Journal of Visualization and Interaction

JoVI review snapshot created at Sep 15, 2026 15:51.

Issue #11 (open): [REVIEW] Recommendations for visual predictive checks in Bayesian workflow

@helske on Jun 19, 2025 08:05:

[opened]

@helske on Jun 19, 2025 08:05:

Conflicts of interest

- [x] I declare that I have no known conflicts of interest with the authors.

Reviewed version

844162b

Review

The paper considers important issue of how to to validate your Bayesian model in terms of its predictive performance. It considers various types of visualizations for (scalar) data, i.e., continuous, count, binary, categorical, and ordinal, and compares the strengths and weaknesses of different visualization types such as kernel density plots and histograms.

The paper is generally well written and easy to follow, and provides good general recommendations which are backed up by research and partly by the personal experiences of the authors. Given the overall tutorial style and aim of the paper, there are few points which could be written in more detail to help readers less experienced with posterior predictive checks and visualizations.

Main comments in no particular order:

Temporal/hierarchical data

It might be a good idea to explicitly state that the paper focuses on visualizations of predictive quantities in case of independent/scalar data, in a sense that there is no discussion on what to do if you have for example time series, panel, or other hierarchical data, where the focus might be on how well the individual posterior predictive trajectories match the original data, or whether the multilevel structure of the data and model should be somehow accounted for. I'm not suggesting that paper should expanded to consider these in detail, but a short note regarding these kind of things could be added to the introduction.

Effect of sample size

There is pretty much no discussion on the effect of the sample size for the different visualizations. When the focus is on visualizing posterior samples, we typically have at least few thousand posterior samples to visualize, and if the samples are say from Gibbs or random-walk Metropolis, even hundreds of thousands or so. On the other hand, the actual observed data sample can be small. Paper could have a short comment on whether some of the methods considered here are more sensitive to these sample sizes.

Description of quantile dot plot

As stated in the paper, quantile dot plots are less known visualization method compared to KDEs and histograms. Therefore a bit more details on it could be helpful, such as: * The intepretation of quantile dot plot could be expanded, for example how crucial is the the number of chosen dots vs. bins in histogram? * In Section 2 when describing the quantile dot plot, it is immediately contrasted with KDE and histograms, ("Compared to KDEs, quantile dot plots have the added benefit..."). This sentence would be better suited in the next subsection where advantages and disadvantages of different methods are listed. * A potential disadvantage of quantile dot plots which is not explicitly mentioned is that being a more recent approach, it is not necessarily available in "all" relevant computing environment (at least not without some external packages/libraries). * It could be just my personal view, but I find the overlaid quantile dot plot of Figure 15c relatively useless, or at least worst of the three choices, as the shape of the observed data dot plot is not well visible. Did you consider drawing 15c so that instead of drawing the posterior samples on top of the quantile dot plot of the observations, you would draw the top dots of the this latter quantile dot plot on top of the posterior samples, akin to the way KDE plot is drawn? Related to this, while trivial to most at least after some trial and error, paper could mention that typically graphical objects corresponding to the observed data is drawn on top of the posterior samples.

PAV-adjusted calibration plot

This is probably the least well known visualization style used in the paper, and it would benefit from more detailed description. Apparently the red line in the figures is the CEP from posterior means, but this is not explicitly stated anywhere. The under/overconfidence statements related to these are also hard to understand given the details, as if CEP is the P(observation | predicted probability of observation = x), then for example looking at the PAV plot of B vs others in Figure 26, given the predicted probability of over 0.76 or so, the CEP is 1, so isn't the model then underconfident and not overconfident as it gives lower probabilities for those certain cases?

Axis labels

Many of the figures have plots which do not have axis labels. While labels such as "Density" might feel unimportant for those frequently drawing these kind of figures, axis labels are important in publication ready visualizations. For example, for those not familiar with rootograms, figure 19b might be confusing as y-axis is labelled as "Count" but x-axis has no label, and in this case the x-axis is actually representing the observed and expected counts, and y-axis is the frequency of the counts.

Minor issues and comments

- In the pdf version, under the equation 4, there is extra Â§ symbol.
- Shouldn't the equation 2 use $\hat f$ instead of f as it is based on the estimated histogram?
- `ggdist` package is written in all capital letters in caption of Figure 1.
- In KDE plot of Figure 17, there is no y axis.
- In Figure 19b, there are two intervals drawn, but there it is not mentioned what these intervals are.
- Also regarding Figure 19b, any reason you preferred to use predicted mean and not predicted median, which could be more natural in this count context? Just something to think about.
- At the start of Section 4, "Assessing the calibration of binary predictions is a common task made difficult by the discreteness of the observed outcome." Discreteness was already an issue in case of count data, perhaps "by the dichotomous nature of observed outcome" would be more precise here?
- Reference to Figure 20a in Section 4.1 should be Figure 21, and Figure 20b in Section 4.2 should be Figure 22, and Figure 20c in Section 4.3 shoujld refer to Figure 23? Is so, then Figure 20 is actually not referred to at all, and could be perhaps removed, or a short description of that figure at the start of Section 4 should be added.
- Regarding the uselessness of the bar graphs in case of binary data, one additional reason to not favor them is that the visualization arguably doesn't bring any benefits here, as same information can be represented more accurately and compactly as a text.
- Figure 24, be explicit on the intervals, are they 95% intervals?
- In the pdf version of the paper, some of the figures are quite far from the corresponding text, so you need to go back and forth multiple pages to.
- The Stan Development Team is in the reference list in form "Team, Stan Development."
- In the Discussion, a short note regarding the availability of software for the recommended

visualizations could be helpful, i.e. instead of copying and modifying the codes of the paper, are there already (or in planning) functionality for example for the suggested PAV-plots in some R packages?

Openness/Transparency

Both the paper and the corresponding visualization and model codes seem to be reproducible with R, although I did not go through all the codes. I did find one minor issue though: In `02_05_predictive_checking.qmd`, when drawing the overlaid quantile dot plot, the `linewidth=0` does not seem to remove the borders of the circles but using `color=NA` does.

For the main README in the repo could list all the R packages needed for running the repository codes.

Submission categories

- ☐ Registered Report
- ☐ Replication Study
- ☑ Empirical Research - Quantitative
- ☐ Empirical Research - Qualitative
- ☐ Systems or design research
- ☐ Commentary
- ☐ Systematic Literature Review

Suggested outcome

Minor revisions: this paper requires some smaller changes, after which I am confident I would be able to endorse it.

Requested changes

I'd mainly like to see bit more expanded description of the quantile dot plot and PAV-adjusted calibration plots, while other issues I've raised are relatively minor points to polish the text, figures, and the message of the paper.

ORCID

https://orcid.org/0000-0001-7130-793X

Edit: small typos.

@Fumeng-Yang on Jun 29, 2025 14:41: Thank you for providing your review!
