Data analytics, Review for Adam Witczak and Jin Hung Toh team.
Reviewed by: Szymon Lis, Filip Olszowski.
20.06.2022 r.

Final results: **23/24**

Problem formulation [0-4 pts]:

- is the problem clearly stated
  1 pt - the probability of an NBA team winning the tournament given the team leader's statistics

- what is the point of creating model, are potential use cases defined
  1 pt - what is the chance for win golden state warriors in final

- where do data comes from, what does it contain
  1 pt - https://www.kaggle.com/datasets/mujinjo/stephen-curry-stats-20092021-in-nba

- is preprocessing step clearly described
  1 pt - adding the column true\false in order to score the result of the match easily

Model [0-4 pts]

- are two different models specified
  1 pt - yes, all parameters are described for both models and they are also justified

- are difference between two models explained
  1 pt - for model 1 the theta values are 550 and 450, for the second model the theta value depends on the alpha and beta coefficients which are from the beta distribution, the values for each model are listed and described

- is the difference in the models justified (e.g. does adding additional parameter makes sense?)
  1 pt - yes, different parameters and different values for each model

- are models sufficiently described (what are formulas, what are parameters, what data are required) [1 pt]
  1 pt - yes

Priors [0-4 pts]

- Is it explained why particular priors for parameters were selected
  1 pt - yes, the leader played 13 seasons (82 games in each) and the number of career wins was assumed based on that

- Have prior predictive checks been done for parameters (are parameters simulated from priors make sense)
  1 pt - the parameter theta that is simulated from the prior they selected makes sense as they guessed the team won 550 out 1000 which is 55%

- Have prior predictive checks been done for measurements (are measurements simulated from priors make sense)
  1 pt - the graphs reflect assumptions

- How prior parameters were selected
  1 pt - yes, the leader played 13 seasons (82 games in each) and the number of career wins was assumed based on that

Posterior analysis (model 1) [0-4 pts]

- were there any issues with the sampling? if there were what kind of ideas for mitigation were used
  1 pt - no issues with the sampling

- are the samples from posterior predictive distribution analyzed
  1 pt - yes, they have the histograms as justification

- are the data consistent with posterior predictive samples and is it sufficiently commented (if they are not then is the justification provided)
  1 pt - data is consistent with posterior predictive samples (graphs) and it is also commented (Comments on Posterior Analysis on Model 1)

- have parameter marginal distributions been analyzed (histograms of individual parameters plus summaries, are they diffuse or concentrated, what can we say about values)
  1 pt - about values: mean value of y is 0.69 therefore, they can say that the chance of lider winning this championship is 69%, the histogram for real stats is narrower

Posterior analysis (model 2) [0-4 pts]

- were there any issues with the sampling? if there were what kind of ideas for mitigation were used
  1 pt - no issues with sampling

- are the samples from posterior predictive distribution analyzed
  1 pt - yes, they have the histograms as justification

- are the data consistent with posterior predictive samples and is it sufficiently commented (if they are not then is the justification provided)
  1 pt - data is consistent with posterior predictive samples (graphs) and it is also commented (Comments on Posterior Analysis on Model 2)

- have parameter marginal distributions been analyzed (histograms of individual parameters plus summaries, are they diffuse or concentrated, what can we say about values)
  1 pt - about values: mean value of y is 0.8 therefore, they can say that the chance of lider winning this championship is 80%, the histogram for real stats is narrower but it is more similar to prior than in the first model

Model comparison [0-4 pts]

- Have models been compared using information criteria
  1 pt - yes, there is visible difference between models using above criteria

- Have result for WAIC been discussed (is there a clear winner, or is there an overlap, were there any warnings)
  1 pt - clear winner for WAIC is model 2, any warnings, slight overlaps

- Have result for PSIS-LOO been discussed (is there a clear winner, or is there an overlap, were there any warnings)
  1 pt - clear winner for LOO is model 2, any warnings, slight overlaps

- Whas the model comparison discussed? Do authors agree with information criteria? Why in your opinion one model better than another
  0 pt - Comparison is discussed in section Comments on both models but they do not highlight what is the exactly reason why the model 2 is better than 1