani School of Data Science and AI
Indian Institute of Technology Madras

Personal webpage | Email

DA5453 — Learning from Human Preferences
First offered at IIT Madras, July–November 2026
Course webpage

First draft, August 2026

AI-use disclaimer. This document was generated with substantial assistance from AI tools and may contain hallucinations, mathematical errors, omissions, or passages that are difficult to read. Nevertheless, I, Suryanarayana Sankagiri, accept full responsibility for the material presented here. Readers are encouraged to contact me about any suspected error or unclear passage.

Preface

These notes accompany Module 1 of DA5453, *Learning from Human Preferences*, and were prepared for the first offering of the course at the Indian Institute of Technology Madras, from July to November 2026. They study how latent utilities may be inferred from noisy comparisons and choices. The Bradley–Terry pairwise model provides a running example through which modeling, statistical estimation, optimization, graph structure, spectral algorithms, and finite-sample analysis are developed within one coherent setting. Course information and updated public materials are available on the course webpage.

The chapters follow the mathematical development of the course and are intended to be read actively. Short exercises reinforce nearby definitions and calculations, while longer exercises complete proofs and develop extensions; sparse hints appear in Appendix A, and two model solutions in Appendix B illustrate the expected level of mathematical writing. Coding challenges are optional. The table of contents, citations, mathematical references, exercises, hints, solutions, and external readings are clickable.

Contents

Notation

Symbols are defined again when they first appear. This table collects the notation used from the beginning of the notes; graph and Markov-chain notation is introduced later, where it is needed.

Symbol	Meaning
Items, utilities, and choice models	
$[n]$	Item set $\{1, \dots, n\}$.
$S \subseteq [n]$	A displayed choice set.
$\theta_i \in \mathbb{R}$	Latent utility of item i; $\theta = (\theta_1, \dots, \theta_n)$.
$w_i = e^{\theta_i} > 0$	Positive BTL weight.
$\sigma(z) = 1/(1 + e^{-z})$	Logistic function.
$p_{ij}(\theta)$	Model probability that i defeats j.
Observed data and estimation	
$Y_{ij}^{(t)}$	Indicator that i defeats j in their t-th comparison.
Y_{ij}	Number of observed victories of i over j.
$N_{ij} = Y_{ij} + Y_{ji}$	Total number of comparisons between i and j.
$\mathcal{D}$, $\mathcal{D}_0$	A generic comparison dataset, and the four-item running dataset introduced in Chapter 2.
$\mathcal{L}(\theta)$	Likelihood of the observed data as a function of θ.
$\ell(\theta) = \log \mathcal{L}(\theta)$	Log-likelihood.
$L(\theta) = -\ell(\theta)$	Negative log-likelihood.
$\theta^\star$, $\widehat{\theta}$	True utility vector and an estimator computed from observed data.
$\star$, $\widehat{\ }$	A superscript $\star$ marks a population or true quantity; a hat marks an empirical quantity. For example, $p_{ij}^\star$ is a true comparison probability and $\widehat{p}_{ij} = Y_{ij}/N_{ij}$ its empirical estimate.
$\mathbb{P}$, $\mathbb{E}$	Probability and expectation under the data-generating model.
Vectors	
$\mathbf{1}$, e_i	All-ones vector and the i-th standard basis vector.
$\langle x, y \rangle$, $\|x\|_1, \|x\|_2, \|x\|_\infty$	Euclidean inner product; the sum, Euclidean, and maximum norms.

Chapter 1

Human preferences and observed choices

Let us begin by asking a few basic questions. What are human preferences? How are they expressed in the choices people make? Why might we want to learn about human preferences from such data? We will use these questions to clarify the learning problem before introducing mathematical models.

1.1 What are human preferences, and how are they expressed in choices?

When we say that a person prefers one alternative to another, we mean that the person favors it in some respect. The alternatives may be explicit, as in "tea or coffee?", or implicit, as in "I prefer not to answer." The preference may depend on the person, the purpose of the decision, and the surrounding circumstances. We need not assume that it forms a complete or permanent ranking of every possible alternative.

Definition 1.1.1 (Preference and choice)**.** A *preference* is a latent disposition to favor some alternatives over others, possibly conditional on a person and a context. A *choice* is an observed action taken when particular alternatives are available.

Let us keep the distinction between these two ideas in view. We cannot observe a person's preference directly; we observe actions that provide evidence about it. A commuter chooses between a bus, a train, and a cab after seeing fares and travel times. A viewer selects one film from a home screen, watches several trailers, or leaves without choosing. A user clicks a search result, perhaps after scrolling past several others. An annotator sees two responses to the same prompt and selects the better one. In every case, the action is shaped both by preference and by the situation in which it was elicited.

We should therefore not expect choices to reproduce a fixed ranking. The same person may act differently on repeated occasions; different people may disagree; and a change in price, prompt, display position, or available alternatives may change the action. A choice is evidence about preference, not a noiseless label for it.

1.1.1 A general description of choice data

Let us describe one interaction abstractly by

$$Z_t = (u_t, c_t, \mathbf{S}_t, \mathbf{X}_t, A_t). \tag{1.1}$$

Here u_t identifies the decision maker when that information is available; c_t is the context, such as a query, prompt, task, location, or time; $\mathbf{S}_t = (i_{t1}, \dots, i_{tK_t})$ records the displayed alternatives; $\mathbf{X}_t$ contains any observed item features, such as price, travel time, or response length; and A_t is the observed action. Components that are not recorded may simply be omitted. A dataset is a collection $\mathcal{D} = \{Z_t\}_{t=1}^{N}$ of such interactions.

We will keep this notation intentionally broad. The display $\mathbf{S}_t$ is an ordered list when position matters and can be treated as a set S_t when it does not. The admissible action space also varies. An interaction may allow exactly one selection, no selection, several selections, or a partial or complete ranking. A binary comparison is the special case $S_t = \{i, j\}$, with the outcome often encoded as

$$Y_{ij}^{(t)} = \begin{cases} 1, & \text{if } i \text{ is selected over } j, \\ 0, & \text{if } j \text{ is selected over } i. \end{cases}$$

By contrast, a multiway choice records $i_t \in S_t$, possibly with an outside option $i_t = \varnothing$. A click log may contain an ordered list and zero, one, or several clicks. These are different observation structures and need not be described by the same model.

Setting	Alternatives and context	Observed action	Structure that matters
LLM response evaluation	Two or more responses to a common prompt	A preferred response, a ranking, or a tie	The prompt is a target shared by the responses
Transport or retail	Modes or products with prices and attributes	One choice, possibly no purchase	Multiway substitution and an outside option
Search or recommendation	A displayed, usually ordered list for a query or user	Zero, one, or several clicks	Position, exposure, and item features
Streaming or a social feed	A sequence of films, posts, or videos	Watch, skip, stop, or several interactions	Order, duration, and sequential behavior

Key idea

Before choosing a statistical model, specify what was available, what was observed, whether order mattered, which side information was recorded, and what the observation leaves unknown.

Observational activity

For each row of the table, decide whether the alternatives are binary or multiway; whether zero, one, or multiple actions are possible; whether order matters; and which person, context, target, and item features are observed. Then identify one conclusion that would be unjustified from the recorded action alone. Compare the resulting answers: the differences indicate why a single observation model cannot be used indiscriminately in every setting.

These distinctions are not merely matters of data format. A non-click can mean that an item was disliked, that it was never noticed, or that the user chose to do nothing. Choosing i from $\{i, j, k\}$ is not the same observation as recording two separate victories $i \succ j$ and $i \succ k$. Likewise, an item shown first may receive more attention even when it is not preferred. We must therefore describe how the choice was presented and recorded before deciding what the observation says about preference.

1.2 Why learn from human preferences?

Let us now ask what we hope to accomplish by learning from preference data. Five recurring goals are to *aggregate*, *rank*, *predict*, *decide*, and *align*. What counts as a successful representation depends on which of these goals is intended.

1.2.1 Aggregation

Aggregation combines the preferences of several people into a collective outcome. An election selects a winner from individual ballots; a committee combines reviewers' judgments; and a group recommender may need to choose one film for viewers with different tastes. The output may be a single alternative or a social ranking of all alternatives.

This problem is not solved merely by collecting more data. Disagreement may represent genuinely different preferences rather than measurement noise, and the aggregation rule determines how those differences are weighted. Majority rule, scoring rules, and other social-choice procedures embody different trade-offs; none is automatically neutral. Chapter 5 of Truong et al. (2025) develops this connection between classical social choice and modern preference-learning systems.

1.2.2 Ranking

Sometimes the desired output is an ordering rather than a complete model of choice. A search engine ranks webpages for a query; a recommender ranks films for a user; and a conference or tournament may infer an approximate ordering from incomplete pairwise outcomes. Preference data are useful here because direct numerical scores may be unavailable or incomparable, while a judgment that i is preferred to j is relatively natural to obtain.

A ranking can be useful even when the associated choice probabilities are not accurately calibrated. Conversely, a global ranking may be inappropriate when relevance depends strongly on the user, query, or displayed alternatives. The model must therefore match the intended ranking task and the mechanism that generated its comparisons or clicks.

1.2.3 Prediction

Prediction asks how a person or population will choose in a situation not yet observed. The alternatives may be familiar but presented in a new combination, or the choice set may contain a genuinely new option. The classic BART study illustrates the second case: a discrete-choice model learned from commuters' existing travel choices and the attributes of their transport options forecast that BART would receive 6.3% of work trips, compared with an observed share of 6.2%; the contemporary official forecast had been about 15% (McFadden 2002).

The point of such a model is not only to describe choices already seen, but to transfer the learned trade-offs—for example, between time, cost, and convenience—to a new decision environment. This use is more demanding than fitting historical data: prediction can fail if the new setting changes an important feature that the model omitted.

1.2.4 Decision-making

Prediction is often an intermediate step toward choosing what to do. Suppose a retailer has ten products but shelf space for only three. A choice model predicts how customers substitute among the products in any proposed assortment; a decision procedure then chooses the assortment that best serves an objective. Selecting the three individually most popular products need not be optimal, because similar products may compete for the same customers while a more varied set serves a broader range of preferences.

The same pattern appears when a platform decides what to recommend, how to order an interface, or which comparisons to request next. The final decision also depends on constraints and on the objective—revenue, user satisfaction, social welfare, or some combination of them. Chapter 4 of Truong et al. (2025) studies how learned preferences guide decisions under uncertainty.

We should not conflate the first four goals. A method may recover the correct ordering while estimating choice probabilities poorly. A predictive model may work for familiar choice sets but be unreliable when used to design a new assortment. Aggregating a population also raises questions that do not arise when modeling one person's repeated choices. Alignment adds a further use: preference data can guide the behavior of the learned system itself.

1.2.5 LLM alignment as an example

To see how preference data can guide the learned system itself, consider a language model. A model trained to predict the next token is not thereby trained to follow a user's intent. Qualities such as usefulness, clarity, and safety are difficult to specify completely through a single hand-written objective, but people can often compare two candidate responses to the same prompt.

An explicit preference record in this setting may take the form

$$(x, y^+, y^-),$$

where x is a prompt and a human evaluator prefers response y^+ to y^-. Many such comparisons can be used to learn a model of the recorded judgments; that model can then help select or train future responses. The InstructGPT pipeline of Ouyang et al. (2022), for example, combined human demonstrations with rankings of model outputs and reinforcement learning from human feedback.

This example also exposes an important qualification. "Human preference" is not a single universal object. Judgments depend on the evaluator, the instructions given to evaluators, the prompt distribution, and the values that the data-collection procedure asks them to apply. Learning from preferences therefore requires care not only in estimation, but also in deciding whose preferences are represented and how they are elicited.

1.2.6 How will we learn?

We will proceed by constructing a probabilistic model for the observed action, schematically

$$\mathbb{P}_\theta(A_t = a \mid u_t, c_t, \mathbf{S}_t, \mathbf{X}_t)\,, \tag{1.2}$$

and then estimate its unknown parameter θ from $\mathcal{D}$. A probabilistic model accommodates variability and disagreement while making explicit which aspects of the observation are being explained.

In the next chapter we begin with a random-utility description of choice and derive concrete probability models from it. The chapters that follow study likelihood-based estimation, the geometry of comparison data, a Markov-chain algorithm for ranking, and finite-sample recovery guarantees. Together they instantiate the progression

observed choices $\longrightarrow$ probabilistic model $\longrightarrow$ estimation algorithm $\longrightarrow$ prediction, ranking, or decision.

Further reading

The freely accessible text by Truong et al. (2025) treats preference learning, decisions, and aggregation in Chapters 2, 4, and 5, respectively, while its foundations chapter discusses recommendation and language-model alignment. The use of ranked human evaluations in instruction-following language models is described by Ouyang et al. (2022); the BART forecasting example is recounted by McFadden (2002).

Chapter 2

Random utility and probabilistic choice

We have described several ways in which choices can be observed. Let us now ask how those choices might arise. A first attempt would assign each alternative a fixed score and predict that the alternative with the larger score is always chosen. This is too rigid. If two responses to the same prompt are compared many times, one response may be preferred more often without being preferred every time. Different evaluators may disagree, and even one evaluator may reverse a choice on another occasion.

We therefore need to represent both a systematic tendency and the variation left unexplained by it. A random-utility model does exactly this. It retains the idea that people choose alternatives they value more highly, but allows the relevant utility to vary across people, contexts, and occasions.

2.1 Random utility

Let $S \subseteq [n]$ be a finite choice set. On a particular occasion, item $i \in S$ has realized utility

$$U_i = \theta_i + \varepsilon_i, \tag{2.1}$$

where $\theta_i \in \mathbb{R}$ is its systematic utility and ε_i is an occasion-specific random term. We assume that the selected item is

$$\arg\max_{i \in S} U_i.$$

The vector θ encodes stable differences represented by the model. The noise variables collect whatever is not represented there: unrecorded context, heterogeneous judgments, fluctuations in attention, or other omitted attributes.

For example, suppose that two responses i and j to one prompt are shown to an evaluator. The difference $\theta_i - \theta_j$ may represent a systematic tendency to favor one response. The terms ε_i and ε_j may reflect which evaluator was sampled, how much that evaluator values concision on this occasion, whether a subtle error was noticed, or some feature of the responses that our model does not record. Conditional on the realized utilities, the larger one is selected; to us, the choice is random because we do not observe all of these components.

This is a rational-choice assumption, but a limited one. We assume that the decision maker selects the alternative with the greatest *realized* utility. We do not assume that θ_i completely describes a person's preference, or that the same alternative must be selected on every repetition. The substantive behavioral assumptions enter through the form chosen for θ and through the joint distribution of the noise terms.

For two items,

$$\mathbb{P}(i \succ j) = \mathbb{P}(U_i > U_j) = \mathbb{P}(\varepsilon_j - \varepsilon_i < \theta_i - \theta_j).$$

Only the utility difference appears. This is natural: adding the same constant to both alternatives cannot change which one has the larger realized utility. If the noise difference has a continuous distribution function F_{ij}, then

$$\mathbb{P}(i \succ j) = F_{ij}(\theta_i - \theta_j). \tag{2.2}$$

The distribution of the noise difference therefore determines how a utility gap is converted into a choice probability.

The qualitative behavior is already clear. Under symmetric noise, equal systematic utilities give probability $1/2$. If $\theta_i - \theta_j$ is large and positive relative to the scale of the noise, i is chosen almost deterministically. When the gap is small, ordinary fluctuations can reverse the ordering, so the outcome is much less predictable. This is precisely the behavior we would want from a probabilistic comparison model.

2.1.1 Gaussian noise and the probit model

Suppose first that the errors are independent standard normal random variables:

$$\varepsilon_i, \varepsilon_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1).$$

Then $\varepsilon_j - \varepsilon_i \sim \mathcal{N}(0,2)$, and hence

$$\mathbb{P}(i \succ j) = \Phi\left(\frac{\theta_i - \theta_j}{\sqrt{2}}\right),$$

where Φ is the standard normal distribution function. This is the *probit model*. With errors of variance τ^2, the denominator becomes $\sqrt{2}\,\tau$.

The normal distribution is centered and symmetric, so a zero utility gap gives probability 1/2. Its cumulative distribution function is increasing and S-shaped: moderate gaps produce uncertain choices, while gaps that are several noise standard deviations wide produce probabilities close to zero or one. Probit is therefore a perfectly sensible random-utility model. For a binary comparison, the normal CDF gives everything we need. The situation changes when a person must choose one alternative from a set of three or more. We then need the probability that one realized utility exceeds *all* the others. With Gaussian errors this probability is valid but generally requires a multidimensional normal integral. It no longer reduces to the one-dimensional formula above.

2.2 Why Gumbel noise?

Suppose a commuter chooses among a car, a bus, and a train. The bus is selected only when its realized utility exceeds the realized utilities of *both* other modes. Three separately specified pairwise probabilities do not by themselves define the probability of this event; we need a joint model for all three realized utilities. Random utility supplies that joint experiment.

Let us therefore ask for a noise model with three properties. It should retain the interpretation that large utility gaps lead to predictable choices; it should define one coherent choice rule for every finite set S; and it should give probabilities simple enough to interpret and estimate. Independent Gumbel noise provides exactly this combination.

The Gumbel distribution is also called the type-I extreme-value distribution because it arises naturally in the study of maxima. That connection makes it a plausible language for a model in which the selected item is the one with the largest realized utility. The assumption remains a modeling choice: it does not assert that every unrecorded influence on a human decision is literally Gumbel distributed. Its practical attraction is that maximizing Gumbel-perturbed utilities leads to a remarkably simple multiway probability formula.

Assume, then, that the noise terms are independent standard Gumbel random variables. A standard Gumbel random variable has distribution function and density

$$F_\varepsilon(x) = \exp(-e^{-x}), \qquad f_\varepsilon(x) = e^{-x}\exp(-e^{-x}).$$

For the realized utility in Equation (2.1), we then have

$$F_i(u) = \mathbb{P}(U_i \le u) = \exp(-e^{\theta_i - u}),$$

and

$$f_i(u) = e^{\theta_i - u}\exp(-e^{\theta_i - u}).$$

Because these distributions are continuous, ties occur with probability zero.

Fix $i \in S$. Conditioning on $U_i = u$, item i is selected exactly when every other realized utility is below u. Independence gives

$$\mathbb{P}(i \mid S) = \int_{-\infty}^{\infty} f_i(u) \prod_{j \in S\setminus\{i\}} F_j(u)\, du \tag{2.3}$$

$$= \int_{-\infty}^{\infty} e^{\theta_i} e^{-u} \exp\left(-e^{-u}\sum_{j\in S} e^{\theta_j}\right) du. \tag{2.4}$$

Let $C = \sum_{j \in S} e^{\theta_j}$ and substitute $z = Ce^{-u}$. The integral evaluates to e^{θ_i}/C. We have therefore proved the following result.

Proposition 2.2.1 (Gumbel-max choice probability)**.** *Under independent Gumbel errors in Equation (2.1),*

$$\mathbb{P}(i \mid S) = \frac{e^{\theta_i}}{\sum_{j \in S} e^{\theta_j}}. \tag{2.5}$$

We will call the family in Equation (2.5) the *Bradley–Terry–Luce model*, abbreviated BTL. The name records two closely related developments. Bradley and Terry introduced the binary comparison model; Luce developed the share-of-weight choice rule for arbitrary finite sets (R. A. Bradley and Terry 1952; Luce 1959). Terminology differs across fields and observation types. To keep the present notes consistent, BTL will refer to the common exponential-weight family, while its two-item restriction will be called the Bradley–Terry model.

Increasing θ_i increases the probability of choosing i, while adding more attractive alternatives to S lowers its share. The Gumbel construction has thus turned noisy utility maximization into a tractable multiway choice model whose likelihood can be optimized.

2.3 The Bradley–Terry pairwise model

For the two-element set $S = \{i, j\}$, Equation (2.5) becomes

$$p_{ij}(\theta) := \mathbb{P}(i \succ j) = \frac{e^{\theta_i}}{e^{\theta_i} + e^{\theta_j}} = \frac{1}{1 + e^{-(\theta_i - \theta_j)}} = \sigma(\theta_i - \theta_j). \tag{2.6}$$

Here $\sigma(z) := (1 + e^{-z})^{-1}$ is the logistic function. The pairwise probability model in Equation (2.6) is the *Bradley–Terry model*. It is the two-item restriction of the BTL family.

The logistic function satisfies

$$\sigma(-z) = 1 - \sigma(z), \qquad \sigma'(z) = \sigma(z)(1 - \sigma(z)).$$

Consequently $p_{ji} = 1 - p_{ij}$, and increasing $\theta_i - \theta_j$ increases the probability that i wins.

Example 2.3.1 (Utility gap)**.** If $\theta_i - \theta_j = \log 3$, then

$$p_{ij} = \frac{3}{3 + 1} = \frac{3}{4}.$$

Thus exponentiated utility differences have an odds interpretation:

$$\frac{p_{ij}}{p_{ji}} = e^{\theta_i - \theta_j}.$$

2.4 The running dataset

We use the following four-item dataset, denoted by $\mathcal{D}_0$, throughout the notes. One may imagine four teams playing repeated matches, four products being compared by users, or four responses being judged in pairs. Every unordered pair in $\{A, B, C, D\}$ is compared ten times.

Pair	Wins by first item	Wins by second item
A, B	$A : 7$	$B : 3$
A, C	$A : 8$	$C : 2$
A, D	$A : 9$	$D : 1$
B, C	$B : 6$	$C : 4$
B, D	$B : 8$	$D : 2$
C, D	$C : 6$	$D : 4$

There are sixty observations in total, and the total observed wins are

$$A : 24, \qquad B : 17, \qquad C : 12, \qquad D : 7.$$

Because the design is balanced, simply counting wins gives the plausible order $A \succ B \succ C \succ D$. The corresponding overall win fractions are

$$\widehat{r}_A = \frac{24}{30} = 0.800, \qquad \widehat{r}_B = \frac{17}{30} \approx 0.567, \qquad \widehat{r}_C = \frac{12}{30} = 0.400, \qquad \widehat{r}_D = \frac{7}{30} \approx 0.233.$$

This apparent success depends heavily on every item facing every other item the same number of times.

To see the problem, suppose that only the adjacent comparisons A–B, B–C, and C–D remain. Their outcomes are 7:3, 6:4, and 6:4. The raw win totals become

$$A : 7, \qquad B : 9, \qquad C : 10, \qquad D : 4.$$

Counting wins would now place C above B and A, even though the recorded head-to-head chain points in the direction $A \succ B \succ C \succ D$. The error comes from treating every win as equally informative while ignoring both the number and the strength of the opponents.

Dividing wins by matches played repairs the unequal match counts but not the unequal schedules. In the same reduced dataset the observed win fractions are

$$0.70, \qquad 0.45, \qquad 0.50, \qquad 0.40$$

for A, B, C, D, respectively. This still puts C above B, because C has the opportunity to collect wins against D, whereas B faces A as well as C. An average win rate does not distinguish a difficult schedule from a weak item.

Selective missingness can be more damaging still. From the ten A–C comparisons, retain both losses by A but only three of its eight wins. The recorded win fraction for A in this surviving sample is then $3/5 = 0.60$. If all ten B–D comparisons are retained, the recorded win fraction for B is $8/10 = 0.80$. If these are the only surviving records involving A and B, a naive comparison of win fractions ranks B above A, even though the complete dataset strongly suggests the reverse. This deliberately extreme construction makes a general point: when the probability that a record is missing depends on its outcome, no ranking method can repair the bias without additional assumptions about the missingness mechanism.

The Bradley–Terry model addresses the first two difficulties by retaining the pair-level outcomes and estimating all item strengths jointly. Information can pass along the observed chain: comparisons of A with B and of B with C contain indirect information about A relative to C. It cannot, however, overcome every defect in the data. If the comparison graph is disconnected, some relative strengths cannot be identified; if outcomes are selectively hidden, the fitted model describes the recorded sample rather than the missing data. These limitations will return when we study identifiability and statistical recovery.

2.5 Weights and utilities

Define $w_i = e^{\theta_i} > 0$. Then

$$\mathbb{P}(i \mid S) = \frac{w_i}{\sum_{j \in S} w_j}, \qquad p_{ij} = \frac{w_i}{w_i + w_j}. \tag{2.7}$$

The weight parameterization makes the share-of-total-strength interpretation explicit. The utility parameterization is more convenient for differentiation and convex optimization because probabilities depend on linear differences $\theta_i - \theta_j$.

Conversely, any positive weights define utilities through $\theta_i = \log w_i$. A common multiplication of all weights corresponds to a common additive shift of all utilities.

2.6 Shift invariance and identifiability

For any constant c,

$$\frac{e^{\theta_i+c}}{\sum_{j\in S} e^{\theta_j+c}} = \frac{e^c e^{\theta_i}}{e^c \sum_{j\in S} e^{\theta_j}} = \frac{e^{\theta_i}}{\sum_{j\in S} e^{\theta_j}}.$$

Hence choice probabilities are unchanged under

$$\theta \mapsto \theta + c\mathbf{1}.$$

The absolute level of utility is not identifiable. Only differences are.

A normalization selects one representative from each equivalence class. Common choices are

$$\mathbf{1}^\top \theta = 0, \qquad \theta_n = 0, \qquad \text{or} \qquad \sum_i w_i = 1.$$

The centered constraint will be used throughout these notes because it treats all items symmetrically and is compatible with Euclidean geometry.

2.7 Independence of irrelevant alternatives

Definition 2.7.1 (IIA)**.** A choice model satisfies *independence of irrelevant alternatives* if, for any i, j and any choice sets S, T containing both items,

$$\frac{\mathbb{P}(i \mid S)}{\mathbb{P}(j \mid S)} = \frac{\mathbb{P}(i \mid T)}{\mathbb{P}(j \mid T)}.$$

For BTL,

$$\frac{\mathbb{P}(i \mid S)}{\mathbb{P}(j \mid S)} = \frac{e^{\theta_i}}{e^{\theta_j}},$$

which is independent of the other alternatives in S. Thus BTL satisfies IIA.

IIA is a statement about an odds ratio, not about equality between pairwise and multiway probabilities. For example, with $\theta_A = 2$, $\theta_B = 1$, and $\theta_C = 0$,

$$\mathbb{P}(A \succ B) = \frac{e^2}{e^2 + e}, \qquad \mathbb{P}(A \mid \{A, B, C\}) = \frac{e^2}{e^2 + e + 1}.$$

The second probability is smaller because a third item can be chosen. What remains constant is the odds of A relative to B.

Let us work through the standard bus example rather than assume the conclusion. A commuter initially chooses between a car and a blue bus. Suppose the two have equal BTL weights. The model then gives

$$\mathbb{P}(\text{car}) = \frac{1}{2}, \qquad \mathbb{P}(\text{blue bus}) = \frac{1}{2}.$$

Now the transport authority adds a red bus whose route, travel time, and service are almost identical to those of the blue bus. A natural behavioral prediction is that the original bus demand will be divided between the two colors, while the overall attraction of traveling by bus rather than car remains roughly the same:

$$\mathbb{P}(\text{car}) \approx \frac{1}{2}, \qquad \mathbb{P}(\text{blue bus}) \approx \frac{1}{4}, \qquad \mathbb{P}(\text{red bus}) \approx \frac{1}{4}.$$

BTL cannot produce these numbers when the three alternatives have equal weights. It instead assigns probability 1/3 to each. Equivalently, IIA forces the car-to-blue-bus odds to remain 1:1 after the red bus is introduced. The model does not recognize that the two buses are much closer substitutes for one another than either is for the car. Their unobserved utilities may share route, comfort, or mode-specific factors, contradicting the assumption of independent noise. The example therefore reveals both the benefit and the price of BTL: IIA gives a clean, estimable formula, but it can impose an unrealistic pattern of substitution when alternatives are similar (Train 2009, Section 3.3.2).

2.8 Luce's representation

IIA essentially characterizes the share-of-weight form.

Theorem 2.8.1 (Luce choice representation)**.** *Let $\mathcal{X}$ be finite. Suppose that for every nonempty $S \subseteq \mathcal{X}$, the probabilities $p(i \mid S)$ are strictly positive, sum to one, and satisfy IIA. Then there exist weights $w_i > 0$ such that*

$$p(i \mid S) = \frac{w_i}{\sum_{j \in S} w_j}.$$

Proof sketch. Fix a reference item 0, set $w_0 = 1$, and define $w_i = p(i \mid \{i, 0\})/p(0 \mid \{i, 0\})$. IIA implies that the odds of i against j equal w_i/w_j in every set containing them. Normalizing these odds over a set S yields the claimed probabilities. The complete argument is developed in Exercise 2.E4. □

Further reading

For a systematic treatment of random-utility models, Gumbel choice, IIA, and richer discrete-choice models, see Train (2009). The freely accessible Stanford text develops the same family in the setting of modern preference learning and explains how its names vary with the observation type (Truong et al. 2025). The binary model is due to R. A. Bradley and Terry (1952); Luce's axiom and its share-of-weight representation are developed in Luce (1959).

Coding challenge 2.C1 — Empirical Gumbel-max

Fix a centered, nonconstant vector of four utilities and repeatedly draw independent Gumbel errors. Compare empirical choice frequencies with Equation (2.5). Repeat after adding a common constant to all utilities and after multiplying all utilities by two. Explain the different outcomes.

2.9 Exercises

Exercise 2.E1 — Gumbel calculations

(a) Verify that $f_\varepsilon = F'_\varepsilon$.

(b) Derive F_i and f_i for $U_i = \theta_i + \varepsilon_i$.

(c) Starting from Equation (2.4), carry out the change of variable explicitly and recover Equation (2.5).

Exercise 2.E2 — Weights, odds, and normalization

(a) Show that arbitrary positive weights define BTL utilities and hence Bradley–Terry pairwise probabilities.

(b) Prove $p_{ij}/p_{ji} = w_i/w_j$.

(c) Compare centering θ, fixing $\theta_n = 0$, and normalizing w. Which symmetries does each choice preserve?

Exercise 2.E3 — IIA does not equate experiments

Take $\theta_A = 2$, $\theta_B = 1$, and $\theta_C = 0$.

(a) Compute $\mathbb{P}(A \succ B)$ and $\mathbb{P}(A \mid \{A, B, C\})$.

(b) Compare the likelihood of one three-way choice of A with the likelihood of two independent outcomes $A \succ B$ and $A \succ C$.

(c) Explain why the observations cannot be substituted for one another.

Exercise 2.E4 — Luce's theorem

Complete the proof of Theorem 2.8.1.

(a) For distinct $i, j \neq 0$, use IIA on $\{i, j, 0\}$ and its two-item subsets to show that $p(i \mid \{i, j\})/p(j \mid \{i, j\}) = w_i/w_j$. Explain separately why the identity holds when one item is the reference item 0.

(b) Extend the odds identity to an arbitrary set containing i, j.

(c) Use normalization to obtain the share-of-weight formula.

Hint available in Appendix A.

Exercise 2.E5 — When IIA is implausible

Give a concrete choice setting, different from the bus example, in which adding a new alternative should change the odds between two existing alternatives. State what contextual or similarity information the BTL model fails to represent.

Likelihood, estimation, and optimization

The Bradley–Terry model runs in one direction: a utility vector θ determines the probabilities of comparison outcomes. Statistical estimation runs in the other direction. We observe who was compared and who won, and we ask which candidate utility vectors explain those observations well.

Consider only the ten comparisons between A and D in $\mathcal{D}_0$. Item A won nine times. A candidate under which $p_{AD} = 0.9$ assigns the observed sequence a factor $0.9^9(0.1)$, whereas a candidate under which $p_{AD} = 0.5$ assigns it 0.5^{10}. The first factor is about forty times larger. This does not by itself estimate all four utilities: the same θ_A must also explain comparisons with B and C, and all the other utilities are shared across their own comparisons. The likelihood is the device that combines these coupled pieces of evidence.

The chapter develops the following route from observations to inferred preferences:

$$\underbrace{(\mathcal{X}, \mathcal{D})}_{\text{items and comparisons}} \longrightarrow \underbrace{L(\theta)}_{\text{statistical objective}} \longrightarrow \underbrace{\widehat{\theta}}_{\text{fitted utilities}} \longrightarrow \underbrace{p_{ij}(\widehat{\theta}) \text{ and a ranking}}_{\text{inferred preferences}}.$$

3.1 From individual outcomes to comparison counts

Recall that $Y_{ij}^{(t)}$ denotes the outcome of the t-th comparison of items i and j, oriented so that

$$Y_{ij}^{(t)} = \begin{cases} 1, & i \text{ defeats } j, \\ 0, & j \text{ defeats } i. \end{cases}$$

Conditional on the compared pair and a candidate parameter θ,

$$Y_{ij}^{(t)} \sim \text{Bernoulli}(p_{ij}(\theta)), \qquad p_{ij}(\theta) = \sigma(\theta_i - \theta_j).$$

We treat the set of compared pairs as fixed and assume that the recorded outcomes are conditionally independent given θ.

For every ordered pair $i \neq j$, define

$$Y_{ij} = \sum_{t=1}^{N_{ij}} Y_{ij}^{(t)}, \qquad N_{ij} = Y_{ij} + Y_{ji} = N_{ji}.$$

Thus Y_{ij} is the number of victories of i over j, and N_{ij} is the total number of their comparisons. We collect the counts in the generic dataset

$$\mathcal{D} = \{(N_{ij}, Y_{ij}, Y_{ji}) : 1 \leq i < j \leq n,\ N_{ij} > 0\}.$$

The symbol $\mathcal{D}_0$ continues to denote the four-item dataset introduced in Chapter 2.

For one pair, the empirical fraction Y_{ij}/N_{ij} estimates its win probability. The Bradley–Terry problem is more structured: all such probabilities must be generated by one common vector θ. Estimation therefore cannot generally be performed one pair at a time.

3.2 From a probability model to a likelihood

For one observed outcome $Y_{ij}^{(t)} = y$,

$$\mathbb{P}_\theta(Y_{ij}^{(t)} = y) = p_{ij}(\theta)^y (1 - p_{ij}(\theta))^{1-y}.$$

Multiplying these probabilities over an ordered sequence of N_{ij} comparisons gives

$$p_{ij}(\theta)^{Y_{ij}} p_{ji}(\theta)^{Y_{ji}}.$$

If only the aggregate count is regarded as the observation, its binomial probability contains the additional factor $\binom{N_{ij}}{Y_{ij}}$. This factor does not depend on θ, so both descriptions produce the same maximum-likelihood estimator (MLE).

After dropping all factors independent of θ, the likelihood may be written as

$$\mathcal{L}(\theta) \propto \prod_{i<j:N_{ij}>0} p_{ij}(\theta)^{Y_{ij}} p_{ji}(\theta)^{Y_{ji}}. \tag{3.1}$$

Maximum likelihood

Each candidate θ assigns probabilities to possible datasets. The likelihood evaluates the probability assigned to the dataset that was actually observed. It is not the probability that θ is true. Maximum likelihood selects the candidate under which the observed data are most probable among the candidates considered.

Taking logarithms converts the product into a sum and does not change the maximizer. After omitting the corresponding additive constant, we write

$$\ell(\theta) = \sum_{i<j:N_{ij}>0} \left[Y_{ij} \log p_{ij}(\theta) + Y_{ji} \log p_{ji}(\theta) \right]. \tag{3.2}$$

It is convenient to minimize the negative log-likelihood $L(\theta) = -\ell(\theta)$.

Maximum likelihood is a general procedure for statistical estimation. A model specifies a family of data distributions indexed by a parameter; the likelihood uses the observed sample to select one member of that family. Whether the resulting estimator exists, is unique, can be computed efficiently, and is close to the data-generating parameter are separate questions.

3.3 A useful form of the loss

For $i < j$, let $d_{ij} = \theta_i - \theta_j$. Since

$$-\log p_{ij}(\theta) = \log(1 + e^{-d_{ij}}),$$

the contribution from pair $\{i, j\}$ is

$$L_{ij}(\theta) = Y_{ij} \log(1 + e^{-d_{ij}}) + Y_{ji} \log(1 + e^{d_{ij}}) \tag{3.3}$$

$$= N_{ij} \log(1 + e^{d_{ij}}) - Y_{ij} d_{ij}. \tag{3.4}$$

The first line preserves the winner–loser interpretation. The second is especially convenient for differentiation.

3.4 Shift invariance and the centered estimator

Chapter 2 showed that all Bradley–Terry probabilities depend only on utility differences. Consequently,

$$L(\theta + c\mathbf{1}) = L(\theta) \qquad \text{for every } c \in \mathbb{R}.$$

If an unconstrained minimizer exists, every common shift of it is another minimizer. We choose one representative by defining the centered estimator

$$\widehat{\theta} \in \operatorname*{arg\,min}_{\theta \in \mathbb{R}^n:\ \mathbf{1}^\top \theta = 0} L(\theta). \tag{3.5}$$

This is a definition conditional on attainment. Centering removes the unavoidable common-shift ambiguity, but it does not by itself guarantee that a finite minimizer exists or that no other flat directions remain. These questions are the subject of Chapter 4.

3.5 The score equations

Let us first differentiate the contribution of one unordered pair. From Equation (3.4),

$$\frac{d}{dd_{ij}} L_{ij} = N_{ij}\sigma(d_{ij}) - Y_{ij} = N_{ij}p_{ij}(\theta) - Y_{ij}.$$

Since $d_{ij} = \theta_i - \theta_j$, this quantity enters coordinate i with a positive sign and coordinate j with a negative sign.

There is a small indexing point worth making explicit. If a pair appears in the log-likelihood with $j < i$, differentiating its contribution with respect to θ_i gives

$$\begin{aligned} -Y_{ji} + N_{ji}p_{ji}(\theta) &= Y_{ij} - N_{ij} + N_{ij}(1 - p_{ij}(\theta)) \\ &= Y_{ij} - N_{ij}p_{ij}(\theta) \end{aligned}$$

for the log-likelihood, and the opposite sign for the negative log-likelihood. Here we used

$$Y_{ij} + Y_{ji} = N_{ij}, \qquad p_{ij}(\theta) + p_{ji}(\theta) = 1.$$

Thus every comparison involving item i has the same predicted-minus-observed form, regardless of its ordering in the sum over $i < j$.

Proposition 3.5.1 (Gradient of the Bradley–Terry loss)**.** *For every item i,*

$$\frac{\partial L}{\partial \theta_i} = \sum_{j \neq i} \left(N_{ij}p_{ij}(\theta) - Y_{ij}\right). \tag{3.6}$$

The corresponding coordinate of the log-likelihood score $\nabla \ell$ is observed wins minus predicted wins; the loss gradient $\nabla L = -\nabla \ell$ is predicted wins minus observed wins. A positive loss-gradient coordinate therefore means that the current model overpredicts wins for item i; a gradient-descent step decreases θ_i.

Summing all coordinates gives zero. Every pair contributes once with each sign, so

$$\mathbf{1}^\top \nabla L(\theta) = 0. \tag{3.7}$$

This is the differential form of shift invariance.

Suppose a finite centered minimizer exists. First-order optimality says that the projection of $\nabla L(\widehat{\theta})$ onto the centered subspace is zero. Equation (3.7) says that the gradient already belongs to that subspace. Hence

$$\nabla L(\widehat{\theta}) = 0.$$

The fitted utilities therefore satisfy the *score equations*

$$\sum_{j \neq i} N_{ij}p_{ij}(\widehat{\theta}) = \sum_{j \neq i} Y_{ij} \qquad \text{for every } i. \tag{3.8}$$

At a finite optimum, the total number of wins predicted for each item equals its total observed wins.

This statement does not say that every fitted pairwise probability equals the corresponding empirical fraction. Such fractions may be mutually incompatible with any single utility vector. Maximum likelihood fits all comparisons simultaneously, and the same θ_i must serve in every pair containing item i.

3.6 From score equations to gradient descent

The score equations are coupled and nonlinear. Except in very small special cases, they are not solved by rearranging one equation at a time. Gradient descent supplies a direct numerical procedure for minimizing L. Given a step size $\eta > 0$, it iterates

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla L(\theta^{(t)}). \tag{3.9}$$

An item with more observed wins than currently predicted has a negative loss-gradient coordinate, so its utility increases in the next step. The update therefore implements the same correction suggested by the score equations.

If $\mathbf{1}^\top \theta^{(0)} = 0$, then Equation (3.7) gives

$$\mathbf{1}^\top \theta^{(t+1)} = \mathbf{1}^\top \theta^{(t)} - \eta \mathbf{1}^\top \nabla L(\theta^{(t)}) = 0.$$

Ordinary gradient descent thus remains in the centered subspace without an explicit projection.

Choose a centered initial vector $\theta^{(0)}$, a step size η, and a tolerance ε.
for $t = 0, 1, 2, \ldots$ **do**
  Compute $g^{(t)} = \nabla L(\theta^{(t)})$ using Equation (3.6).
  if $\left\| g^{(t)} \right\|_2 \le \varepsilon$ **then**
    stop
  end if
  Set $\theta^{(t+1)} = \theta^{(t)} - \eta g^{(t)}$.
end for

A fixed step size must be small enough for the local linear approximation to be reliable. In practice one may use a line search or an adaptive method. These are choices of optimization algorithm; the statistical estimator remains the minimizer of the specified loss.

3.7 What convexity does, and does not, guarantee

Definition 3.7.1 (Convexity, strict convexity, and strong convexity)**.** Let f be defined on a convex set. It is *convex* if

$$f((1-t)x + ty) \le (1-t)f(x) + tf(y) \qquad (0 \le t \le 1).$$

It is *strictly convex* if the inequality is strict whenever $x \ne y$ and $0 < t < 1$. A differentiable function is μ-strongly convex if

$$f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|y - x\|_2^2 \qquad \text{for all } x, y.$$

Convexity rules out spurious local minima. Strict convexity gives at most one minimizer, but does not guarantee that a minimizer exists. Strong convexity is a quantitative curvature condition; it implies strict convexity and turns a small gradient into a parameter-error bound.

The Bradley–Terry negative log-likelihood is convex. Chapter 4 derives its Hessian and shows exactly how the comparison graph determines its flat directions. The common-shift direction is always flat, so positive curvature can hold only after normalization. Even then, pointwise strict convexity need not provide one uniform strong-convexity constant over an unbounded parameter space.

For a representative optimization statement, suppose a convex differentiable function has a minimizer and a β-Lipschitz gradient, meaning that

$$\|\nabla f(x) - \nabla f(y)\|_2 \le \beta \|x - y\|_2 \qquad \text{for all } x, y.$$

With a suitable fixed step size, such as $0 < \eta \le 1/\beta$, gradient descent obtains

$$f(x^{(t)}) - f(x^\star) = O(1/t).$$

If the function is additionally μ-strongly convex on the relevant subspace, the convergence is geometric. For example, with $\eta = 1/\beta$,

$$\left\|x^{(t)} - x^\star\right\|_2^2 \le \left(1 - \frac{\mu}{\beta}\right)^t \left\|x^{(0)} - x^\star\right\|_2^2.$$

These conclusions require their stated smoothness, curvature, and existence assumptions. Convexity alone does not settle the existence question.

3.8 Optimization error and statistical error

Let $\theta^\star$ denote the utility vector that generated the data. Two different errors arise:

$$\underbrace{\left\|\theta^{(t)} - \widehat{\theta}\right\|_2}_{\text{optimization error}} \qquad \text{and} \qquad \underbrace{\left\|\widehat{\theta} - \theta^\star\right\|_2}_{\text{statistical error}}.$$

The first measures how accurately an algorithm has solved the empirical optimization problem. It can be reduced by computation. The second remains even at the exact empirical minimizer because the observed outcomes are random. It is controlled by sample size, comparison design, model curvature, and sampling noise. Chapter 6 develops one finite-sample bound.

We can now complete the estimation pipeline promised at the start of the chapter:

$$\mathcal{D} \xrightarrow{\text{Bradley–Terry likelihood}} L(\theta) \xrightarrow{\text{gradient method}} \widehat{\theta} \xrightarrow{p_{ij}(\widehat{\theta})=\sigma(\widehat{\theta}_i-\widehat{\theta}_j)} \text{probabilities and a ranking.}$$

The remaining structural question is whether the middle object is finite and well determined. Chapter 4 answers it using two graphs associated with the comparison data.

Further reading

For the general principle of maximum likelihood, see Fan (2016). For convexity, smoothness, strong convexity, and gradient methods, see Chapters 3 and 9 of the freely accessible text by Boyd and Vandenberghe (2004). Section 2.3 of Truong et al. (2025) contains runnable Bradley–Terry maximum-likelihood code and a discussion of model fitting.

Coding challenge 3.C1 — Centered Bradley–Terry maximum likelihood

Implement Equation (3.9) for $\mathcal{D}_0$. Check the analytic gradient against a finite-difference approximation at several randomly chosen centered vectors. During optimization, record $L(\theta^{(t)})$, $\left\|\nabla L(\theta^{(t)})\right\|_2$, and $\mathbf{1}^\top\theta^{(t)}$. Compare a stable step size with one that is deliberately too large, and explain the observed behavior. Report the fitted ranking and fitted pairwise probabilities.

3.9 Exercises

Exercise 3.E1 — Sequence data and count data

Suppose one ordered sequence of five comparisons of i with j is $1, 0, 1, 1, 0$. Write its likelihood exactly. Then write the probability of observing the aggregate count $Y_{ij} = 3$, identify the additional combinatorial factor, and explain why the two likelihoods have the same maximizer.

Exercise 3.E2 — Likelihood of the running dataset

Write Equation (3.2) explicitly for the six unordered pairs in $\mathcal{D}_0$. Then write the contribution of pair $\{A, D\}$ in both forms appearing in Equation (3.4).

Exercise 3.E3 — Gradient and one centered update: model solution

Consider three items with comparison records

$$A\text{–}B : 3{:}1, \qquad A\text{–}C : 2{:}2, \qquad B\text{–}C : 3{:}1,$$

where the first count in each pair belongs to the first named item. In the order (A, B, C):

(a) compute the observed total wins;

(b) compute $\nabla L(0)$;

(c) verify that the gradient is centered;

(d) compute one gradient-descent update from the origin and interpret its ordering.

A model solution appears in Appendix B.

Exercise 3.E4 — Score equations need not fit every pair

Three items are compared four times per pair. The empirical results satisfy

$$\widehat{p}_{AB} = \widehat{p}_{BC} = \widehat{p}_{CA} = \frac{3}{4}.$$

Show that no Bradley–Terry utility vector can reproduce all three empirical fractions exactly. Nevertheless, show that $\theta = 0$ satisfies the score equations. Use the graph results of Chapter 4, after reading that chapter, to explain why it is the unique centered MLE.

Exercise 3.E5 — Centered first-order optimality

Let $\mathcal{C} = \{\theta : \mathbf{1}^\top \theta = 0\}$. Prove that first-order optimality of $\widehat{\theta}$ on $\mathcal{C}$ implies $\nabla L(\widehat{\theta}) \in \operatorname{span}\{\mathbf{1}\}$. Combine this with Equation (3.7) to obtain $\nabla L(\widehat{\theta}) = 0$.

Exercise 3.E6 — Shift invariance and the gradient

Prove $L(\theta + c\mathbf{1}) = L(\theta)$ directly from Equation (3.2). Differentiate the identity with respect to c at $c = 0$ to recover Equation (3.7).

Exercise 3.E7 — Convexity distinctions

For each function below, list every applicable property among convexity, strict convexity, and strong convexity on the stated domain:

$$f_1(x) = x^4 \text{ on } \mathbb{R}, \qquad f_2(x, y) = (x - y)^2 \text{ on } \mathbb{R}^2, \qquad f_3(x, y) = (x - y)^2 \text{ on } \{x + y = 0\}.$$

Use the examples to explain how restricting a function to the centered subspace can remove a flat direction.

Exercise 3.E8 — Why a small gradient needs curvature

For $a > 0$, let $f_a(x) = \frac{a}{2}x^2$. If $|f_a'(x)| \le \varepsilon$, derive the best bound on the distance from x to the minimizer. Explain why there is no parameter-error bound independent of the lower-curvature constant a.

Existence and uniqueness of the Bradley–Terry estimator

Chapter 3 defined the centered estimator

$$\widehat{\theta} \in \arg\min_{\theta \in \mathcal{C}} L(\theta), \qquad \mathcal{C} = \{\theta \in \mathbb{R}^n : \mathbf{1}^\top \theta = 0\}.$$

Before trusting the output of an optimization algorithm, we should ask two questions, in this order. Is the minimum attained at a finite point? If it is, is that point unique?

The distinction is easy to miss. The following one-dimensional examples keep it visible.

Function on $\mathbb{R}$	Existence	Uniqueness
$f(t) = t^2$	the minimum is attained at 0	the minimizer is unique
$f(t) = e^t$	the infimum 0 is approached as $t \to -\infty$, but never attained	there is at most one minimizer, but none exists
$f(t) = 0$	every point is a minimizer	the minimizer is not unique

Both t^2 and e^t are strictly convex. Strict convexity therefore gives *at most* one minimizer; it does not make a minimizer appear.

The Bradley–Terry loss has analogues of the last two failures. If the observed comparisons split into disconnected groups, one group can be shifted relative to another without changing the loss: this is a flat direction and an identifiability failure. If all evidence across some cut points one way, an entire group can be pushed upward while the loss keeps decreasing: this is an escaping direction and an existence failure. Two different graphs will make these statements precise.

4.1 Strict convexity and directional curvature

A function f is strictly convex on a convex set if

$$f((1-t)x + ty) < (1-t)f(x) + tf(y)$$

whenever $x \neq y$ and $0 < t < 1$. If two distinct points were both minimizers, strict convexity would make every point strictly between them better, which is impossible.

For a twice-differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, the Hessian is

$$\nabla^2 f(\theta) = \left(\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \right)_{i,j=1}^{n}.$$

To see what this matrix measures, fix a direction x and restrict the function to the line $\theta + tx$:

$$h(t) = f(\theta + tx).$$

The chain rule gives

$$h''(0) = x^\top \nabla^2 f(\theta) x. \tag{4.1}$$

For a unit vector x, this is curvature per unit displacement in direction x. For a general vector it also contains the scale $\|x\|_2^2$.

A symmetric matrix H is positive semidefinite if $x^\top Hx \geq 0$ for every x, and positive definite on a collection of directions if the inequality is strict for every nonzero direction in that collection. A positive-semidefinite Hessian gives convexity. A Hessian that is positive definite on the directions relevant to a convex domain gives strict convexity there. This is not the same as a uniform strong-convexity bound: the latter requires one positive lower bound that works at every point of the domain.

4.2 One compared pair

Fix an unordered pair $\{i, j\}$, and write

$$d_{ij} = \theta_i - \theta_j, \qquad L_{ij}(\theta) = N_{ij} \log(1 + e^{d_{ij}}) - Y_{ij} d_{ij}.$$

As a scalar function of $d = d_{ij}$,

$$\frac{dL_{ij}}{dd} = N_{ij}\sigma(d) - Y_{ij}, \qquad \frac{d^2 L_{ij}}{dd^2} = N_{ij}\sigma(d)(1 - \sigma(d)).$$

Moreover,

$$d_{ij} = (e_i - e_j)^\top \theta, \qquad \nabla d_{ij} = e_i - e_j.$$

Since d_{ij} is linear, the multivariate chain rule yields

$$\nabla^2 L_{ij}(\theta) = N_{ij} p_{ij}(\theta)(1 - p_{ij}(\theta))(e_i - e_j)(e_i - e_j)^\top. \tag{4.2}$$

For example, when $i = 1$, $j = 2$, and $n = 3$, the outer product is

$$(e_1 - e_2)(e_1 - e_2)^\top = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Only the difference between coordinates i and j is affected.

For any direction x,

$$x^\top \nabla^2 L_{ij}(\theta) x = N_{ij} p_{ij}(\theta)(1 - p_{ij}(\theta))\big(x^\top (e_i - e_j)\big)^2 \tag{4.3}$$

$$= N_{ij} p_{ij}(\theta)(1 - p_{ij}(\theta))(x_i - x_j)^2. \tag{4.4}$$

One observed pair therefore supplies curvature only in directions that change the utility difference between its endpoints.

4.3 From compared pairs to the comparison graph

Suppose first that every pair is observed. Any nonconstant direction changes at least one compared difference, so some pair detects it. The all-ones direction remains flat because it changes no difference.

At the other extreme, suppose comparisons occur only within two groups and never across them. Raising every utility in the first group while lowering every utility in the second leaves all observed differences unchanged. A single comparison joining the groups removes that particular freedom. We do not need every possible pair; we need enough pairs to link all items.

Definition 4.3.1 (Comparison graph)**.** The undirected comparison graph is $G = ([n], E)$, where

$$\{i, j\} \in E \quad \Longleftrightarrow \quad N_{ij} > 0.$$

It is connected if every two vertices are joined by an undirected path.

The graph records which utility differences are touched by the data, but not who won. Summing Equation (4.2) over its edges gives

$$\nabla^2 L(\theta) = \sum_{\{i,j\}\in E} a_{ij}(\theta)(e_i - e_j)(e_i - e_j)^\top, \qquad a_{ij}(\theta) = N_{ij}p_{ij}(\theta)(1 - p_{ij}(\theta)). \tag{4.5}$$

At every finite θ, each observed-edge weight is strictly positive. The quadratic form is

$$x^\top \nabla^2 L(\theta) x = \sum_{\{i,j\}\in E} a_{ij}(\theta)(x_i - x_j)^2. \tag{4.6}$$

Every term is nonnegative, which proves convexity of the loss.

4.4 Nullspace, identifiability, and uniqueness

The sum in Equation (4.6) is zero exactly when $x_i = x_j$ on every observed edge. Equality then propagates along paths.

Proposition 4.4.1 (Hessian nullspace)**.** *At every finite θ,*

$$\ker \nabla^2 L(\theta) = \{x \in \mathbb{R}^n : x \text{ is constant on each connected component of } G\}.$$

In particular, if G is connected, then

$$\ker \nabla^2 L(\theta) = \operatorname{span}\{\mathbf{1}\}.$$

The conclusion in the connected case is first that all coordinates of a zero-curvature direction are equal, not that they are all zero. If the direction is also centered, then $0 = \mathbf{1}^\top x = nc$, so the common value c is zero. Thus the Hessian is positive definite on the centered subspace.

Corollary 4.4.2 (Uniqueness conditional on existence)**.** *If the comparison graph is connected, the Bradley–Terry negative log-likelihood is strictly convex on $\mathcal{C}$. Consequently, if a finite centered minimizer exists, it is unique.*

If the graph has k connected components, each component may be shifted independently. Centering removes one global degree of freedom but leaves $k - 1$ unidentified relative shifts.

Connectedness gives pointwise positive curvature in every nonzero centered direction. It does not give one positive lower bound over the entire unbounded centered subspace. Indeed, the logistic factor

$$p(1 - p)$$

is largest at $p = 1/2$ and approaches zero as p approaches zero or one. A nearly deterministic fitted comparison may make the ordering clear while only weakly pinning down the magnitude of an already large utility gap. This is why Chapter 6 imposes a bounded dynamic range before claiming a uniform curvature bound.

4.5 Weighted graph Laplacians

A matrix of the form

$$\mathsf{L}_G(a) = \sum_{\{i,j\}\in E} a_{ij}(e_i - e_j)(e_i - e_j)^\top, \qquad a_{ij} > 0,$$

is a weighted graph Laplacian. Equation (4.5) therefore says that the Bradley–Terry Hessian is a Laplacian of the comparison graph, with parameter-dependent weights.

The graph determines which directions can be constrained; the weights determine how strongly. For a connected graph, the smallest positive Laplacian eigenvalue measures the weakest direction orthogonal to $\mathbf{1}$. A path and a complete graph are both connected and hence both identify centered utilities, but the path has much weaker global directions. This quantitative distinction returns in Chapter 6.

4.6 Why finite existence is a different question

Consider two items compared ten times, with item 1 winning every comparison. Under centering, write $\theta(t) = (t, -t)$. Then

$$L(\theta(t)) = 10 \log(1 + e^{-2t}). \tag{4.7}$$

The loss is positive at every finite t and decreases to zero as $t \to \infty$. It has an infimum but no finite minimizer. The comparison graph is connected and the centered loss is strictly convex, yet existence fails.

The reason is directional evidence. Every result supports item 1 being above item 2, and no reverse outcome resists a larger gap. If both outcomes occur, with $a > 0$ wins by item 1 and $b > 0$ wins by item 2, the pair loss

$$a \log(1 + e^{-d}) + b \log(1 + e^{d})$$

diverges in both tails and has its finite minimum at $d = \log(a/b)$. Wins in both directions on every compared pair are therefore sufficient for finite existence, but they are not necessary. With three items, for example, the win arrows $1 \to 2$, $2 \to 3$, and $3 \to 1$ already prevent every item or group from being raised without encountering contrary evidence.

This phenomenon is the comparison-model analogue of separation in logistic regression. A penalty or prior can produce a finite estimate, but it changes the estimator: a penalized or Bayesian estimate is not the unregularized MLE.

4.7 The directed win graph

To retain the direction of the observed evidence, draw an arrow from each observed winner to the corresponding loser.

Definition 4.7.1 (Win graph)**.** The directed win graph $G^{\mathrm{w}} = ([n], E^{\mathrm{w}})$ contains the arc

$$i \to j \quad \Longleftrightarrow \quad Y_{ij} > 0.$$

It is strongly connected if every ordered pair of vertices is joined by a directed path.

For one observed win $i \succ j$,

$$-\log p_{ij}(\theta) = -\log\left(\frac{1}{1 + e^{\theta_j - \theta_i}}\right) = \log(1 + e^{\theta_j - \theta_i}).$$

There are Y_{ij} such observations. Summing over all ordered win directions gives the equivalent loss representation

$$L(\theta) = \sum_{i \neq j} Y_{ij} \log(1 + e^{\theta_j - \theta_i}). \tag{4.8}$$

For an arc $i \to j$, increasing $\theta_i - \theta_j$ decreases its loss term; placing the observed loser far above the observed winner makes that term large.

Now consider a nonempty proper group S. If no win arrow enters S, then no item outside S has ever defeated an item inside it. Raising every utility in S relative to those outside improves every cross-group outcome and leaves all within-group differences unchanged. Strong connectivity rules out precisely this kind of one-way cut: every nonempty proper set must have an entering and a leaving directed path, and hence at least one win arrow in each direction across the cut.

4.8 Ford's criterion

Theorem 4.8.1 (Ford's criterion for a finite centered MLE)**.** *Let $n \geq 2$, and suppose the undirected comparison graph is connected. The centered Bradley–Terry MLE exists, and is then unique, if and only if the directed win graph is strongly connected.*

Strong connectivity of the win graph already implies connectedness of the comparison graph. The connectedness assumption is displayed separately because it isolates the condition responsible for uniqueness from the directed condition responsible for finite attainment. The criterion is due to Ford (1957).

Proof sketch. Suppose first that the win graph is not strongly connected. Contract each strongly connected component—a maximal group of mutually reachable vertices—to one vertex. The resulting directed acyclic graph, which has no directed cycle, has a source component S, meaning that no win arc enters S. Since the comparison graph is connected, S is a proper subset with at least one comparison across the cut (S, S^c); every realized cross-cut win points from S to S^c.

The direction

$$q_S = \mathbf{1}_S - \frac{|S|}{n}\mathbf{1}$$

where $\mathbf{1}_S$ is the indicator vector of S. This direction is centered and raises all items in S by one unit relative to those outside. Along $\theta + tq_S$, within-group loss terms remain unchanged and every cross-cut term decreases, with at least one decreasing strictly. Thus every finite centered point can be improved, so no finite minimizer exists.

Conversely, suppose the win graph is strongly connected. For a centered θ, write $R(\theta) = \max_i \theta_i - \min_i \theta_i$, choose a minimum-utility vertex u and a maximum-utility vertex v, and let

$$u = i_0 \to i_1 \to \cdots \to i_k = v, \qquad k \le n - 1,$$

be a simple directed path. Its utility increments telescope:

$$\sum_{r=0}^{k-1} (\theta_{i_{r+1}} - \theta_{i_r}) = \theta_v - \theta_u = R(\theta).$$

At least one win arc on the path therefore has

$$\theta_{i_{r+1}} - \theta_{i_r} \ge \frac{R(\theta)}{n-1}.$$

On that arc the observed loser lies above the observed winner. Since $\log(1 + e^x) \ge x$ for $x \ge 0$,

$$L(\theta) \ge \frac{R(\theta)}{n-1}.$$

For centered θ,

$$R(\theta) \ge \|\theta\|_\infty \ge \frac{\|\theta\|_2}{\sqrt{n}}.$$

Hence $L(\theta) \to \infty$ whenever a centered parameter escapes to infinity. This property is called coercivity on $\mathcal{C}$.

The sublevel set at height $L(0)$,

$$\{\theta \in \mathcal{C} : L(\theta) \le L(0)\}$$

is nonempty, closed, and bounded, and is therefore compact. Continuity gives a finite minimizer on this set. Connectedness and Corollary 4.4.2 give uniqueness. The individual graph and compactness steps are developed in Exercise 4.E7. □

4.9 The running dataset and the two graph conditions

In $\mathcal{D}_0$, every pair is compared, so the comparison graph is complete. In addition, every pair contains wins in both directions. The win graph therefore contains both orientations of every comparison edge and is strongly connected. The centered MLE consequently exists and is unique. Bidirectional wins on every edge are more than the theorem requires; a directed cycle can already be strongly connected.

The chapter's two graphs serve different purposes.

Object	What it records	Relevant condition	What the condition explains
comparison graph	which pairs were observed	connected	uniqueness, conditional on finite existence
win graph	which directed outcomes occurred	strongly connected	finite existence

Further reading

The finite-existence criterion is due to Ford (1957). The paper by Shah, Balakrishnan, J. Bradley, et al. (2016), included in the course project reading list, develops topology-dependent estimation bounds for pairwise-comparison models. For computational approaches to the Bradley–Terry MLE, see Newman (2023); for the convex-analytic language used here, see Boyd and Vandenberghe (2004).

4.10 Exercises

Exercise 4.E1 — Directional curvature on small graphs

For a single observed edge $\{1, 2\}$ on three vertices, write the Hessian up to its positive scalar weight and find its nullspace. Repeat for the path 1–2–3 with arbitrary positive edge weights. Explain which additional direction the second edge constrains.

Exercise 4.E2 — Disconnected components

Suppose the comparison graph has $k \geq 2$ connected components. Prove that the loss is unchanged when an arbitrary constant is added within each component. After imposing $\mathbf{1}^\top \theta = 0$, construct $k - 1$ linearly independent centered flat directions and explain why relative component levels remain unidentified.

Exercise 4.E3 — Nullspace of the Hessian

Prove Proposition 4.4.1 in both directions. Deduce positive definiteness on the centered subspace when G is connected. Hint available in Appendix A.

Exercise 4.E4 — One reverse win

Two items are compared $a + b$ times. Item 1 wins a times and item 2 wins b times. For $a, b > 0$, minimize the pair loss as a function of $d = \theta_1 - \theta_2$ and obtain $d = \log(a/b)$. Describe what happens when $a = 0$ or $b = 0$, and relate the answer to the two-vertex win graph.

Exercise 4.E5 — Diagnose the graphs

For each directed graph below, form its underlying undirected comparison graph and decide whether a finite unique centered MLE is guaranteed:

(a) the directed cycle $1 \to 2 \to 3 \to 1$;

(b) two strongly connected groups with at least one cross-group arrow, and with every such arrow pointing from the first group to the second;

(c) a connected undirected graph with both orientations present on every edge.

For every case that fails, state whether the obstruction is a flat direction, an escaping direction, or both.

Exercise 4.E6 — A dominant item

Every pair among A, B, C, D is compared. Item A defeats each of B, C, D every time, while comparisons among B, C, D contain wins in both directions. Starting from an arbitrary centered vector θ, define

$$\theta(t) = \theta + t(3, -1, -1, -1)^\top, \qquad t \geq 0.$$

Show that all contributions involving A strictly decrease with t, while the remaining contributions are unchanged. Conclude that every finite centered point can be improved and hence that no finite centered MLE exists.

Exercise 4.E7 — Complete Ford-criterion proof

Complete the proof of Theorem 4.8.1.

(a) Form the *condensation* by contracting every strongly connected component to a single vertex. Show that the resulting graph is acyclic and, when the original graph is not strongly connected, has a source component that is a nonempty proper subset of the vertices.

(b) Use connectedness of the comparison graph to prove that at least one comparison crosses the corresponding cut. Verify that $q_S = \mathbf{1}_S - (|S|/n)\mathbf{1}$ is centered and determine how every term in Equation (4.8) changes along $\theta + tq_S$.

(c) Under strong connectivity, prove the path-increment bound used in the proof sketch.

(d) Use $\log(1 + e^x) \geq x$ for $x \geq 0$ to establish the range lower bound on L.

(e) Prove that a centered sequence with diverging Euclidean norm has diverging range.

(f) Use continuity and compactness of a sublevel set to prove attainment, and then use strict convexity to prove uniqueness.

Hint available in Appendix A.

Coding challenge 4.C1 — Flat and escaping directions

Plot $L(\theta + tr)$ against t for three self-contained datasets: any dataset under the global shift $r = \mathbf{1}$; a disconnected two-component design under a centered component-shift direction; and the dominant-item data of exercise 4.E6 under $r = (3, -1, -1, -1)^\top$. Use positive and negative values of t. Identify every flat curve and every direction that approaches an unattained infimum.

Rank Centrality

Maximum likelihood estimates utilities by optimizing a statistical objective. Rank Centrality takes a different route. It turns the observed comparisons into a random walk and uses the stationary probability of each item as its score. The method is spectral rather than likelihood-based, although under the population Bradley–Terry model it targets the same positive weights.

Let us begin with the ranking idea, and introduce the required Markov-chain language only after the walk has been constructed.

5.1 Why use a random walk?

Let $G = ([n], E)$ denote the undirected comparison graph, and let $\mathcal{N}(i) = \{j : \{i, j\} \in E\}$ be the neighbors of item i. The graph tells us which items were compared. An ordinary random walk on this undirected graph would move from an item to one of its neighbors, perhaps uniformly. Such a walk sees the comparison schedule but not the outcomes. On the complete comparison graph of $\mathcal{D}_0$, all four vertices have the same degree, so the ordinary walk has a uniform stationary distribution. It cannot distinguish A, which won twenty-four times, from D, which won seven times.

A ranking walk must use who defeated whom. Recall that the win graph contains the arrow $i \to j$ when i defeated j. Following those arrows would move probability from winners toward losers, the opposite of what we want. Rank Centrality reverses them. When the walker is at item i, it tends to move toward item j in proportion to how often j defeated i.

For an observed pair, define the empirical win fraction

$$\widehat{p}_{ij} = \frac{Y_{ij}}{N_{ij}}.$$

The reversed arrow from i to j receives weight $\widehat{p}_{ji}$. In the A, D comparisons of $\mathcal{D}_0$, for example,

$$\widehat{p}_{AD} = 0.9, \qquad \widehat{p}_{DA} = 0.1.$$

The walk should therefore move from D toward A much more readily than from A toward D.

5.2 Turning reversed win weights into probabilities

The outgoing reversed weights at a vertex need not sum to one and may sum to more than one. We therefore scale all of them by a common constant. Let $\deg(i)$ be the degree of vertex i in the undirected comparison graph, let

$$d_{\max} = \max_i \deg(i),$$

and choose $d \geq d_{\max}$. In this chapter, d denotes this common normalizing constant, not a utility difference. Define, for $i \neq j$,

$$\widehat{P}_{ij} = \begin{cases} \dfrac{1}{d}\widehat{p}_{ji}, & \{i, j\} \in E, \\ 0, & \{i, j\} \notin E, \end{cases} \qquad \widehat{P}_{ii} = 1 - \sum_{j \neq i} \widehat{P}_{ij}. \tag{5.1}$$

The off-diagonal row sum satisfies

$$\sum_{j \neq i} \widehat{P}_{ij} = \frac{1}{d} \sum_{j \in \mathcal{N}(i)} \widehat{p}_{ji} \leq \frac{\deg(i)}{d} \leq 1.$$

Thus every diagonal entry is nonnegative, and every row sums to one. The unused probability is placed on the diagonal, where it becomes the probability that the walker stays at its current item. Choosing $d > d_{\max}$ makes every diagonal entry positive.

For the pair A, D with $d = 4$,

$$\widehat{P}_{DA} = \frac{1}{4}\widehat{p}_{AD} = \frac{9}{40}, \qquad \widehat{P}_{AD} = \frac{1}{4}\widehat{p}_{DA} = \frac{1}{40}.$$

The larger transition points from the frequently defeated item toward the frequent winner, as intended.

5.3 The Markov-chain toolkit

A finite, time-homogeneous Markov chain $X_0, X_1, \ldots$ on state space $\mathcal{X} = \{1, \ldots, n\}$ is specified by transition probabilities

$$P_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i).$$

The next state depends on the present state but, conditional on it, not on the earlier path. Each entry is nonnegative and each row sums to one; such a matrix is called row-stochastic.

We represent state distributions as row vectors. If μ_t is the distribution of X_t, then

$$\mu_{t+1} = \mu_t P, \qquad \mu_t = \mu_0 P^t. \tag{5.2}$$

Example 5.3.1 (A three-state chain)**.** Consider

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{pmatrix}$$

with state order (R, S, T). These are generic Markov states, not items from $\mathcal{D}_0$. From R, the chain stays at R or moves to S, each with probability $1/2$. Starting from $\mu_0 = (1, 0, 0)$,

$$\mu_1 = \mu_0 P = (1/2, 1/2, 0).$$

State T cannot be reached in one step but can be reached in two through $R \to S \to T$.

5.3.1 Stationary distributions

Definition 5.3.2 (Stationary distribution)**.** A probability row vector π is stationary for P if

$$\pi P = \pi, \qquad \pi_i \geq 0, \qquad \sum_i \pi_i = 1. \tag{5.3}$$

If $X_0 \sim \pi$, then every later X_t has the same distribution. The state may change, but the probability distribution over states does not. Algebraically, π is a normalized left eigenvector of P with eigenvalue one. Under suitable conditions, π_i is also the long-run fraction of time spent in state i.

5.3.2 Irreducibility, periodicity, and convergence

A chain is *irreducible* if every state can eventually reach every other:

$$\forall i, j, \quad \exists t \geq 0 \text{ such that } (P^t)_{ij} > 0.$$

Equivalently, the directed graph of positive-probability transitions is strongly connected. Every finite irreducible chain has a unique stationary distribution, and all its coordinates are strictly positive.

Irreducibility alone does not guarantee convergence of $\mu_0 P^t$. The chain

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is irreducible and has stationary distribution $(1/2, 1/2)$, but a point mass alternates forever. Formally, the period of state i is

$$\gcd\{t \geq 1 : (P^t)_{ii} > 0\}.$$

In an irreducible chain all states have the same period. The chain is *aperiodic* when this common period is one. A positive self-loop at any state is sufficient for a finite irreducible chain to be aperiodic.

If a finite chain is irreducible and aperiodic, then

$$\mu_0 P^t \longrightarrow \pi \qquad \text{for every initial distribution } \mu_0. \tag{5.4}$$

Irreducibility gives a unique positive stationary distribution; aperiodicity is additionally needed for convergence at successive time steps.

Reducibility does not automatically imply nonuniqueness. For instance,

$$P = \begin{pmatrix} 1 & 0 \\ q & 1-q \end{pmatrix}, \qquad 0 < q \leq 1,$$

has the unique stationary distribution $(1, 0)$, because the first state is the only closed class, meaning the only set of states that the chain cannot leave. What fails is positivity on every state. This distinction matters when stationary probabilities are later converted to log-utilities.

5.3.3 Detailed balance

Definition 5.3.3 (Detailed balance)**.** A probability vector π satisfies detailed balance for P if

$$\pi_i P_{ij} = \pi_j P_{ji} \qquad \text{for every } i, j. \tag{5.5}$$

Detailed balance equates the stationary probability flow in the two directions across every pair. It implies stationarity because, for every j,

$$(\pi P)_j = \sum_i \pi_i P_{ij} = \sum_i \pi_j P_{ji} = \pi_j \sum_i P_{ji} = \pi_j,$$

where the final sum is row j of P. Detailed balance is sufficient for stationarity, but it is not its definition; chains that do not satisfy detailed balance may also have stationary distributions.

5.4 The role of the normalizing constant

The choice of d sets the time scale of the walk. If d and d' are both valid common normalizers, and I denotes the identity matrix, then

$$\widehat{P}(d') = I + \frac{d}{d'}(\widehat{P}(d) - I). \tag{5.6}$$

It follows immediately that

$$\pi \widehat{P}(d) = \pi \quad \Longleftrightarrow \quad \pi \widehat{P}(d') = \pi.$$

Thus the set of stationary distributions, and hence the ranking, is independent of the valid common d. A larger d places more mass on the diagonal and generally makes the walk evolve more slowly. Choosing $d > d_{\max}$ is useful because the resulting self-loops guarantee aperiodicity whenever the chain is irreducible.

5.5 The population oracle

The empirical chain is constructed from noisy win fractions. To understand its target, imagine first that the exact Bradley–Terry probabilities are known. Let

$$w_i^\star = e^{\theta_i^\star} > 0, \qquad p_{ij}^\star = \frac{w_i^\star}{w_i^\star + w_j^\star}.$$

On the same comparison graph, define

$$P_{ij}^\star = \begin{cases} \frac{1}{d} p_{ji}^\star = \frac{1}{d} \frac{w_j^\star}{w_i^\star + w_j^\star}, & i \neq j,\ \{i,j\} \in E, \\ 0, & i \neq j,\ \{i,j\} \notin E, \end{cases}$$

and set $P_{ii}^\star = 1 - \sum_{j \neq i} P_{ij}^\star$.

Theorem 5.5.1 (Population Rank Centrality)**.** *Suppose the undirected comparison graph is connected. Then $P^\star$ has the unique stationary distribution*

$$\pi_i^\star = \frac{w_i^\star}{\sum_k w_k^\star}.$$

Proof. For every observed edge,

$$\pi_i^\star P_{ij}^\star = \frac{w_i^\star}{\sum_k w_k^\star} \frac{1}{d} \frac{w_j^\star}{w_i^\star + w_j^\star} = \pi_j^\star P_{ji}^\star.$$

Detailed balance therefore gives stationarity. On every observed edge, the population transition probability is positive in both directions. A connected comparison graph consequently makes $P^\star$ irreducible, so the stationary distribution is unique. □

This calculation explains why the arrows were reversed: stronger items attract more stationary mass. If the comparison graph is disconnected, the normalized Bradley–Terry weights still satisfy detailed balance and hence form a stationary vector, but they are not uniquely recovered by the walk. Probability mass can be redistributed among disconnected components without creating any flow between them.

It is useful to keep the empirical and population objects visibly separate.

	Empirical	Population or oracle
pairwise quantity	$\widehat{p}_{ij} = Y_{ij}/N_{ij}$	$p_{ij}^\star = w_i^\star/(w_i^\star + w_j^\star)$
transition matrix	$\widehat{P}$, computed from data	$P^\star$, used to identify the target
stationary vector	$\widehat{\pi}$	$\pi^\star \propto w^\star$ when the graph is connected

5.6 The running dataset

For $\mathcal{D}_0$, the maximum degree is three. Choose $d = 4$ and use the item order (A, B, C, D). We already computed the two transitions between A and D. The complete row from A is

$$\left(1 - \frac{3+2+1}{40}, \frac{3}{40}, \frac{2}{40}, \frac{1}{40}\right) = \frac{1}{40}(34, 3, 2, 1),$$

because B, C, D defeated A three, two, and one times. At D, the walker moves toward A, B, C according to their wins over D, giving

$$\frac{1}{40}(9, 8, 6, 17).$$

The first and last rows of the empirical transition matrix are therefore

$$\widehat{P} = \frac{1}{40}\begin{pmatrix} 34 & 3 & 2 & 1 \\ * & * & * & * \\ * & * & * & * \\ 9 & 8 & 6 & 17 \end{pmatrix}. \tag{5.7}$$

The asterisks mark entries intentionally left for Exercise 5.E6.

Starting uniformly,

$$\mu_0 = (1/4, 1/4, 1/4, 1/4), \qquad \mu_1 = \mu_0 \widehat{P} = \frac{1}{160}(58, 44, 34, 24).$$

The first iterate already has the order $A > B > C > D$, but it is not yet the Rank Centrality score. Completing the two missing rows and continuing the walk gives the stationary distribution. That calculation, including an exact answer and the resulting ranking, is left to Exercise 5.E6.

Stationary probabilities estimate positive Bradley–Terry weights, not additive utilities. When every coordinate is positive, a centered utility estimate is obtained by taking logarithms and removing their mean:

$$\widehat{\theta}_i^{\mathrm{RC}} = \log \widehat{\pi}_i - \frac{1}{n}\sum_{k=1}^{n} \log \widehat{\pi}_k. \tag{5.8}$$

Exercise 5.E6 also applies this conversion to $\mathcal{D}_0$.

5.7 Power iteration

The stationary distribution may be found by solving linear equations or by repeated multiplication:

Choose an initial probability vector μ_0 and tolerance ε.
for $t = 0, 1, 2, \ldots$ **do**
  Set $\mu_{t+1} = \mu_t \widehat{P}$.
  if $\|\mu_{t+1} - \mu_t\|_1 \le \varepsilon$ **then**
    stop
  end if
end for

When the chain is irreducible and aperiodic, Equation (5.4) guarantees convergence from every initial distribution. For a large sparse comparison graph, one iteration is a sparse matrix–vector multiplication. This can be attractive computationally, but a poorly mixing chain may require many iterations; there is no universal speed ordering between a spectral method and likelihood optimization.

5.8 Finite data, connectivity, and stability

For fixed items, independent repeated comparisons give

$$\widehat{p}_{ij} \longrightarrow p^\star_{ij}, \qquad \widehat{P} \longrightarrow P^\star.$$

If the population chain is well connected, its stationary distribution is stable under small perturbations, so one expects $\widehat{\pi} \approx \pi^\star$. This is a perturbation argument, not a finite-sample identity. Four features govern its quality:

- more comparisons reduce the noise in the empirical win fractions;
- connectivity allows information to travel across all items;
- faster mixing and better conditioning improve stationary-distribution stability;
- a large utility dynamic range makes one-way empirical outcomes more likely.

The empirical transition graph reverses every arc of the observed win graph:

$$\widehat{P}_{ij} > 0 \quad \Longleftrightarrow \quad Y_{ji} > 0 \qquad (i \neq j).$$

Reversing all arrows preserves strong connectivity. Hence the empirical Rank Centrality chain is irreducible exactly when the win graph is strongly connected. Self-loops can remove periodicity, but they cannot repair a lack of reachability.

At finite sample sizes, maximum likelihood and Rank Centrality need not agree. Their roles are summarized below.

	Bradley–Terry maximum likelihood	Rank Centrality
operation	minimize a negative log-likelihood	construct $\widehat{P}$ and find $\widehat{\pi}$
population target	$\theta^\star$, up to a common shift	$w^\star / \sum_i w_i^\star$
finite sample	likelihood estimator	spectral estimator; generally different from the MLE

If an empirical stationary vector has a zero coordinate, its logarithm is $-\infty$, echoing the separation phenomenon of Chapter 4. One may modify the estimator by smoothing or teleportation—adding a small common restart transition—but that modification must be stated: it is no longer the unaltered Rank Centrality construction.

Further reading

Rank Centrality and its finite-sample analysis were introduced by Negahban et al. (2017), a paper in the course reading list. The stationary-distribution viewpoint also motivates likelihood-oriented spectral iterations; see Maystre and Grossglauser (2015). For finite Markov chains, reversibility, and mixing, see the freely accessible book by Levin et al. (2017). Computational connections to other pairwise-ranking methods are discussed by Newman (2023).

5.9 Exercises

Exercise 5.E1 — Three-state chain: model solution

For the matrix in Example 5.3.1:

(a) verify row-stochasticity;

(b) draw its transition graph and establish irreducibility and aperiodicity;

(c) solve $\pi P = \pi$, $\sum_i \pi_i = 1$;

(d) verify detailed balance.

A model solution appears in Appendix B.

Exercise 5.E2 — Why the ordinary walk ignores outcomes

Let G be a connected undirected graph on $n \geq 2$ vertices. The simple random walk moves from i to each neighbor with probability $1/\deg(i)$. Prove by detailed balance that its stationary distribution is

$$\pi_i = \frac{\deg(i)}{\sum_k \deg(k)}.$$

Explain why this score depends on the comparison schedule but not on any observed winner.

Exercise 5.E3 — Changing the normalizer

Starting from Equation (5.1), prove Equation (5.6). Deduce that every valid common d gives the same stationary distribution. Explain why a very large d may nevertheless be computationally undesirable.

Exercise 5.E4 — Uniqueness without irreducibility

For $P = \left(\begin{smallmatrix} 1 & 0 \\ q & 1-q \end{smallmatrix}\right)$ with $q \in (0, 1]$, solve for every stationary distribution and compute $\mu_0 P^t$ for a general initial distribution $\mu_0 = (\alpha, 1-\alpha)$. State precisely which conclusion supplied by irreducibility fails and which conclusions still hold.

Exercise 5.E5 — Periodicity and averaged iterates: optional

For the two-state alternating chain, show that ordinary iterates need not converge. Show that the Cesàro averages $T^{-1} \sum_{t=0}^{T-1} \mu_0 P^t$ converge to $(1/2, 1/2)$.

Exercise 5.E6 — Complete the Rank Centrality calculation on $\mathcal{D}_0$

Fill the two rows marked by asterisks in Equation (5.7) and verify every row sum. Check the displayed first power iterate. Then solve exactly for the stationary distribution, compute the centered log-utilities in Equation (5.8), and report the resulting ranking. Hint available in Appendix A.

Exercise 5.E7 — Connected design, reducible empirical chain

Three items are compared only on the edges A–B and B–C. In the observed sample, A wins every comparison with B, and B wins every comparison with C. Take $d = 2$, construct the empirical Rank Centrality chain, and determine its stationary behavior. Contrast it with the population chain on the same connected comparison graph under finite Bradley–Terry utilities.

Exercise 5.E8 — Population detailed balance

Repeat the detailed-balance calculation in Theorem 5.5.1 without assuming that the comparison graph is connected. Identify exactly which conclusion holds on every graph and where connectedness is needed.

Coding challenge 5.C1 — Spectral and likelihood estimates

Implement Rank Centrality for $\mathcal{D}_0$ using power iteration and compare the centered log-score with the centered MLE from coding challenge 3.C1. Then simulate increasing numbers of comparisons from a fixed Bradley–Terry parameter. For each sample size, test irreducibility. On reducible trials, identify the closed classes and determine whether any strictly positive stationary distribution exists. Compare centered log-scores only on irreducible trials, unless you explicitly introduce and name a smoothing modification. Plot the discrepancy between the two estimators against comparisons per observed pair.

A finite-sample recovery guarantee

Chapters 3 and 4 constructed the Bradley–Terry maximum-likelihood estimator and determined when the unbounded centered problem has a finite, unique solution. We now ask a statistical question. If repeated datasets are generated from the same true utility vector, their win counts—and hence their fitted utilities—will differ. How close is the estimate from one finite random sample to the truth?

The proof develops a pattern that appears throughout statistical estimation:

$$\boxed{\text{estimation error} \lesssim \frac{\text{random score noise}}{\text{likelihood curvature}}.}$$

Here $\lesssim$ means bounded above up to a numerical constant. The numerator measures how much sampling fluctuations tilt the empirical loss at the truth. The denominator measures how strongly the loss resists movement away from it.

6.1 Model, design, and estimator

Assume that every unordered pair is compared independently m times. For $i < j$ and $r = 1, \dots, m$, let

$$Y_{ij}^{(r)} = \begin{cases} 1, & i \text{ defeats } j, \\ 0, & j \text{ defeats } i, \end{cases} \qquad Y_{ij}^{(r)} \sim \text{Bernoulli}(p_{ij}^\star),$$

where

$$p_{ij}^\star = \sigma(\theta_i^\star - \theta_j^\star).$$

All variables corresponding to distinct comparison outcomes are independent. For $j < i$, define $Y_{ij}^{(r)} = 1 - Y_{ji}^{(r)}$, so the symbol always records whether the first index defeats the second. Set

$$Y_{ij} = \sum_{r=1}^{m} Y_{ij}^{(r)}, \qquad i \neq j.$$

The total number of observed outcomes is

$$N = m\binom{n}{2}. \tag{6.1}$$

For $0 \leq b < \infty$, define the centered, bounded parameter set

$$\Theta_b = \left\{ \theta \in \mathbb{R}^n : \mathbf{1}^\top \theta = 0, \quad \max_i \theta_i - \min_i \theta_i \leq b \right\}. \tag{6.2}$$

Assume $\theta^\star \in \Theta_b$, and define

$$\widehat{\theta} \in \arg\min_{\theta \in \Theta_b} L(\theta). \tag{6.3}$$

Throughout this chapter, $\widehat{\theta}$ denotes this range-constrained estimator; it need not equal the unbounded MLE studied in Chapter 4.

The two constraints have different purposes. Centering removes the common-shift nonidentifiability. The dynamic-range bound prevents fitted probabilities from becoming arbitrarily saturated and supplies uniform curvature throughout the segment joining $\theta^\star$ and $\widehat{\theta}$. It also makes the estimator finite for every

realized sample: even under a complete design, a finite sample can contain one-way outcomes and need not satisfy Ford's condition.

The set Θ_b is convex and compact. Convexity follows because every pairwise difference along a line segment is a convex combination of the endpoint differences. Centering and range at most b imply $|\theta_i| \le b$ for every coordinate, which gives boundedness; the set is also closed. Continuity of L therefore guarantees a minimizer.

For every $\theta \in \Theta_b$, all utility differences lie in $[-b, b]$. Define the minimum logistic curvature on this interval by

$$c_b := \min_{|z| \le b} \sigma(z)(1 - \sigma(z)) = \sigma(b)(1 - \sigma(b)) = \frac{e^b}{(1 + e^b)^2} > 0. \tag{6.4}$$

6.2 The recovery theorem

Theorem 6.2.1 (Finite-sample Bradley–Terry recovery)**.** *Let $n \ge 2$, $m \ge 1$, $0 \le b < \infty$, and $\delta \in (0, 1)$. Under the complete comparison design and assumptions above, with probability at least $1 - \delta$,*

$$\left\| \widehat{\theta} - \theta^\star \right\|_2 \le \frac{2\sqrt{2}}{c_b} \sqrt{\frac{\log(2n/\delta)}{m}}. \tag{6.5}$$

The constant is less important than the structure of the bound.

- **Comparisons per pair.** The error decreases as $m^{-1/2}$. Four times as many comparisons per pair divide the bound by two.
- **Confidence.** A smaller failure probability affects the bound only through $\log(1/\delta)$.
- **Dynamic range.** As b grows, c_b shrinks. Nearly deterministic comparisons reveal the likely sign of a gap more readily than its precise magnitude.
- **Utility versus ranking recovery.** The theorem controls utility error. Exact recovery of a strict ordering additionally requires the true utility gaps to be separated from zero.

The root-mean-square error per item obeys

$$\frac{1}{\sqrt{n}} \left\| \widehat{\theta} - \theta^\star \right\|_2 \le \frac{2\sqrt{2}}{c_b} \sqrt{\frac{\log(2n/\delta)}{mn}} = \frac{2}{c_b} \sqrt{\frac{(n-1)\log(2n/\delta)}{N}}. \tag{6.6}$$

Up to logarithmic and dynamic-range factors, its square has the familiar parametric form n/N: there are $n - 1$ identifiable utility degrees of freedom and N binary observations. The final expression is only a rewriting of the complete-design theorem. It does not prove that an arbitrary sparse design with the same N performs equally well.

6.3 Proof architecture

Write

$$\Delta = \widehat{\theta} - \theta^\star.$$

Both endpoints are centered, so $\mathbf{1}^\top \Delta = 0$. Along the segment joining them, define

$$h(t) = L(\theta^\star + t\Delta), \qquad 0 \le t \le 1.$$

Sampling noise generally makes $h'(0) = \langle \nabla L(\theta^\star), \Delta \rangle$ nonzero, so the empirical minimizer is displaced from the truth. A lower bound on $h''(t)$ limits that displacement. Figure 6.1 illustrates this one-dimensional geometry.

The proof rests on one deterministic lemma and one probabilistic lemma.

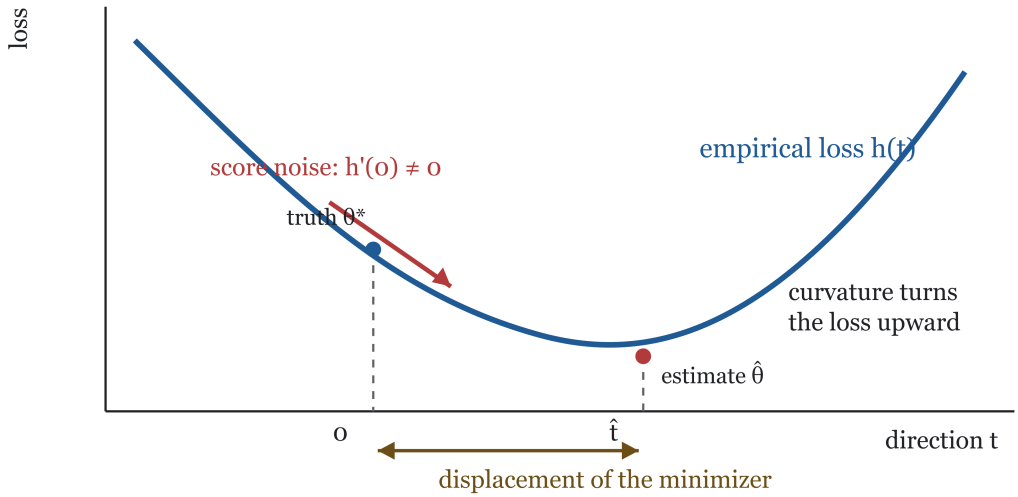

Figure 6.1: Sampling noise produces a nonzero slope at the true parameter, while curvature limits the displacement of the empirical minimizer.

Lemma 6.3.1 (Uniform curvature)**.** *For every $\theta \in \Theta_b$ and every centered direction u,*

$$u^\top \nabla^2 L(\theta) u \geq mnc_b \|u\|_2^2 . \tag{6.7}$$

Lemma 6.3.2 (Concentration of the score)**.** *With probability at least $1-\delta$,*

$$\|\nabla L(\theta^\star)\|_2 \leq \sqrt{2mn(n-1)\log\frac{2n}{\delta}}. \tag{6.8}$$

Lemma 6.3.1 is a deterministic statement about the model, the parameter region, and the comparison design. Lemma 6.3.2 controls the random first-order fluctuation of the empirical loss at the truth.

Proof sketch, assuming the lemmas. Optimality of $\widehat{\theta}$ gives the basic inequality

$$L(\theta^\star + \Delta) \leq L(\theta^\star).$$

The segment between the two points lies in the convex set Θ_b. Taylor's theorem applied to h yields a point $\widetilde{\theta} = \theta^\star + \xi\Delta$, with $\xi \in (0,1)$, such that

$$L(\widehat{\theta}) = L(\theta^\star) + \langle \nabla L(\theta^\star), \Delta \rangle + \frac{1}{2}\Delta^\top \nabla^2 L(\widetilde{\theta})\Delta.$$

Combining the two displays gives

$$\frac{1}{2}\Delta^\top \nabla^2 L(\widetilde{\theta})\Delta \leq -\langle \nabla L(\theta^\star), \Delta \rangle . \tag{6.9}$$

Uniform curvature and Cauchy–Schwarz then imply

$$\frac{1}{2} mnc_b \|\Delta\|_2^2 \leq -\langle \nabla L(\theta^\star), \Delta \rangle \leq \|\nabla L(\theta^\star)\|_2 \|\Delta\|_2 .$$

If $\Delta \neq 0$, cancel one factor and insert the score-concentration bound; the remaining simplification gives Equation (6.5). The case $\Delta = 0$ is immediate. Exercise 6.E5 develops the argument with the exact constant. □

This proof is an instance of a common estimation template. The basic inequality compares the empirical objective at the estimator and the truth; curvature turns displacement into excess loss; concentration controls the random score; and Cauchy–Schwarz combines the two.

6.4 Proof of the curvature lemma

For the complete design, every $N_{ij} = m$. The Hessian quadratic form from Chapter 4 gives, for $\theta \in \Theta_b$,

$$\begin{aligned} u^\top \nabla^2 L(\theta) u &= \sum_{i<j} m p_{ij}(\theta)(1 - p_{ij}(\theta))(u_i - u_j)^2 \\ &\geq m c_b \sum_{i<j} (u_i - u_j)^2. \end{aligned}$$

The complete graph satisfies the identity

$$\sum_{i<j} (u_i - u_j)^2 = n \|u\|_2^2 - (\mathbf{1}^\top u)^2. \tag{6.10}$$

It follows by writing the left side as $\frac{1}{2}\sum_{i,j}(u_i - u_j)^2$ and expanding. For centered u, the final term vanishes, and hence

$$u^\top \nabla^2 L(\theta) u \geq m n c_b \|u\|_2^2 .$$

This proves Lemma 6.3.1. Exercise 6.E3 asks for the expansion without skipped steps.

The bound contains three factors:

$$\underbrace{m}_{\text{repetitions per edge}} \times \underbrace{n}_{\text{complete-graph geometry}} \times \underbrace{c_b}_{\text{minimum logistic curvature}} .$$

The Hessian acts as an information matrix: directions with large curvature are well constrained by the data, while flat or weakly curved directions are poorly estimated.

6.5 A brief concentration primer

At the true parameter, the score $\nabla L(\theta^\star)$ has expectation zero, but its realized value is not generally zero. A concentration inequality bounds the probability that this random fluctuation is large. A sum of M independent, centered, bounded variables typically fluctuates on the scale $\sqrt{M}$; equivalently, its average fluctuates on the scale $M^{-1/2}$.

We use the following elementary form of Hoeffding's inequality.

Theorem 6.5.1 (Hoeffding, bounded symmetric form)**.** *Let $Z_1, \ldots, Z_M$ be independent, mean-zero random variables with $Z_r \in [-1, 1]$. Then for every $t > 0$,*

$$\mathbb{P}\left(\left| \sum_{r=1}^{M} Z_r \right| \geq t \right) \leq 2 \exp\left(-\frac{t^2}{2M} \right). \tag{6.11}$$

To make this probability at most δ, one may set

$$t = \sqrt{2M \log(2/\delta)}.$$

The inequality is distribution-free within its assumptions; it does not require a parametric model for the summands beyond independence, centering, and boundedness.

6.6 Proof sketch of the score-concentration lemma

Let $g = \nabla L(\theta^\star)$. For each item i,

$$g_i = \sum_{j \neq i} (m p_{ij}^\star - Y_{ij}) \tag{6.12}$$

$$= \sum_{j \neq i} \sum_{r=1}^{m} \left(p_{ij}^\star - Y_{ij}^{(r)} \right). \tag{6.13}$$

For fixed i, the $m(n-1)$ summands correspond to distinct comparison outcomes. They are independent, have mean zero, and lie in $[-1,1]$. Hoeffding's inequality therefore gives

$$\mathbb{P}(|g_i| \geq t) \leq 2\exp\left(-\frac{t^2}{2m(n-1)}\right).$$

Choose the coordinatewise failure probability as δ/n, apply a union bound over the n coordinates, and then use $\|g\|_2 \leq \sqrt{n}\,\|g\|_\infty$. This proves Lemma 6.3.2; Exercise 6.E4 develops the exact threshold and constants.

The coordinates g_i are not independent: one comparison affects two coordinates with opposite signs. The union bound does not require independence between the events being combined; it requires only a valid bound for each coordinate separately.

6.7 What the complete design contributes

The complete-design assumption enters the curvature proof through the exact identity Equation (6.10). It is a sufficient and algebraically clean design, not a necessary one.

For a general fixed design, define the count-weighted design Laplacian

$$\mathsf{L}_G = \sum_{i<j} N_{ij}(e_i - e_j)(e_i - e_j)^\top. \tag{6.14}$$

The same Hessian calculation gives, for $\theta \in \Theta_b$,

$$u^\top \nabla^2 L(\theta) u \geq c_b u^\top \mathsf{L}_G u.$$

A disconnected design leaves centered directions unidentified. Among connected designs, the smallest positive eigenvalue of L_G measures the weakest remaining curvature. A path has a slowly varying direction with small edge-difference energy, whereas a complete graph constrains every centered direction strongly. Sparse but well-connected graphs can still be effective; connectivity alone does not make all designs statistically equivalent.

A general-graph recovery theorem must control both this graph-dependent curvature and the corresponding score noise. That additional spectral analysis is beyond the present theorem.

6.8 Lower-bound intuition

Near zero,

$$\sigma(\epsilon) = \frac{1}{2} + \frac{\epsilon}{4} + O(\epsilon^3).$$

Thus distinguishing a zero utility gap from a gap ϵ requires distinguishing Bernoulli probabilities separated by order ϵ. Standard binary-testing intuition says that this needs order $1/\epsilon^2$ independent observations. The resulting resolution is of order $m^{-1/2}$, matching the dependence in the upper bound. Exercise 6.E8 develops the calculation.

6.9 A curvature calculation for the running dataset

The maximum of $p(1-p)$ is $1/4$, attained at $p = 1/2$. In $\mathcal{D}_0$, $\widehat{p}_{AD} = 0.9$. At a parameter whose fitted probability equals this empirical fraction, the local logistic curvature factor is

$$\widehat{p}_{AD}(1-\widehat{p}_{AD}) = 0.09 = 0.36 \times \frac{1}{4}.$$

The pair then supplies only 36% of the local curvature of a balanced pair with the same number of comparisons. This does not mean that the evidence for $A \succ D$ is weak. It means that evidence about the sign of a gap and information about the precise magnitude of an already large gap are different notions.

6.10 From fixed designs to adaptive learning

The comparison design in this chapter is supplied in advance and does not depend on observed outcomes. The goal is to estimate utilities from that fixed dataset. Later modules consider an online learner that chooses comparisons sequentially, using earlier outcomes. The objective may then be identification or cumulative regret—the accumulated loss relative to a benchmark decision—rather than global utility error. The broad noise-versus-curvature idea often survives, but the concentration argument must account for dependence created by adaptive data collection.

Further reading

For concentration inequalities, including Hoeffding's inequality, see Chapter 2 of the freely accessible text by Vershynin (2026); the original bounded-sum inequality is due to Hoeffding (1963). For sharp topology-dependent estimation bounds in pairwise-comparison models, see Shah, Balakrishnan, J. Bradley, et al. (2016). The finite-sample analysis of Rank Centrality is in Negahban et al. (2017). Robust and nonparametric alternatives to the Bradley–Terry model are discussed by Shah and Wainwright (2018) and Shah, Balakrishnan, Guntuboyina, et al. (2017), both in the course project reading list.

6.11 Exercises

Exercise 6.E1 — Geometry of the parameter set

Prove that Θ_b is convex and compact. Show that every point on the line segment between $\theta^\star$ and $\widehat{\theta}$ remains in Θ_b.

Exercise 6.E2 — Why constrain the estimator?

Two items are compared m times, and item 1 wins every comparison. Under centering and the constraint $\max_i \theta_i - \min_i \theta_i \le b$, find the constrained maximum-likelihood estimate. Contrast it with the unbounded centered problem and explain why the constrained estimator is finite for this sample.

Exercise 6.E3 — Complete-graph identity

Prove Equation (6.10) by expanding $\frac{1}{2}\sum_{i,j}(u_i - u_j)^2$. State its value when $\mathbf{1}^\top u = 0$.

Exercise 6.E4 — Hoeffding and a union bound

Starting from Equation (6.13), derive Equation (6.8). Your solution must identify the number, range, mean, and independence structure of the summands, and must state where the union bound and the ℓ_∞-to-ℓ_2 conversion are used. Hint available in Appendix A.

Exercise 6.E5 — Reconstruct the recovery proof

Prove Theorem 6.2.1 without consulting its proof.

(a) Derive Equation (6.9) from optimality and a one-dimensional Taylor expansion.

(b) Establish Equation (6.7) using exercise 6.E3.

(c) Use exercise 6.E4 and Cauchy–Schwarz to obtain the precise constant in Equation (6.5).

(d) Derive the total-sample form Equation (6.6).

Hint available in Appendix A.

Exercise 6.E6 — From utility error to exact ranking

Suppose the true utilities have no ties and let

$$\gamma = \min_{i \neq j} \left| \theta_i^\star - \theta_j^\star \right| > 0.$$

Show that $\left\| \widehat{\theta} - \theta^\star \right\|_\infty < \gamma/2$ is sufficient for the estimated ordering to be exact. Use $\|x\|_\infty \leq \|x\|_2$ and Theorem 6.2.1 to derive a sufficient lower bound on m in terms of c_b, γ, n, and δ.

Exercise 6.E7 — Dynamic range

Show directly that $c_b = e^b/(1+e^b)^2$ decreases for $b \geq 0$. Compute c_1 and c_4. Quantify both the change in the error-bound multiplier $1/c_b$ and the change in the sample size $1/c_b^2$ required to maintain a fixed target error according to the theorem.

Exercise 6.E8 — Lower-bound challenge

Compute $\sigma(0)$, $\sigma'(0)$, and $\sigma''(0)$, and prove

$$\sigma(\epsilon) = \frac{1}{2} + \frac{\epsilon}{4} + O(\epsilon^3).$$

Assume without proof that distinguishing Bernoulli parameters separated by η requires order $1/\eta^2$ samples. Deduce the smallest order of a Bradley–Terry utility gap that can generally be resolved from m comparisons.

Exercise 6.E9 — A quantitative topology comparison

Let

$$v_i = i - \frac{n+1}{2}, \qquad u = \frac{v}{\|v\|_2}.$$

Show that u is centered and has unit norm. For the path $1\text{–}2\text{–}\cdots\text{–}n$, prove

$$\sum_{i=1}^{n-1} (u_{i+1} - u_i)^2 = \frac{12}{n(n+1)}.$$

For the complete graph, use Equation (6.10) to show that the corresponding edge-difference energy is n. Interpret the contrast.

Coding challenge 6.C1 — Empirical recovery rate

Fix a centered $\theta^\star \in \Theta_b$. For several values of m, simulate complete-design Bradley–Terry data and compute the constrained MLE over the same set Θ_b used in the theorem. Repeat each experiment enough times to display uncertainty bars for the mean ℓ_2 error. On logarithmic axes, plot error against m and superimpose an $m^{-1/2}$ reference slope. Repeat with a larger dynamic range, and report any optimization failures rather than silently discarding them.

Hints for selected exercises

Hints are intentionally brief. Return to the exercise and complete the missing steps before consulting another source.

Exercise 2.E4: Luce's theorem

For any set containing i, j, IIA fixes the ratio $p(i \mid S)/p(j \mid S)$. Once every pairwise ratio equals w_i/w_j, write $p(i \mid S) = c_S w_i$ and determine c_S by summing over S.

↩ Return to Exercise 2.E4.

Exercise 4.E3: Hessian nullspace

If the quadratic form is zero, every nonnegative summand in Equation (4.6) is zero. Therefore $x_i = x_j$ on every edge. Propagate this equality along paths. For the reverse direction, start with a vector constant on each connected component.

↩ Return to Exercise 4.E3.

Exercise 4.E7: Finite MLE

For necessity, note that the coefficient of t in $\theta_i(t) - \theta_j(t)$ is one across a cut from S to S^c. For sufficiency, telescope the utilities along a directed path from a minimum to a maximum. On the centered subspace, $\min_i \theta_i \le 0 \le \max_i \theta_i$, so every coordinate is bounded in absolute value by the range.

↩ Return to Exercise 4.E7.

Exercise 5.E6: Rank Centrality matrix

For an off-diagonal entry in row i, use the frequency with which the destination j defeated the current state i, and then divide by $d = 4$. After the off-diagonal entries are filled, compute the diagonal from the row sum. To solve for stationarity exactly, clear the common factor $1/40$ and solve the resulting linear equations together with $\sum_i \pi_i = 1$. Take logarithms only after all stationary coordinates have been verified to be positive.

↩ Return to Exercise 5.E6.

Exercise 6.E4: Score concentration

For fixed i, use $M = m(n-1)$ in Hoeffding's inequality. Choose the coordinatewise failure probability to be δ/n, then solve for t. On the simultaneous event, use $\|g\|_2 \le \sqrt{n}\,\|g\|_\infty$.

↩ Return to Exercise 6.E4.

Exercise 6.E5: Recovery theorem

Apply Taylor's theorem to $h(t) = L(\theta^\star + t\Delta)$. The intermediate point remains in Θ_b. After inserting the curvature lower bound, use $-\langle g, \Delta\rangle \le \|g\|_2 \, \|\Delta\|_2$ and cancel only after treating the case $\Delta = 0$.

↩ Return to Exercise 6.E5.

Selected model solutions

The purpose of these solutions is to illustrate complete mathematical presentation. A good solution states the relevant formula, substitutes the data carefully, justifies each conclusion, and interprets the result.

B.1 Gradient and one centered update

This is a solution to Exercise 3.E3.

The comparison records are

$$A\text{–}B : 3{:}1, \qquad A\text{–}C : 2{:}2, \qquad B\text{–}C : 3{:}1.$$

The total wins are therefore

$$\begin{aligned} W_A &= 3 + 2 = 5, \\ W_B &= 1 + 3 = 4, \\ W_C &= 2 + 1 = 3. \end{aligned}$$

At the origin every utility difference is zero, so $p_{ij}(0) = 1/2$. Each item participates in eight comparisons, and the model therefore predicts four wins for each item.

Using the predicted-minus-observed formula Equation (3.6),

$$\nabla L(0) = \begin{pmatrix} 4 - 5 \\ 4 - 4 \\ 4 - 3 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

The coordinates sum to zero, as required by shift invariance:

$$-1 + 0 + 1 = 0.$$

Starting at $\theta^{(0)} = 0$, one gradient-descent step with step size $\eta > 0$ gives

$$\theta^{(1)} = -\eta \nabla L(0) = \eta(1, 0, -1)^\top.$$

This iterate is centered and orders the items as

$$A \succ B \succ C.$$

Item A won more often than the model predicted at the origin, so its utility increases. Item C won less often than predicted, so its utility decreases. Item B's observed and predicted totals agree at the origin, so its coordinate does not move in the first step.

↩ Return to Exercise 3.E3.

B.2 A three-state Markov chain

This is a solution to Exercise 5.E1.

Consider, in state order (R, S, T),

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{pmatrix}.$$

Every entry is nonnegative, and the row sums are 1, so P is a transition matrix.

The positive transitions include $R \leftrightarrow S$ and $S \leftrightarrow T$. Hence every state can reach every other state, and the chain is irreducible. Each state has a positive self-loop, so the irreducible chain is aperiodic.

Let $\pi = (a, b, c)$. The stationarity equations $\pi P = \pi$ give

$$\begin{aligned} a &= \tfrac{1}{2}a + \tfrac{1}{4}b, \\ b &= \tfrac{1}{2}a + \tfrac{1}{2}b + \tfrac{1}{2}c, \\ c &= \tfrac{1}{4}b + \tfrac{1}{2}c, \end{aligned}$$

together with $a + b + c = 1$. The first and third equations give $b = 2a$ and $b = 2c$, so $a = c$. Normalization gives

$$\pi = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right).$$

Detailed balance holds. Across edge R, S,

$$\pi_R P_{RS} = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{4} = \pi_S P_{SR}.$$

The same calculation holds across S, T, and both directions between R, T have zero flow. Thus detailed balance supplies an independent verification of stationarity.

↩ Return to Exercise 5.E1.

References and further reading

Boyd, Stephen and Lieven Vandenberghe (2004). *Convex Optimization*. Cambridge University Press. URL: https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf.

Bradley, Ralph Allan and Milton E. Terry (1952). "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons". In: *Biometrika* 39.3/4, pp. 324–345. DOI: 10.2307/2334029.

Fan, Zhou (2016). *Maximum Likelihood Estimation: STATS 200 Lecture 13*. URL: https://web.stanford.edu/class/archive/stats/stats200/stats200.1172/Lecture13.pdf.

Ford Lester R., Jr. (1957). "Solution of a Ranking Problem from Binary Comparisons". In: *The American Mathematical Monthly* 64.8, Part 2, pp. 28–33. DOI: 10.2307/2308513.

Hoeffding, Wassily (1963). "Probability Inequalities for Sums of Bounded Random Variables". In: *Journal of the American Statistical Association* 58.301, pp. 13–30. DOI: 10.1080/01621459.1963.10500830.

Levin, David A., Yuval Peres, and Elizabeth L. Wilmer (2017). *Markov Chains and Mixing Times*. 2nd ed. American Mathematical Society. URL: https://pages.uoregon.edu/dlevin/MARKOV/.

Luce, R. Duncan (1959). *Individual Choice Behavior: A Theoretical Analysis*. Wiley.

Maystre, Lucas and Matthias Grossglauser (2015). "Fast and Accurate Inference of Plackett–Luce Models". In: *Advances in Neural Information Processing Systems*. Vol. 28, pp. 172–180. URL: https://proceedings.neurips.cc/paper/2015/hash/2a38a4a9316c49e5a833517c45d31070-Abstract.html.

McFadden, Daniel (2002). "The Path to Discrete-Choice Models". In: *ACCESS* 20, pp. 2–7. URL: https://www.accessmagazine.org/wp-content/uploads/sites/7/2016/07/access20-01-the-path-to-discrete-choice-models.pdf.

Negahban, Sahand, Sewoong Oh, and Devavrat Shah (2017). "Rank Centrality: Ranking from Pairwise Comparisons". In: *Operations Research* 65.1, pp. 266–287. URL: https://arxiv.org/abs/1209.1688.

Newman, M. E. J. (2023). "Efficient Computation of Rankings from Pairwise Comparisons". In: *Journal of Machine Learning Research* 24.238, pp. 1–25. URL: https://jmlr.org/papers/v24/22-1086.html.

Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F. Christiano, Jan Leike, and Ryan Lowe (2022). "Training Language Models to Follow Instructions with Human Feedback". In: *Advances in Neural Information Processing Systems* 35, pp. 27730–27744. URL: https://arxiv.org/abs/2203.02155.

Shah, Nihar B., Sivaraman Balakrishnan, Joseph Bradley, Abhay Parekh, Kannan Ramchandran, and Martin J. Wainwright (2016). "Estimation from Pairwise Comparisons: Sharp Minimax Bounds with Topology Dependence". In: *Journal of Machine Learning Research* 17.58, pp. 1–47. URL: https://jmlr.org/papers/v17/15-189.html.

Shah, Nihar B., Sivaraman Balakrishnan, Adityanand Guntuboyina, and Martin J. Wainwright (2017). "Stochastically Transitive Models for Pairwise Comparisons: Statistical and Computational Issues". In: *IEEE Transactions on Information Theory* 63.2, pp. 934–959. URL: https://arxiv.org/abs/1510.05610.

Shah, Nihar B. and Martin J. Wainwright (2018). "Simple, Robust and Optimal Ranking from Pairwise Comparisons". In: *Journal of Machine Learning Research* 18.199, pp. 1–38. URL: https://jmlr.org/papers/v18/16-206.html.

Train, Kenneth E. (2009). *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge University Press. URL: https://eml.berkeley.edu/books/choice2.html.

Truong, Sang T., Andreas Haupt, and Sanmi Koyejo (2025). *Machine Learning from Human Preferences*. Stanford University. URL: https://mlhp.stanford.edu/ (visited on 08/14/2026).

Vershynin, Roman (2026). *High-Dimensional Probability: An Introduction with Applications in Data Science*. 2nd ed. Cambridge University Press. URL: https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-2.pdf.