Variational Learning of Fractional Posteriors

Kian Ming A. Chai
DSO National Laboratories

Edwin V. Bonilla
CSIRO's Data61

42nd International Conference on Machine Learning
ICML 2025

Chai conducted the research while on sabbatical leave.

Given a probabilistic model with likelihood $p(D|z)$ and prior $p(z)$,

we want log evidence $\mathcal{L}_{\text{evd}} \stackrel{\text{def}}{=} \log \mathbb{E}_{z \sim p}[p(D|z)]$

and a posterior $q(z)$ combining data D with prior.

Given a probabilistic model with likelihood $p(D|z)$ and prior $p(z)$,

we want log evidence $\mathcal{L}_{\text{evd}} \stackrel{\text{def}}{=} \log \mathbb{E}_{z\sim p}[p(D|z)]$

and a posterior $q(z)$ combining data D with prior.

A common approach to approximate the Bayes posterior is

$$q^* = \arg\max_{q\in\mathcal{Q}} \mathcal{L}_{\text{ELBO}}$$

$$\mathcal{L}_{\text{evd}} \geq \mathcal{L}_{\text{ELBO}} \stackrel{\text{def}}{=} \mathbb{E}_{z\sim q}[\log p(D|z)] - \mathbb{E}_{z\sim q}\left[\log \frac{q(z)}{p(z)}\right].$$

Given a probabilistic model with likelihood $p(D|z)$ and prior $p(z)$,

we want log evidence $\mathcal{L}_{\text{evd}} \stackrel{\text{def}}{=} \log \mathbb{E}_{z\sim p}[p(D|z)]$

and a posterior $q(z)$ combining data D with prior.

A common approach to approximate the Bayes posterior is

$$q^* = \arg\max_{q\in\mathcal{Q}} \mathcal{L}_{\text{ELBO}}$$

$$\mathcal{L}_{\text{evd}} \geq \mathcal{L}_{\text{ELBO}} \stackrel{\text{def}}{=} \mathbb{E}_{z\sim q}[\log p(D|z)] - \mathbb{E}_{z\sim q}\left[\log \frac{q(z)}{p(z)}\right].$$

This work introduces a one-parameter $\gamma \in (0, 1)$ lower bound

$$\mathcal{L}_{\text{evd}} \geq \mathcal{L}_{\gamma} \stackrel{\text{def}}{=} \frac{1}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[p(D|z)^{1-\gamma}\right] - \frac{\gamma}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[\left(\frac{q(z)}{p(z)}\right)^{(1-\gamma)/\gamma}\right]$$

Given a probabilistic model with likelihood $p(D|z)$ and prior $p(z)$,

we want log evidence $\mathcal{L}_{\mathrm{evd}} \stackrel{\mathrm{def}}{=} \log \mathbb{E}_{z\sim p}[p(D|z)]$

and a posterior $q(z)$ combining data D with prior.

A common approach to approximate the Bayes posterior is

$$q^* = \arg\max_{q \in \mathcal{Q}} \mathcal{L}_{\mathrm{ELBO}}$$

$$\mathcal{L}_{\mathrm{evd}} \geq \mathcal{L}_{\mathrm{ELBO}} \stackrel{\mathrm{def}}{=} \mathbb{E}_{z\sim q}[\log p(D|z)] - \mathbb{E}_{z\sim q}\left[\log \frac{q(z)}{p(z)}\right].$$

This work introduces a one-parameter $\gamma \in (0,1)$ lower bound

$$\mathcal{L}_{\mathrm{evd}} \geq \mathcal{L}_{\gamma} \stackrel{\mathrm{def}}{=} \frac{1}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[p(D|z)^{1-\gamma}\right] - \frac{\gamma}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[\left(\frac{q(z)}{p(z)}\right)^{(1-\gamma)/\gamma}\right]$$

Given a probabilistic model with likelihood $p(D|z)$ and prior $p(z)$,

we want log evidence $\mathcal{L}_{\text{evd}} \stackrel{\text{def}}{=} \log \mathbb{E}_{z\sim p}[p(D|z)]$

and a posterior $q(z)$ combining data D with prior.

A common approach to approximate the Bayes posterior is

$$q^* = \arg\max_{q\in\mathcal{Q}} \mathcal{L}_{\text{ELBO}}$$

$$\mathcal{L}_{\text{evd}} \geq \mathcal{L}_{\text{ELBO}} \stackrel{\text{def}}{=} \mathbb{E}_{z\sim q}[\log p(D|z)] - \mathbb{E}_{z\sim q}\left[\log \frac{q(z)}{p(z)}\right].$$

This work introduces a one-parameter $\gamma \in (0,1)$ lower bound

$$\mathcal{L}_{\text{evd}} \geq \mathcal{L}_{\gamma} \stackrel{\text{def}}{=} \frac{1}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[p(D|z)^{1-\gamma}\right] - \frac{\gamma}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[\left(\frac{q(z)}{p(z)}\right)^{(1-\gamma)/\gamma}\right]$$

Given a probabilistic model with likelihood $p(D|z)$ and prior $p(z)$,

we want log evidence $\mathcal{L}_{\mathrm{evd}} \stackrel{\mathrm{def}}{=} \log \mathbb{E}_{z\sim p}[p(D|z)]$

and a posterior $q(z)$ combining data D with prior.

A common approach to approximate the Bayes posterior is

$$q^* = \arg\max_{q\in\mathcal{Q}} \mathcal{L}_{\mathrm{ELBO}}$$

$$\mathcal{L}_{\mathrm{evd}} \geq \mathcal{L}_{\mathrm{ELBO}} \stackrel{\mathrm{def}}{=} \mathbb{E}_{z\sim q}[\log p(D|z)] - \mathbb{E}_{z\sim q}\left[\log \frac{q(z)}{p(z)}\right].$$

This work introduces a one-parameter $\gamma \in (0,1)$ lower bound

$$\mathcal{L}_{\mathrm{evd}} \geq \mathcal{L}_{\gamma} \stackrel{\mathrm{def}}{=} \frac{1}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[p(D|z)^{1-\gamma}\right] - \frac{\gamma}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[\left(\frac{q(z)}{p(z)}\right)^{(1-\gamma)/\gamma}\right]$$

Given a probabilistic model with likelihood $p(D|z)$ and prior $p(z)$,

we want log evidence $\mathcal{L}_{\text{evd}} \stackrel{\text{def}}{=} \log \mathbb{E}_{z \sim p}[p(D|z)]$

and a posterior $q(z)$ combining data D with prior.

A common approach to approximate the Bayes posterior is

$$q^* = \arg\max_{q \in \mathcal{Q}} \mathcal{L}_{\text{ELBO}}$$

$$\mathcal{L}_{\text{evd}} \geq \mathcal{L}_{\text{ELBO}} \stackrel{\text{def}}{=} \mathbb{E}_{z \sim q}[\log p(D|z)] - \mathbb{E}_{z \sim q}\left[\log \frac{q(z)}{p(z)}\right].$$

This work introduces a one-parameter $\gamma \in (0, 1)$ lower bound

$$\mathcal{L}_{\text{evd}} \geq \mathcal{L}_{\gamma} \stackrel{\text{def}}{=} \frac{1}{1-\gamma} \log \mathbb{E}_{z \sim q}\left[p(D|z)^{1-\gamma}\right] - \frac{\gamma}{1-\gamma} \log \mathbb{E}_{z \sim q}\left[\left(\frac{q(z)}{p(z)}\right)^{(1-\gamma)/\gamma}\right]$$

Given a probabilistic model with likelihood $p(D|z)$ and prior $p(z)$,

we want log evidence $\mathcal{L}_{\mathrm{evd}} \stackrel{\mathrm{def}}{=} \log \mathbb{E}_{z\sim p}[p(D|z)]$

and a posterior $q(z)$ combining data D with prior.

A common approach to approximate the Bayes posterior is

$$q^* = \arg\max_{q\in\mathcal{Q}} \mathcal{L}_{\mathrm{ELBO}}$$

$$\mathcal{L}_{\mathrm{evd}} \geq \mathcal{L}_{\mathrm{ELBO}} \stackrel{\mathrm{def}}{=} \mathbb{E}_{z\sim q}[\log p(D|z)] - \mathbb{E}_{z\sim q}\left[\log \frac{q(z)}{p(z)}\right].$$

This work introduces a one-parameter $\gamma \in (0,1)$ lower bound

$$\mathcal{L}_{\mathrm{evd}} \geq \mathcal{L}_{\gamma} \stackrel{\mathrm{def}}{=} \frac{1}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[p(D|z)^{1-\gamma}\right] - \frac{\gamma}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[\left(\frac{q(z)}{p(z)}\right)^{(1-\gamma)/\gamma}\right]$$

Given a probabilistic model with likelihood $p(D|z)$ and prior $p(z)$,

we want log evidence $\mathcal{L}_{\rm evd} \stackrel{\rm def}{=} \log \mathbb{E}_{z\sim p}[p(D|z)]$

and a posterior $q(z)$ combining data D with prior.

A common approach to approximate the Bayes posterior is

$$q^* = \arg\max_{q\in\mathcal{Q}} \mathcal{L}_{\rm ELBO}$$

$$\mathcal{L}_{\rm evd} \geq \mathcal{L}_{\rm ELBO} \stackrel{\rm def}{=} \mathbb{E}_{z\sim q}[\log p(D|z)] - \mathbb{E}_{z\sim q}\left[\log \frac{q(z)}{p(z)}\right].$$

This work introduces a one-parameter $\gamma \in (0,1)$ lower bound

$$\mathcal{L}_{\rm evd} \geq \mathcal{L}_{\gamma} \stackrel{\rm def}{=} \frac{1}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[p(D|z)^{1-\gamma}\right] - \frac{\gamma}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[\left(\frac{q(z)}{p(z)}\right)^{(1-\gamma)/\gamma}\right]$$

Given a probabilistic model with likelihood $p(D|z)$ and prior $p(z)$,

we want log evidence $\mathcal{L}_{\mathrm{evd}} \stackrel{\mathrm{def}}{=} \log \mathbb{E}_{z\sim p}[p(D|z)]$

and a posterior $q(z)$ combining data D with prior.

A common approach to approximate the Bayes posterior is

$$q^* = \arg\max_{q \in \mathcal{Q}} \mathcal{L}_{\mathrm{ELBO}}$$

$$\mathcal{L}_{\mathrm{evd}} \geq \mathcal{L}_{\mathrm{ELBO}} \stackrel{\mathrm{def}}{=} \mathbb{E}_{z\sim q}[\log p(D|z)] - \mathbb{E}_{z\sim q}\left[\log \frac{q(z)}{p(z)}\right].$$

This work introduces a one-parameter $\gamma \in (0,1)$ lower bound

$$\mathcal{L}_{\mathrm{evd}} \geq \mathcal{L}_{\gamma} \stackrel{\mathrm{def}}{=} \frac{1}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[p(D|z)^{1-\gamma}\right] - \frac{\gamma}{1-\gamma} \log \mathbb{E}_{z\sim q}\left[\left(\frac{q(z)}{p(z)}\right)^{(1-\gamma)/\gamma}\right]$$

$$q^* = \arg\max_{q \in \mathcal{Q}} \mathcal{L}_{\gamma} \implies q^* \underset{\sim}{\propto} p(D|z)^{\gamma} p(z).$$

The derivation of the bound uses Hölder's inequality.

The derivation of the bound uses Hölder's inequality.

The regularisation term is the Rényi divergence.

The derivation of the bound uses Hölder's inequality.

The regularisation term is the Rényi divergence.

The bound generalises and includes $\mathcal{L}_{\mathrm{ELBO}}$:

$$\lim_{\gamma \to 1} \mathcal{L}_\gamma = \mathcal{L}_{\mathrm{ELBO}}. \qquad \text{(by L'Hôpital rule)}$$

The derivation of the bound uses Hölder's inequality.

The regularisation term is the Rényi divergence.

The bound generalises and includes $\mathcal{L}_{\mathrm{ELBO}}$:

$$\lim_{\gamma \to 1} \mathcal{L}_{\gamma} = \mathcal{L}_{\mathrm{ELBO}}. \qquad \text{(by L'Hôpital rule)}$$

Optimising the bound gives an approximation to the fractional posterior,

which can help with robustness to model misspecifications.

The derivation of the bound uses Hölder's inequality.

The regularisation term is the Rényi divergence.

The bound generalises and includes $\mathcal{L}_{\mathrm{ELBO}}$:

$$\lim_{\gamma \to 1} \mathcal{L}_\gamma = \mathcal{L}_{\mathrm{ELBO}}. \qquad \text{(by L'Hôpital rule)}$$

Optimising the bound gives an approximation to the fractional posterior,

which can help with robustness to model misspecifications.

The tightness of the bound depends on both $\mathcal{Q}$ and γ.

The derivation of the bound uses Hölder's inequality.

The regularisation term is the Rényi divergence.

The bound generalises and includes $\mathcal{L}_{\mathrm{ELBO}}$:

$$\lim_{\gamma \to 1} \mathcal{L}_\gamma = \mathcal{L}_{\mathrm{ELBO}}. \qquad \text{(by L'Hôpital rule)}$$

Optimising the bound gives an approximation to the fractional posterior,

which can help with robustness to model misspecifications.

The tightness of the bound depends on both $\mathcal{Q}$ and γ.

A new tool in the toolbox of objectives.

Two-component GMM calibration at $\alpha = 0.05$. $\mathcal{L}_{0.8}$ gives best calibrated posteriors.

Obj.	μ_1 coverage	μ_2 coverage
$\mathcal{L}_{0.7}$	0.9694	0.9606
$\mathcal{L}_{0.8}$	0.9568	0.9458
$\mathcal{L}_{0.9}$	0.9438	0.9334
ELBO	0.9278	0.9182

Two-component GMM calibration at $\alpha = 0.05$. $\mathcal{L}_{0.8}$ gives best calibrated posteriors.

Obj.	μ_1 coverage	μ_2 coverage
$\mathcal{L}_{0.7}$	0.9694	0.9606
$\mathcal{L}_{0.8}$	0.9568	0.9458
$\mathcal{L}_{0.9}$	0.9438	0.9334
ELBO	0.9278	0.9182

Optimising VAE on MNIST. $\mathcal{L}_{0.1}$ gives highest test objective.

Obj.	Test Objective
$\mathcal{L}_{0.1}$	1677.2
$\mathcal{L}_{0.5}$	1672.8
$\mathcal{L}_{0.9}$	1639.3
ELBO	1583.2

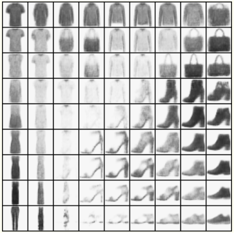
Two-component GMM calibration at $\alpha = 0.05$. $\mathcal{L}_{0.8}$ gives best calibrated posteriors.

Obj.	μ_1 coverage	μ_2 coverage
$\mathcal{L}_{0.7}$	0.9694	0.9606
$\mathcal{L}_{0.8}$	0.9568	0.9458
$\mathcal{L}_{0.9}$	0.9438	0.9334
ELBO	0.9278	0.9182

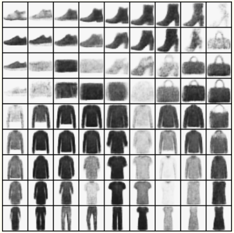
Optimising VAE on MNIST. $\mathcal{L}_{0.1}$ gives highest test objective.

Obj.	Test Objective
$\mathcal{L}_{0.1}$	1677.2
$\mathcal{L}_{0.5}$	1672.8
$\mathcal{L}_{0.9}$	1639.3
ELBO	1583.2

Generation from prior in VAE for Fashion-MNIST. $\mathcal{L}_{10^{-5}}$ gives clearer images and a lower Fréchet inception distance (FID).



ELBO; FID=83.5



$\mathcal{L}_{10^{-5}}$; FID=68.8