PoissonVAE: combining Bayesian Inference with Predictive Coding results in Amortized Sparse Coding

Hadi Vafaii (vafaii@berkeley.edu)
Redwood Center for Theoretical Neuroscience
Herbert Wertheim School of Optometry & Vision Science
Berkeley, CA 94720, USA

Jacob L. Yates (yates@berkeley.edu)
Herbert Wertheim School of Optometry & Vision Science
Helen Wills Neuroscience Institute
Berkeley, CA 94720, USA

Abstract

Variational autoencoders (VAE) employ Bayesian inference to interpret sensory inputs, mirroring processes that occur in primate vision across both ventral (Higgins et al., 2021) and dorsal (Vafaii, Yates, & Butts, 2023) pathways. Despite their success, traditional VAEs rely on continuous latent variables, which significantly deviates from the discrete nature of biological neurons.

Here, we developed the Poisson VAE ($\mathcal{P}$-VAE), a novel architecture that combines principles of predictive coding with a VAE that encodes inputs into discrete spike counts. Combining Poisson-distributed latent variables with predictive coding introduces a metabolic cost term in the model loss function, suggesting a relationship with sparse coding. We explored this connection, training a $\mathcal{P}$-VAE with a linear decoder and an overcomplete latent space on natural image patches, contrasting it with a traditional Gaussian VAE. Unlike the Gaussian VAE, which learned features similar to principal component analysis, $\mathcal{P}$-VAE exhibited Gabor-like feature selectivity, reminiscent of sparse coding patterns. Notably, $\mathcal{P}$-VAE with a linear decoder effectively implements "Amortized Sparse Coding," where inference over neural activations is achieved through the VAE encoder.

Our work provides an interpretable computational framework to study brain-like sensory processing and paves the way for a deeper understanding of perception as an inferential process.

Keywords: Bayesian Inference; Predictive Coding; Sparse Coding; Analysis by Synthesis; Variational Autoencoder

Introduction

Brains have access to noisy, incomplete sensory data, which necessitates an active process to infer the underlying causes of sensory information—a concept similarly reflected in probabilistic generative models like variational autoencoders (VAE) (Kingma & Welling, 2014; Rezende, Mohamed, & Wierstra, 2014). Consequently, VAEs have emerged as promising computational models of visual perception (Higgins et al., 2021; Vafaii et al., 2023; Storrs, Anderson, & Fleming, 2021; Csikor, Meszéna, & Orbán, 2023). However, in contrast to the discrete nature of spiking biological neurons, VAEs are typically parameterized using continuous, Gaussian-distributed latent variables, significantly limiting their biological realism and interpretability.

We address this gap by introducing the Poisson VAE ($\mathcal{P}$-VAE; Fig. 1a), a generative model with a discrete, Poisson-distributed latent space. $\mathcal{P}$-VAE brings together key concepts in neuroscience, such as rate coding and predictive coding, and links them to modern machine learning. Combining Poisson-distributed latents with predictive coding leads to the emergence of a metabolic cost term in the model loss function. This property reveals an unintentional but welcome connection to sparse coding (Olshausen & Field, 1996), which we verify empirically by training a linear $\mathcal{P}$-VAE (Fig. 1b) on natural image patches. In the remainder of the paper, we cover necessary background information, develop the $\mathcal{P}$-VAE theory, and end with the empirical results.

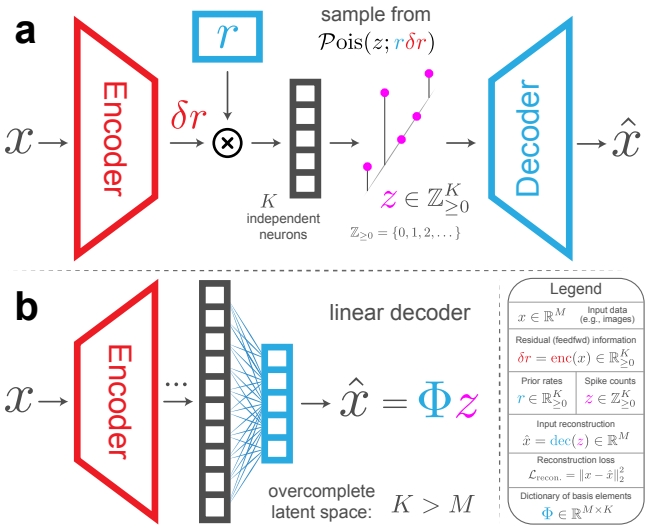

Figure 1: Introducing the Poisson VAE ($\mathcal{P}$-VAE). **(a)** Model architecture. Colored shapes depict learnable model parameters, including the prior firing rates, r. We color code the model's inference and generative components using red and blue, respectively. The $\mathcal{P}$-VAE encodes its inputs in discrete spike counts, z, significantly enhancing its bio-realism. **(b)** "Amortized Sparse Coding" as a special case of the $\mathcal{P}$-VAE.

Background, Methods, and Results

We consider a probabilistic generative model $p(\boldsymbol{x}, \boldsymbol{z})$ of observed data $\boldsymbol{x} \in \mathbb{R}^M$, and K-dimensional latent variables $\boldsymbol{z}$, with a data generative process given by $p(\boldsymbol{x}) = \int p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})\,d\boldsymbol{z}$. VAEs learn an approximate posterior, $q(\boldsymbol{z}|\boldsymbol{x})$, from data by minimizing the following loss function:

$$\mathcal{F}(q) = -\mathbb{E}_{\boldsymbol{z}\sim q(\boldsymbol{z}|\boldsymbol{x})}\Big[\log p(\boldsymbol{x}|\boldsymbol{z})\Big] + \mathcal{D}_{\mathrm{KL}}\Big(q(\boldsymbol{z}|\boldsymbol{x})\,\Big\|\,p(\boldsymbol{z})\Big). \quad (1)$$

The first term, the "reconstruction loss", describes how well the VAE reconstructs the data $\boldsymbol{x}$, and typically uses the mean squared error (Fig. 1). The second term, the "KL term", is the Kullback-Leibler divergence, $\mathcal{D}_{\mathrm{KL}}$, of the approximate posterior from a prior $p(\boldsymbol{z})$. Practitioners have complete autonomy in choosing the form of these probability distributions, however, the vast majority of VAE literature uses Gaussian distributions, $p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{1})$ and $q(\boldsymbol{z}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{\sigma}^2(\boldsymbol{x}))$, where $\boldsymbol{\mu}(\boldsymbol{x})$ and $\boldsymbol{\sigma}^2(\boldsymbol{x})$ are produced by the encoder network.

Poisson Variational Autoencoder ($\mathcal{P}$-VAE). Our key innovation is integrating Poisson-distributed latents into VAEs. We set $p(\boldsymbol{z}) = \mathcal{P}\mathrm{ois}(\boldsymbol{z}; \boldsymbol{r}_{\mathrm{prior}})$ and $q(\boldsymbol{z}|\boldsymbol{x}) = \mathcal{P}\mathrm{ois}(\boldsymbol{z}; \boldsymbol{r}_{\mathrm{post.}}(\boldsymbol{x}))$ as our prior and approximate posterior distributions, respectively, where $\mathcal{P}\mathrm{ois}(\boldsymbol{z}; \boldsymbol{\lambda}) = \prod_{i=1}^K \lambda_i^{z_i} e^{-\lambda_i}/z_i!$, and $\boldsymbol{z} \in \mathbb{Z}_{\geq 0}^K$ are discrete spike count variables. The prior rates, $\boldsymbol{r}_{\mathrm{prior}}$, are learnable parameters, and $\boldsymbol{r}_{\mathrm{post.}}(\boldsymbol{x})$ are produced by pushing the data sample $\boldsymbol{x}$ through the encoder neural network. In this paper, we will color code the encoder- and decoder-related parameters using red and blue, respectively (Fig. 1).

Predictive coding and the $\mathcal{P}$-VAE. Predictive coding posits that the cortex maintains predictions of incoming sensory information, and only prediction errors are propagated up the cortical hierarchy (Rao & Ballard, 1999). This idea is seamlessly incorporated into $\mathcal{P}$-VAE by assuming multiplicative interactions between representation units, $\boldsymbol{r}_{\mathrm{prior}}$, and the feedforward encoding information, $\boldsymbol{r}_{\mathrm{post.}}(\boldsymbol{x})$. Explicitly, let $\boldsymbol{r}_{\mathrm{prior}} \to \boldsymbol{r}$, and $\boldsymbol{r}_{\mathrm{post.}}(\boldsymbol{x}) \to \boldsymbol{r}\boldsymbol{\delta r}(\boldsymbol{x})$. Thus, the encoder only produces the residual information, $\boldsymbol{\delta r}(\boldsymbol{x})$ (Fig. 1a). We plug these assumptions into the general VAE loss function in eq. (1) to obtain the $\mathcal{P}$-VAE objective:

$$\mathcal{L}_{\mathcal{P}\mathrm{VAE}} = \mathbb{E}_{\boldsymbol{z}\sim\mathcal{P}\mathrm{ois}(\boldsymbol{z};\boldsymbol{r}\boldsymbol{\delta r})}\Big[\|\boldsymbol{x} - \mathrm{dec}(\boldsymbol{z})\|_2^2\Big] + \sum_{i=1}^K r_i f(\delta r_i), \quad (2)$$

where $f(y) = 1 - y + y\log y$. The second term in eq. (2) stems from the KL term in eq. (1), and its specific form is uniquely determined by our choice of Poisson distributions. This term is minimized by reducing the representation unit firing rates ($\boldsymbol{r} \approx \boldsymbol{0}$). Because firing rates are non-negative, it is reminiscent of the L_1 penalty used in sparse coding and reflects a metabolic cost for spiking.

Amortized Sparse Coding as a special case of the $\mathcal{P}$-VAE. The relationship between $\mathcal{P}$-VAE and sparse coding can be seen directly if (i) the decoder of $\mathcal{P}$-VAE is linear; and, (ii) the latent space is overcomplete (Fig. 1b). In this case, the decoder generates an image identically to the linear generative model from Olshausen and Field (1996): $\hat{\boldsymbol{x}} = \boldsymbol{\Phi}\boldsymbol{z}$, where $\boldsymbol{\Phi} \in \mathbb{R}^{M\times K}$ is a dictionary of basis elements with $K > M$. The key difference is that in the $\mathcal{P}$-VAE, inference over latents $\boldsymbol{z}$ is performed by the encoder, as opposed to numerically optimized. In this sense, $\mathcal{P}$-VAE with a linear decoder instantiates amortized sparse coding.

We trained a linear $\mathcal{P}$-VAE on 16×16 natural image patches extracted from the DOVES dataset (Bovik, Cormack, Linde, & Rajashekar, 2009), comparing it with the Gaussian counterpart that features a continuous latent space. Figure 2 shows the learned dictionaries. As expected, the Gaussian model's basis elements resemble those of principal component analysis (PCA), aligning with previous results that linear Gaussian VAEs are equivalent to probabilistic PCA (Tipping & Bishop, 1999). In sharp contrast, the $\mathcal{P}$-VAE learned Gabor-like feature selectivity, reminiscent of sparse coding patterns.

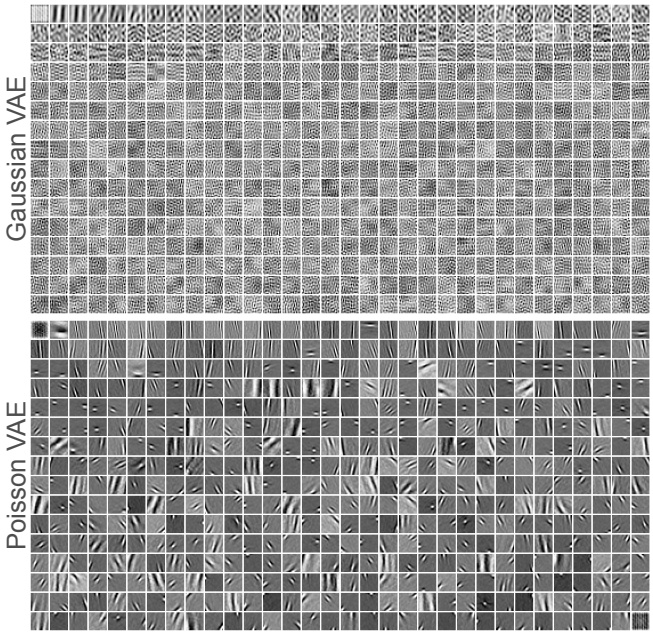

Figure 2: Learned basis elements (512 total, each made of $16\times16 = 256$ pixels. In other words, $\boldsymbol{\Phi} \in \mathbb{R}^{256\times512}$). Features are ordered from top-left to bottom-right, in ascending order of their associated variance, $\boldsymbol{\sigma}^2$, and, prior firing rate, $\boldsymbol{r}$, for Gaussian and Poisson VAEs, respectively.

Conclusions

We introduced the Poisson Variational Autoencoder ($\mathcal{P}$-VAE), which: **(1)** encodes inputs into discrete spike counts, making it a bio-realistic candidate model for sensory processing; **(2)** brings together major theoretical concepts in neuroscience, such as predictive and sparse coding, under the umbrella of variational Bayesian inference; and **(3)** sets the stage for developing deep hierarchical models to advance our understanding of perception as hierarchical Bayesian inference.

Acknowledgments

This work was supported by grants from the NIH R00EY032179 (to J.Y.).

References

Bovik, A., Cormack, L., Linde, I. V. D., & Rajashekar, U. (2009). Doves: a database of visual eye movements. *Spatial Vision*, *22*(2), 161 - 177. doi: 10.1163/156856809787465636

Csikor, F., Meszéna, B., & Orbán, G. (2023). Top-down perceptual inference shaping the activity of early visual cortex. *bioRxiv*. doi: 10.1101/2023.11.29.569262

Higgins, I., Chang, L., Langston, V., Hassabis, D., Summerfield, C., Tsao, D., & Botvinick, M. (2021). Unsupervised deep learning identifies semantic disentanglement in single inferotemporal face patch neurons. *Nature Communications*, *12*(1), 6456. doi: 10.1038/s41467-021-26751-5

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational bayes.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609. doi: 10.1038/381607a0

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. doi: 10.1038/4580

Rezende, D. J., Mohamed, S., & Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning* (pp. 1278–1286). Retrieved from https://proceedings.mlr.press/v32/rezende14.html

Storrs, K. R., Anderson, B. L., & Fleming, R. W. (2021). Unsupervised learning predicts human perception and misperception of gloss. *Nature Human Behaviour*, *5*(10), 1402–1417. doi: 10.1038/s41562-021-01097-6

Tipping, M. E., & Bishop, C. M. (1999, 01). Probabilistic principal component analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *61*(3), 611-622. doi: 10.1111/1467-9868.00196

Vafaii, H., Yates, J. L., & Butts, D. A. (2023). Hierarchical VAEs provide a normative account of motion processing in the primate brain. In *Thirty-seventh conference on neural information processing systems.* Retrieved from https://openreview.net/forum?id=1wOkHN9JK8