*We thank all the three reviewers for their constructive feedback. Please find our answers to major questions raised. Other points will be dealt with in the revised version. Code will be made available by the camera-ready deadline.*

Intuitions and flexibility of the proposed approach (Reviewer #1 and #2): We agree that we should have provided more intuition for the newly introduced density in the main document. The density builds up on Beta distributions as they are the marginals of the Dirichlet distributed random variable $U \sim \text{Dir}(\alpha)$. We then multiply $U$ with an independent random variable $G \sim \text{Beta}(a, b)$. The resulting random variable $Y = UG$ follows a Beta-Liouville distribution, which allows to account for negative dependence, inherited from the Dirichlet distribution through a Beta stick-breaking construction, as well as positive dependence via a common Beta-factor. Our contribution is now how to transform this distribution that lives within the simplex to one that has support on the full hypercube, while also allowing for efficient sampling and log-density evaluations. This discussion will be added to the main document to improve its clarity. We want to add that it is not obvious that eq. (6) is a density, but this follows from the proof of Proposition 1 by taking $f = 1$ in eq. (9) therein. Fig. 1 and 2 shows that our proposed family is quite flexible and expressive while being cheap to sample from. However, we will also try to include a more detailed numerical study of the proposed family of densities. Eventually, the flexibility of the variational approximation can be increased using different complementary work. We have illustrated in Section 6.2 that one can use a mixture of copula-like densities, that can also enhance the flexibility of the marginal distribution. Similarly, one could use the new density within a semi-implicit variational framework whose parameters are the output of a neural network conditional on some latent mixing variable.

Predefined budget of dependency parameters (Reviewer #2): Thank you for pointing out this very interesting question. We think that this point is relevant but would require a lot of ideas and results which are out the scope of the document and would distract the reader from the main ideas suggested in the paper.

Clarification of the transformation $\mathscr{H}$ (Reviewer #2 and #3): The main reason to consider $U = \mathscr{H}(V)$ with $V \sim c_\theta$ was 1) numerical stability since we need to compute quantile functions only on the interval $[\epsilon, 1-\epsilon]$ using this transformation 2) to increase the flexibility of our proposed family. We suggest to take initially at random $\delta \in [0,1]^d$ for the transformation $\mathscr{H}$ such that $\mathbb{P}(\delta_i = \epsilon) = p$ and $\mathbb{P}(\delta_i = 1-\epsilon) = 1-p$ with $p, \epsilon \in (0,1)$ (in our experiments $\epsilon = 0.01$ and $p = 1/2$). We found that choosing a different (large enough) value of $\epsilon$ tends to yield no large difference, as this choice will get balanced by a different value of the standard deviation of the Gaussian marginal transformation. However, we observed that the parameter $p$ can impact the representative power of the variational distribution and the best and most sensible choice was $p = 1/2$ since it leads to a balanced proportion of components of $\delta$ equal to $\epsilon$ and $1-\epsilon$. A more detailed discussion will be added to the document on this point.

Ablation studies and different normalizing flows (Reviewer #1 and #3): We remark that using a Gaussian mean-field distribution with rotations basically yields still a Gaussian approximation, assuming that the small effect of the transformation $\mathscr{H}$ can be neglected. For $X = \mathcal{O}X'$ with a mean-field distribution $X' \sim \mathcal{N}(\mu, \Lambda)$ and some rotation matrix $\mathcal{O}$, we have $X \sim \mathcal{N}(\mathcal{O}\mu, \mathcal{O}\Lambda\mathcal{O}^\top)$. We have shown in sections 6.1 and 6.2 that using a copula-like density instead of an independent-copula base distribution can be beneficial, as the latter corresponds to the results for the full-rank Gaussian case with one rotation in dimension $d = 2$. Following the reviews, we have also performed an analogous study for the BNNs with the basic message being that the transformation $\mathscr{H}$ is essential for the copula-like density, whereas application of additional rotations yields only a smaller improvement. Rotating a mean-field Gaussian also performed less competitively. We will include more details in the revised version and for brevity refer to the MNIST results in the next section. Replacing the rotations with non-linear transformations like Inverse Autoregressive Flows would be an interesting idea to explore further, but we did not have the chance to explore this before the deadline.

High-dimensional BNN and the rotation trick (Reviewer #1 and #3): The variational density for the BNNs includes the rotations. We have not stated this explicitly and this might have caused confusion as we also referred to the density without the rotations as copula-like in the ablation study for the logistic regression. We will make the appropriate correction in the revised document. For the considered BNN with 200k latent variables, the complexity increases by a factor of around 13 compared to a mean-field model. We expect that more ad-hoc tricks can be used to adjust the rotations to some computational budget. For instance, one could consider the series of sparse rotations $\mathcal{O}_1, \cdots, \mathcal{O}_k$, but with $2^k < d$, thereby allowing for rotations of the more adjacent latent variables only. We acknowledge that the format and size Figure 3 is not suitable. We plan to illustrate our result using only one figure presenting $\text{Cor}(\text{vec}(W^3_{\cdot,0}, W^3_{\cdot,1}))$ (corresponding to Fig. 3(a)) with better colors and with not all the neurons. We will also present our results by making bigger and more designed plots in the supplementary material. The motivation for the model in the last experiment was to illustrate that the proposed approach can be used in high-dimensional structured Bayesian models without having to specify more model-specific dependency assumptions in the variatonal family. Our experimental MNIST results show that for this given model, a copula-like approach can perform better (error rate 1.70% with rotations and 1.78% without) than a Gaussian family (3.4% with rotations but without $\mathscr{H}$; 3.82% in the mean-field; a low-rank covariance structure did not converge in our implementation). The proposed density seems a useful alternative to a Gaussian one for the considered model and we have not analysed if it compares favourably in other models mentioned in Table 4, which has been included to indicate that the prediction errors are roughly in line with current work for fully connected networks.