

1 Reviewer 1:

2
3 Thank you. We understand the comments re wordiness and will overhaul the manuscript to focus on precision. Given
4 that there was low confidence in the review and no substantive strengths or weaknesses noted, other than wordiness, we
5 will just give quick answers to your specific questions:
6

7 G indicates a parameter-indexed family of deterministic functions, so sampling from $q_\theta(z)$ for a choice of G is trivial,
8 so long as sampling the base random variable p_0 is trivial. q_0 is a typo, which should read p_0 . ϕ are the parameters of
9 the parameter network. Yes, F is an arbitrary mapping from η to θ approximated by a neural network. Line 130 contains
10 a typo, it should read: $q_\phi(z; \eta)$ rather than $q_{f_\phi}(z; \eta)$. EFNs learn exponential family models, where distributions
11 are indexed by natural parameters $\eta \in H$. A distribution p_η must be chosen in order to stochastically optimize the
12 EFN. Specifying a prior distribution on η , is a tool for prioritizing the approximation of particular exponential family
13 distributions of the model class.
14
15

16 Reviewer 2:

17
18 Thank you for the thorough review. We agree on all your points that “the paper is well written, method well motivated...
19 [but] the experimental section needs more work”. We have spent the past two months, and will continue to spend the
20 next month, fleshing out the experimental section, which we agree was inadequate as it was (one small correction:
21 Figure 5 was not synthetic data; it was neural data gathered from experiments. Nonetheless your point stands that there
22 was too much synthetic data). In brief, some updated experiments so far, and some intended experiments to come:

23 We will do analyses, which provide a satisfying understanding of why density networks (NF) do not scale for many of
24 our tested exponential family distributions, and how the parameter network input differentially affects performance
25 when learning single distributions. We will demonstrate the bias-variance tradeoff of NFs and EFNs: EFNs act as
26 regularizers on model space. We have taken our analyses of neural data a step further to demonstrate the computational
27 savings of using an EFN to learn a log-gaussian Poisson model of spiking responses to drift grating stimuli.
28
29

30 Reviewer 3:

31
32 Respectfully, the reviewer seems to have misunderstood the goal of the paper and the method. The reviewer writes “The
33 paper proposes a variational inference method for exponential family (EF) models”, but this is not correct. The paper
34 attempts to learn a *model*, not a distribution from within that model (which is the problem of variational inference, as
35 in VB, SVAE, etc). As stated in the abstract, “In most all of these settings [VB, SVAE, VI], the goal has been to find a
36 particular member of that model family... Much less attention, however, has been paid to the problem of learning a
37 model itself.” We will clarify this point further in the manuscript as it is indeed the central point of the work, but we
38 request that this review be reconsidered.