BWF Badminton

Bayesian Data Analysis Project

Team: TheSurvivors

Introduction

- Goal: distribution of the outcomes in a badminton tournament
- Approach: Bayesian data analysis on historical data

Tanongsak
SAENSOMBOONSUK [1]

21-14, 21-12

Kantaphon
WANGCHAROEN

20-22, 21-10, 21-4

Cao Cuong
PHAM

21-15, 21-17

Ihsan Maulana
MUSTOFA

23-21, 21-6

Kantaphon
WANGCHAROEN

21-19, 11-4 Retirement

Ihsan Maulana
MUSTOFA

21-17, 21-14

Kantaphon
WANGCHAROEN

21-14, 21-13

Ranking spread

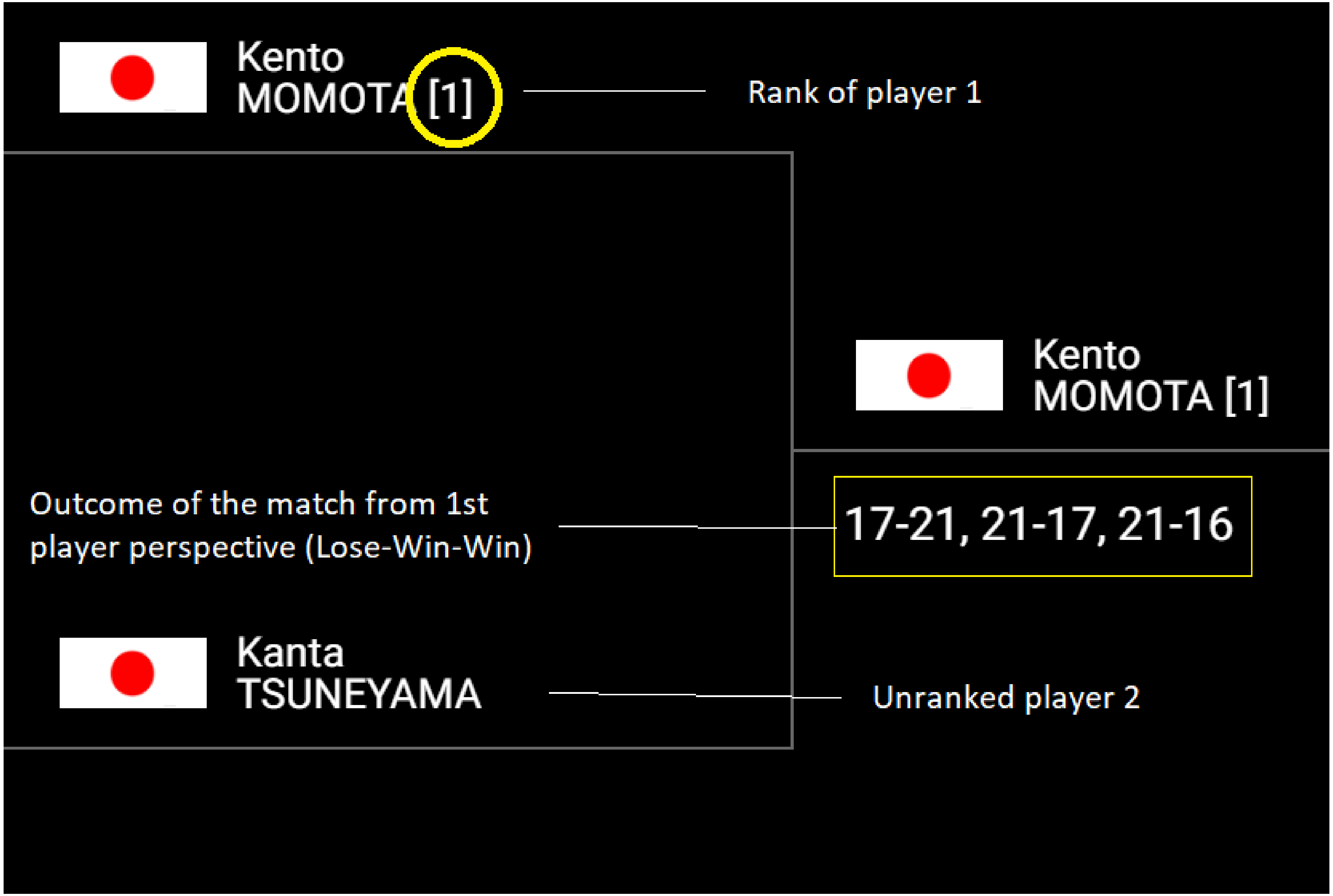
- Ranking spreads = rank(2^{nd} player) – rank(1^{st} player)
- For example:
  - Spread(from 1st rank player to 8th rank player) = 8 - 1 = 7
  - Spread(from 2nd rank player to unrank player) = 12 - 2 = 10

Win degree

- Win degrees = {1,2,3,4,5,6}

  1 Lose Lose → Lose
  2 Lose Win Lose → Lose
  3 Win Lose Lose → Lose
  4 Lose Win Win → Win
  5 Win Lose Win → Win
  6 Win Win → Win

Example of spread 11 and win degree 4



Preprocess

- Objective: 1-d space collection
- Reduction rules:
  - Same spreads, higher win degree correlates to higher value (arrow A)
  - Same win degrees, lower spread correlates to higher value (arrow B)
- How:
  - Starts with an extreme value
  - Increase by step 1

A

B

	6	5	4	3	2	1
-11	28	27	26	25	24	23
-10	27	26	25	24	23	22
-9	26	25	24	23	22	21
-8	25	24	23	22	21	20
⋮	·	·	·	·	·	·
⋮	·	·	·	·	·	·
8	9	8	7	6	5	4
9	8	7	6	5	4	3
10	7	6	5	4	3	2
11	6	5	4	3	2	1

Dataset

- Outcome of the dataset after preprocessing (67x5)

	Tournament 1	Tournament 2	Tournament 3	Tournament 4	Tournament 5
Match 1	6.0	17.0	16.0	6.0	1.0
Match 2	15.0	15.0	17.0	13.0	17.0
Match 3	16.0	7.0	4.0	17.0	5.0
Match 4	15.0	13.0	17.0	17.0	13.0
Match 5	19.0	20.0	20.0	15.0	17.0

Models

- Priors
  - Uniform (weak prior)
  - Inverse Gamma (on variance, conjugate prior to the normal likelihood)

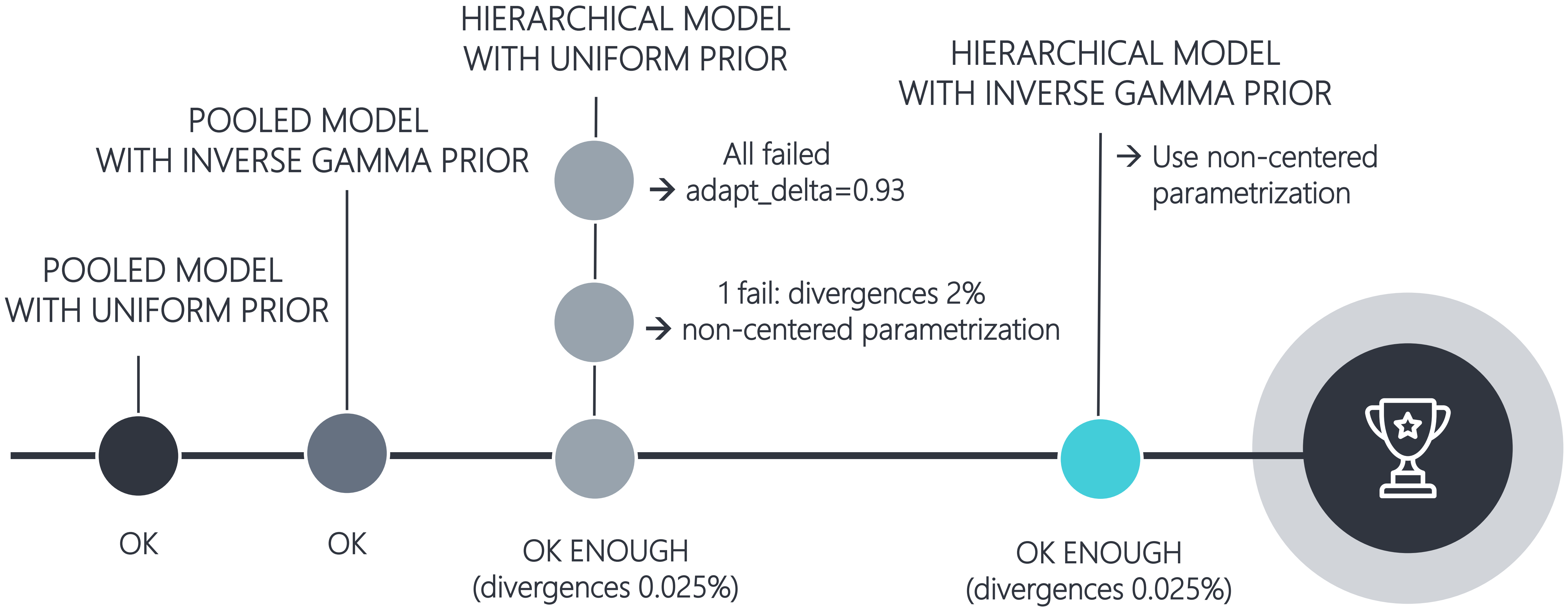
- Likelihood
  - Normal

- Models
  - Pooled with uniform prior
  - Pooled with inverse gamma prior
  - Hierarchical with uniform prior
  - Hierarchical with inverse gamma prior

Convergence diagnostics 1/2

- Pre-conditions
  - Stan's default parameters
  - adapt_delta=0.9
- Validation criteria
  - Rhat < 1.1
  - Effective sample size high
  - Divergences 0

Convergence diagnostics 2/2



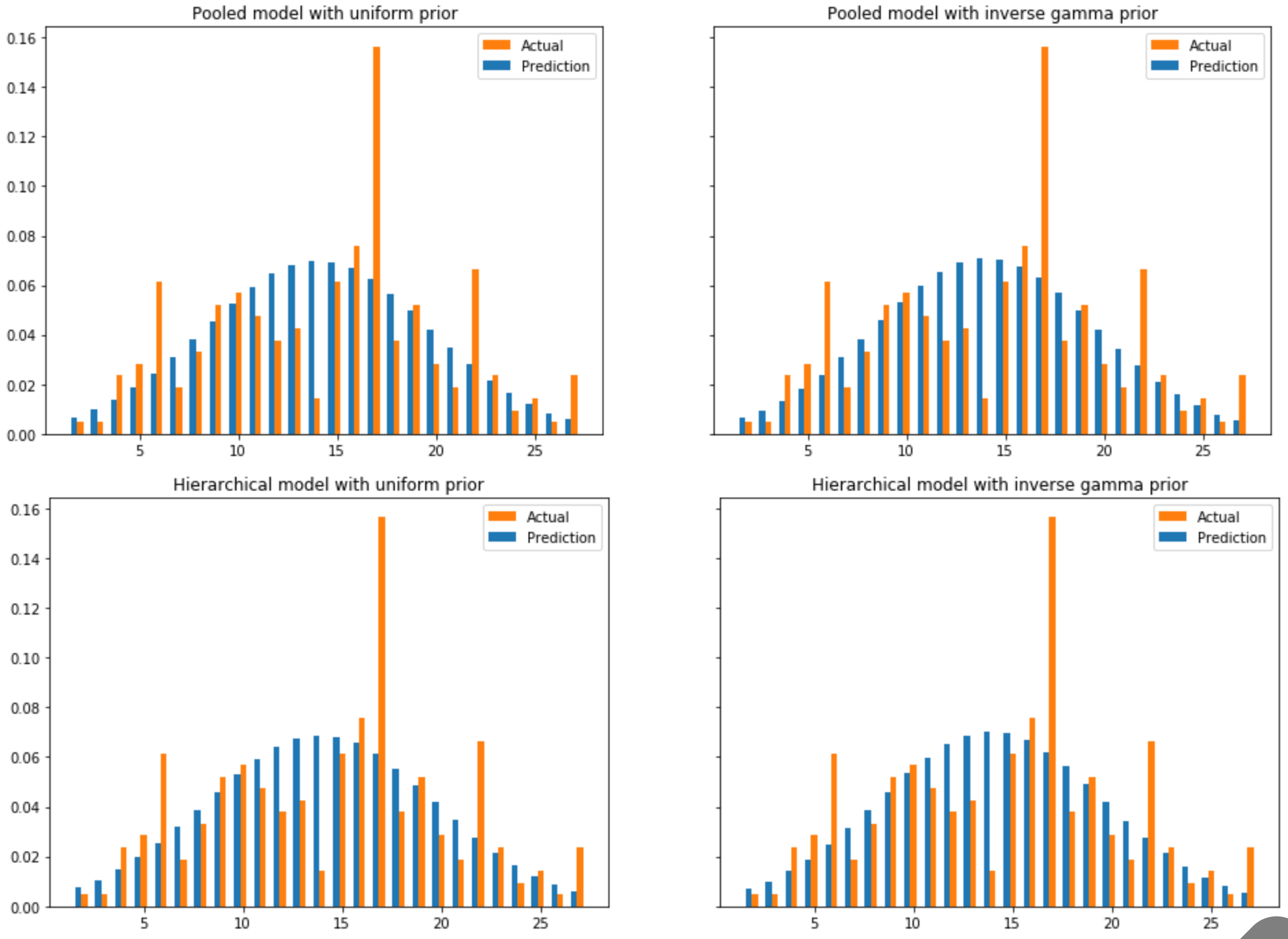
Model comparison with PSIS-LOO

- All the models are reliable (very low k-values)
- Model with best predictive accuracy is Pooled model with inverse gamma prior (highest PSIS-LOO value)

	Models	Psisloo	P_eff	Max k value	Min k value	Mean k value
0	Pooled model with uniform prior	-1056.45	1.67	-0.07	-0.25	-0.18
1	Pooled model with inverse gamma prior	-1056.38	1.64	-0.01	-0.16	-0.11
2	Hierarchical model with uniform prior	-1057.25	2.96	0.10	-0.14	-0.05
3	Hierarchical model with inverse gamma prior	-1057.39	3.10	0.11	-0.18	-0.04

Posterior predictive checking

- Similar trend
- Errors are considerable why?



Conclusions 1/2

- Problems
  - Direct inference of a single match
  - Divergences in hierarchical model
- Improvements
  - Alternative models: Binomial, Multinomial
  - Joint distribution of some parameters (absolute ranking + win degree)
  - Sensitivity analysis for the prior and model

Conclusions 2/2

- Badminton domain perspective:
  - Visible correlation between spread and win degree
  - Extreme outcome (towards 1 or 28) not expected
- Statistical inference perspective
  - Given the domain knowledge, one would expect the distribution of the estimand to be a normal distribution
  - Highly data-driven
  - Hierarchical model ends up as pooled model

THANK YOU