Zig-zag computational bounds and further implementation details

Abstract

Full implementation details for zig-zag algorithms used in "Exact Sampling of Gibbs Measures with Estimated Losses" by Frazier, Knoblauch, Jewson and Drovandi.

1 Preliminaries

Firstly we recall Cinlar's Algorithm (Cinlar, 1975) for sampling from inhomogeneous Poisson process.

Algorithm 1: Cinlar's Algorithm (Cinlar, 1975) to sample $\mathbb{P}(\tau_j \geq t) = \exp(-\Lambda(t))$

1. Initialisation: $s = 0$.
2. Generate $u \sim U(0, 1)$
3. Set $s \mapsto s - log(u)$
4. Return $t \mapsto \inf\{v : \Lambda(v) \geq s\}$

We also recall Poisson superposition which helps to implement Cinlar's Algorithm for some of the applications to come

Theorem 1 (Poisson superposition). *Let $\Lambda^{(1)} : \mathbb{R}_+ \mapsto \mathbb{R}_+$ and $\Lambda^{(2)} : \mathbb{R}_+ \mapsto \mathbb{R}_+$ be continuous for $t \geq 0$. Let $\tau_1^{(1)}, \tau_2^{(1)}, \ldots$ be a increasing finite or infinite sequence of points of a Poisson process with rate function $(\Lambda^{(1)}(t))_{t\geq 0}$, and $\tau_1^{(2)}, \tau_2^{(2)}, \ldots$ be another increasing finite or infinite sequence of points of a Poisson process with rate function $(\Lambda^{(2)}(t))_{t\geq 0}$.*

Then the ordered infinite sequence obtained from the union of $\tau_1^{(1)}, \tau_2^{(1)}, \ldots$ *and* $\tau_1^{(2)}, \tau_2^{(2)}, \ldots$, *say,* $\tilde{\tau}_1, \tilde{\tau}_2, \ldots$ *say, form a nonhomogeneous Poisson process with rate function* $(\Lambda^{(1)}(t) + \Lambda^{(2)}(t))_{t\geq 0}$.

2 Implementation Details

This section contains additional implementation details for the experiments presented in Section 5 of the paper. We start with an additional example where the zig-zag computational bounds are simple to derive.

2.1 Student's-t location model with the MMD

Consider estimating the location parameter $\theta = \{\mu\}$ of a Student's-t likelihood model with fixed scale σ and degree of freedom ν. We place a Gaussian prior on θ with mean $\mu_0 = 0$ and variance $s_0^2 = 25$ and consider sampling from the MMD-Bayes posterior using the RBF kernel with hyperparameter $\gamma = 1$. We *indirectly* sample $u_{1:b} \sim t_\nu(\mu, \sigma^2)$ by first sampling $v_{1:b}$ with $v_k = \{v_{k1}, v_{k2}, v_{k3}\} \sim p_v$ and $p_v(v) = unif(v_1; 0, 1)unif(v_2; 0, 1)\mathcal{G}(v_3; \nu/2, \nu/2)$ and setting

$$u_k = G_\mu(v; \eta, \sigma^2) = \mu + \sqrt{\sigma^2/v_3}\sqrt{-2\log(v_1)}cos(2\pi v_2).$$

Lemma 1 provides computational bounds for this example and Corollary 1 shows how these can be sampled from.

Lemma 1 (Computational Bounds for MMD-Bayes for Student's-t location model)**.** *Consider the MMD-Bayes posterior for a Student's-t location model with Gaussian priors with mean* μ_0 *and variance* s_0^2 *on unknown location parameter* $\theta = \{\mu\}$ *and fixed degrees of* *freedom ν and scale parameter σ. The zig-zag switching rate is given by*

$$\hat{\lambda}(\mu,\nu) := \max\left(0, \nu\frac{\mu-\mu_0}{s_0^2} + 2\nu\frac{\omega}{n}\sum_{i=1}^{n}\frac{1}{b}\sum_{k=1}^{b}\frac{(G_\mu(v_k;\eta,\sigma^2)-y_i)}{\sqrt{2\pi}\gamma^{3/2}}\exp\left(-\frac{(G_\mu(v_k;\eta,\sigma^2)-y_i)^2}{2\gamma}\right)\right)$$

where $v_1,\ldots,v_m \sim p_v$,

$$G_\mu(v;\eta,\sigma^2) = \mu + \sqrt{\sigma^2/v_3}\sqrt{-2\log(v_1)}cos(2\pi v_2),$$

and $p_v(v) = unif(v_1;0,1)unif(v_2;0,1)\mathcal{G}(v_3;\eta/2,\eta/2)$.

Computational bounds are given by

$$\hat{\Lambda}_j(\mu+\nu\cdot t,\nu) = \frac{|\mu-\mu_0|}{s_0^2} + \frac{2\omega}{\sqrt{2\pi}\gamma}\exp\left(-\frac{1}{2}\right) + \frac{1}{s_0^2}t.$$

Proof. Firstly, we can *indirectly* sample $u_{1:b} \sim t_\nu(\mu,\sigma^2)$ by first sampling $v_{1:b}$ with $v_k = \{v_{k1}, v_{k2}, v_{k3}\} \sim p_v$ and $p_v(v) = unif(v_1;0,1)unif(v_2;0,1)\mathcal{G}(v_3;\eta/2,\eta/2)$ and setting $u_k = G_\mu(v_k;\eta,\sigma^2)$

$$G_\mu(v;\eta,\sigma^2) = \mu + \sqrt{\sigma^2/v_3}\sqrt{-2\log(v_1)}cos(2\pi v_2).$$

As a result, for the MMD-Bayes posterior with RBF kernel with hyperparameter γ and

Gaussian prior has

$$
\begin{aligned}
&\Psi_n(\mu) := -\log \pi(\mu) + w\mathsf{L}_n^\gamma(\mu) \\
&= \frac{(\mu-\mu_0)^2}{2s_0^2} - 2\frac{w}{n}\sum_{i=1}^n \mathbb{E}_{v\sim p(\cdot)}\left[\mathcal{K}_\gamma(y_i, G_\mu(v;\eta,\sigma^2))\right] + \omega\mathbb{E}_{v,v'\sim p(\cdot)}\left[\mathcal{K}_\gamma(G_\mu(v;\eta,\sigma^2), F_{\eta,\sigma^2}(v';\mu))\right] \\
&= \frac{(\mu-\mu_0)^2}{2s_0^2} - 2\frac{w}{n}\sum_{i=1}^n \mathbb{E}_{v\sim p(\cdot)}\left[\frac{1}{\sqrt{2\pi\gamma}}\exp\left(\frac{(y_i - G_\mu(v;\eta,\sigma^2)))^2}{2\gamma}\right)\right] \\
&+ \omega\mathbb{E}_{v,v'\sim p(\cdot)}\left[\frac{1}{\sqrt{2\pi\gamma}}\exp\left(\frac{(G_\mu(v;\eta,\sigma^2) - F_{\eta,\sigma^2}(v';\mu)))^2}{2\gamma}\right)\right]
\end{aligned}
$$

The zig-zag requires the evaluation of

$$
\hat{\lambda}(\mu,\nu) := \max\left(0, \nu\left(-\frac{\partial}{\partial\mu}\log(\pi(\mu)) + \omega\varphi_{b,n}(\mu)\right)\right)
$$

and bounding of $\hat{\lambda}(\mu + \nu\cdot t, \nu)$, where

$$
\begin{aligned}
&\varphi_{b,n}(\mu) := 2\frac{1}{n}\sum_{i=1}^n \frac{1}{b}\sum_{k=1}^b \frac{(G_\mu(v_k;\eta,\sigma^2) - y_i)}{\sqrt{2\pi}\gamma^{3/2}}\exp\left(-\frac{(G_\mu(v_k;\eta,\sigma^2) - y_i)^2}{2\gamma}\right) \\
&- \frac{1}{b(b-1)}\sum_{k=1}^b\sum_{k'\neq k}\left(\frac{(G_\mu(v_k;\eta,\sigma^2) - G_\mu(v_{k'};\eta,\sigma^2))}{\sqrt{2\pi}\gamma^{3/2}}\exp\left(-\frac{(G_\mu(v_k;\eta,\sigma^2) - G_\mu(v_{k'};\eta,\sigma^2))^2}{2\gamma}\right)\right. \\
&\left.-\frac{(G_\mu(v_k;\eta,\sigma^2) - G_\mu(v_{k'};\eta,\sigma^2))}{\sqrt{2\pi}\gamma^{3/2}}\exp\left(-\frac{(G_\mu(v_k;\eta,\sigma^2) - G_\mu(v_{k'};\eta,\sigma^2))^2}{2\gamma}\right)\right) \\
&= \frac{\mu-\mu_0}{s_0^2} + 2\frac{1}{n}\sum_{i=1}^n \frac{1}{b}\sum_{k=1}^b \frac{(G_\mu(v_k;\eta,\sigma^2) - y_i)}{\sqrt{2\pi}\gamma^{3/2}}\exp\left(-\frac{(G_\mu(v_k;\eta,\sigma^2) - y_i)^2}{2\gamma}\right)
\end{aligned}
$$

with $v_1,\ldots,v_N \sim p_v$. Above uses the fact that

$$
\begin{aligned}
\frac{\partial}{\partial z}\mathcal{K}_\gamma(z,x) &= -\frac{(z-x)}{\sqrt{2\pi}\gamma^{3/2}}\exp\left(-\frac{(z-x)^2}{2\gamma}\right) \\
\frac{\partial}{\partial\mu}G_\mu(v;\eta,\sigma^2) &= 1
\end{aligned}
$$

and the cancellation of the second MMD term happens as $\frac{\partial}{\partial\mu}G_\mu(v;\eta,\sigma^2)$ does not depend on v.

Now, using the fact that for the RBF kernel $\left|\frac{\partial}{\partial z}\mathcal{K}_\gamma(y, y')\right| \leq \frac{1}{\sqrt{2\pi\gamma}} \exp\left(-\frac{1}{2}\right)$ for any value of y or y' we can bound

$$\begin{aligned}
\hat{\lambda}(\mu + \nu \cdot t, \nu) &= \max\left(0, \nu\left(-\frac{\partial}{\partial\mu}\log(\pi(\mu + \nu \cdot t)) + \omega\varphi_{b,n}(\mu + \nu \cdot t)\right)\right) \\
&\leq \frac{|\mu - \mu_0|}{s_0^2} + \frac{2\omega}{\sqrt{2\pi}\gamma}\exp\left(-\frac{1}{2}\right) + \frac{1}{s_0^2}t \\
&= \hat{\Lambda}(\mu + \nu \cdot t, \nu).
\end{aligned}$$

$\square$

Corollary 1. *Sampling stopping times according to the computational bound in Lemma 1 is achieved by sampling $u \sim U[0, 1]$ and setting*

$$\tau^* := \frac{\sqrt{a^2 - 2log(u)b} - a}{b}$$

with

$$\begin{aligned}
a &:= \frac{|\mu - \mu_0|}{s_0^2} + \frac{2\omega}{\sqrt{2\pi}\gamma}\exp\left(-\frac{1}{2}\right) \\
b &:= \frac{1}{s_0^2}.
\end{aligned}$$

Proof. Sampling stopping times according to the computational bounds in Lemma 1 require the evaluation of

$$\begin{aligned}\int_0^t \hat{\Lambda}(\mu + \nu \cdot s, \nu) ds &= \int_0^t a + bsds \\ &= \left[as + \frac{b}{2}s^2\right]_0^t \\ &= at + \frac{b}{2}t^2\end{aligned}$$

with

$$\begin{aligned} a &:= \frac{|\mu - \mu_0|}{s_0^2} + \frac{2\omega}{\sqrt{2\pi}\gamma} \exp\left(-\frac{1}{2}\right) \\ b &:= \frac{1}{s_0^2}\end{aligned}$$

and therefore the stopping times can be simulated following Cinlar's Algorithm (Cinlar, 1975) (Algorithm 1) by sampling $u \sim U[0,1]$ and solving

$$\begin{aligned} t &:= \inf\left\{\nu : a\nu + \frac{b}{2}\nu^2 \geq -\log(u)\right\} \\ &= \inf\left\{\nu : \frac{b}{2}\nu^2 + a\nu + \log(u) \geq 0\right\} \\ &= \frac{-a \pm \sqrt{a^2 - 4\log(u)\frac{b}{2}}}{b_j} \\ &= \frac{\sqrt{a^2 - 2\log(u)b} - a}{b}.\end{aligned}$$

□

2.2 Outlier Robust linear regression with the MMD

Here we provide implementation details for Section 5.2.1 of the paper.

Firstly, the Gaussian regression model has an implicit representation as:

$$G_\theta(v;x) = \beta^\top x + \sqrt{\sigma^2}\sqrt{-2\log(v_1)}cos(2\pi v_2).$$

This representation makes clear that we can *indirectly* sample $u_{1:b} \sim \mathcal{N}(x^\top\beta, \sigma^2)$ by first sampling $v_{1:b}$ with $u_k = \{v_{k1}, v_{k2}\} \sim p_v$ and $p_v(v) = unif(v_1; 0,1)unif(v_2; 0,1)$ and setting $u_k = G_\theta(v_k; x)$

Given this, we can conduct implicit sampling based on the following implicit representation of the MMD, which was referred to by Alquier and Gerber (2024) as the 'tilde'-type estimator (their Eq. (5) and (6)):

$$\begin{aligned}
\mathsf{L}_{m,n}(\theta, u_{1:b}) &= \mathsf{L}^{\mathcal{K}_\gamma}_{m,n}(y_{1:n}, G_\theta(v_{1:b}; x)) \\
&= -\sum_{i=1}^{n}\left\{2\frac{1}{b}\sum_{j=1}^{b}\mathcal{K}_\gamma(y_i, G_\theta(v_k; x_i)) \right.\\
&\left. -\frac{1}{b(b-1)}\sum_{k=1}^{b}\sum_{k'\neq k}^{b}\mathcal{K}_\gamma(G_\theta(v_k; x_i), G_\theta(v_{k'}; x_i))\right\}
\end{aligned}$$

where $\mathcal{K}_\gamma(y, y') = \frac{1}{\sqrt{2\pi\gamma}}\exp\left(-\frac{(y-y')^2}{2\gamma}\right)$ is the radial basis function (RBF) kernel. We place Gaussian priors with mean $\mu_0 = 0$ and variance $s_0^2 = 25$ on each β_j, $j = 1, \ldots, d$ and an inverse-gamma prior on σ^2 with shape $a_0 = 2$ and scale $b_0 = 0.5$. We reprarametrise $\sigma^2 \mapsto \log(\sigma)$ so that the zig-zag can sample from $\theta = \{\beta, \log(\sigma)\} \in \mathbb{R}^{p+1}$.

Lemma 2 provides computational bonds for this example and Corollary 2 shows how sampling from these is implemented.

Lemma 2 (Computational Bounds for MMD-Bayes for Gaussian regression model)**.** *Consider the MMD-Bayes posterior for a Gaussian regression model with Gaussian priors with mean μ_0 and variance s_0^2 on each regression coefficient β_j, $j = 1, \ldots, d$ and an inverse-gamma prior on scale parameters σ^2 with shape a_0 and scale b_0. We reprarametrise*

$\sigma^2 \mapsto \log(\sigma)$ *so that the zig-zag can sample from* $\theta = \{\beta, \log(\sigma)\} \in \mathbb{R}^{p+1}$. *The zig-zag switching rates are given by*

$$\hat{\lambda}_j(\theta,\nu) := \max\left(0, \nu_j \frac{\beta_j - \mu_0}{s_0^2} + 2\nu_j \frac{w}{n}\sum_{i=1}^{n}\frac{1}{b}\sum_{k=1}^{b}\frac{(G_\theta(v_k;x_i)-y_i)}{\sqrt{2\pi}\gamma^{3/2}}\exp\left(-\frac{(G_\theta(v_k;x_i)-y_i)^2}{2\gamma}\right)x_{ij}\right)$$

$j = 1, \ldots, d$ *and*

$$\begin{aligned}
&\hat{\lambda}_{d,d+1}(t;\theta,\nu) := \max\left(0, 2\nu_{p+1}a_0 - 2\nu_{p+1}b_0\exp(-2\log(\sigma))\right.\\
&-\nu_{p+1}\frac{w}{n}\sum_{i=1}^{n}\left\{-2\frac{1}{b}\sum_{k=1}^{b}\frac{(G_\theta(v_k;x_i)-y_i)}{\sqrt{2\pi}\gamma^{3/2}}\exp\left(-\frac{(G_\theta(v_k;x_i)-y_i)^2}{2\gamma}\right)\right.\\
&\times\exp(\log(\sigma))\sqrt{-2\log(v_{k1})}cos(2\pi v_{k2})\\
&+\frac{1}{b(b-1)}\sum_{k=1}^{b}\sum_{k'\neq k}\exp(\log(\sigma))\left(\sqrt{-2\log(v_{k1})}cos(2\pi v_{k2}) - \sqrt{-2\log(u_{k'1})}cos(2\pi u_{k'2})\right)\\
&\left.\left.\times\frac{(G_\theta(v_k;x_i)-G_\theta(v_{k'};x_i))}{\sqrt{2\pi}\gamma^{3/2}}\exp\left(-\frac{(G_\theta(v_k;x_i)-G_\theta(v_{k'};x_i))^2}{2\gamma}\right)\right\}\right)
\end{aligned}$$

where $v_1, \ldots, v_m \sim p_v$ *and*

$$G_\theta(v;x) = x^\top\beta + \sqrt{\sigma^2}\sqrt{-2\log(v_1)}cos(2\pi v_2).$$

and $p_v(v) = unif(v_1;0,1)unif(v_2;0,1)$.

Computational bounds are given by

$$\hat{\Lambda}_j(\theta+\nu\cdot t,\nu) := \frac{|\beta_j-\mu_0|}{s_0^2} + \frac{2\frac{\omega}{n}\sum_{i=1}^{n}|x_{ij}|}{\sqrt{2\pi}\gamma}\exp\left(-\frac{1}{2}\right) + \frac{1}{s_0^2}t$$

for $j = 1, \ldots, d$ *and*

$$\hat{\Lambda}_{p+1}(\theta+\nu\cdot t,\nu) := a_0 + 2b_0\exp(-2(\log(\sigma)+\nu_{p+1}t)) + \frac{2\omega}{\sqrt{2\pi}\sigma^2}\exp\left(-\frac{1}{2}\right)\exp(\log(\sigma)+\nu_{p+1}t)R$$

where R is an assumed upper bound on $\max_{v_1}\{\sqrt{-2\log(v_1)}\}$

The bounds in Lemma 2 are slightly different for $\theta_{p+1} = \log(\sigma)$ from those in Corollary 3 of the main paper. While Corollary 3 proves that an upper bound exists, obtaining those bounds requires solving an intractable optimisation. Lemma 2 is therefore an approximation of Corollary 3. As v_1 has support arbitrarily close to 0 $\sqrt{-2\log(v_1)}$ is in principle unbounded. However, Corollary 3 guarantees that as long as we take R big enough this will in fact be an upper bound, i.e. looser than the bounds in Corollary 3. We find that setting $R = \sqrt{-2\log 10^{-6/N}}$ provides a bound that is never exceeded. This bound comes from letting 10^{-6} be a tolerated probability that all N observations simulated observations be such that $\sqrt{-2\log v_1} > R$. We keep track of the maximum value of $\hat{\lambda}_{p+1}(\theta + \nu \cdot t, \nu)/\Lambda_{p+1}(\theta + \nu \cdot t, \nu)$ throughout the sampling and check that this is always less than 1.

Proof. Firstly, we can *indirectly* sample $u_{1:b} \sim \mathcal{N}(x^\top\beta, \sigma^2)$ by first sampling $v_{1:b}$ with $v_k = \{v_{k1}, v_{k2}\} \sim p_v$ and $p_0(v) = unif(v_1; 0, 1)unif(v_2; 0, 1)$ and setting $u_k = G_\theta(v_k; x)$

$$G_\theta(v; x) = x^\top\beta + \sqrt{\sigma^2}\sqrt{-2\log(v_1)}cos(2\pi v_2).$$

As a result, for the MMD-Bayes posterior

$$\begin{aligned}
\Psi_n(\theta) &:= -\sum_{j=1}^{p} \log \pi(\beta_j) - \log \pi(\log(\sigma)) + w\mathsf{L}_n^{\mathcal{K}_\gamma}(\theta) \\
&= \sum_{j=1}^{p} \frac{(\beta_j - \mu_0)^2}{2s_0^2} + 2a_0 \log(\sigma) + b_0 \exp(-2\log(\sigma)) \\
&- \frac{w}{n} \sum_{i=1}^{n} \{2\mathbb{E}_{v\sim p_v}\left[\mathcal{K}_\gamma(y_i, G_\theta(v; x_i))\right] - \mathbb{E}_{v,v'\sim p_v}\left[\mathcal{K}_\gamma(G_\theta(v; x_i)), G_\theta(v'; x_i)))\right]\} \\
&= \sum_{j=1}^{p} \frac{(\beta_j - \mu_0)^2}{2s_0^2} + 2a_0 \log(\sigma) + b_0 \exp(-2\log(\sigma)) \\
&- \frac{w}{n} \sum_{i=1}^{n} \left\{2\mathbb{E}_{v\sim p_v}\left[\frac{1}{\sqrt{2\pi\gamma}} \exp\left(\frac{(y_i - G_\theta(v; x_i))^2}{2\gamma}\right)\right]\right. \\
&\left. -\mathbb{E}_{v,v'\sim p_v}\left[\frac{1}{\sqrt{2\pi\gamma}} \exp\left(\frac{(G_\theta(v; x_i)) - G_\theta(v'; x_i)))^2}{2\gamma}\right)\right]\right\}
\end{aligned}$$

The zig-zag requires the evaluation of

$$\hat{\lambda}_j(\theta, \nu) := \max\left(0, \nu\left(-\frac{\partial}{\partial\theta_j} \log(\pi(\theta)) + \omega\varphi_{b,n}(\theta)\right)\right)$$

and bounding of $\hat{\lambda}_j(\theta + \nu \cdot t, \nu)$ for $j = 1, \ldots p + 1$.

For the regression coefficients: For $j = 1, \ldots p$ we have that

$$\begin{aligned}
[\varphi_{m,n}(\theta)]_j &= -\frac{1}{n} \sum_{i=1}^{n} \left\{-2\frac{1}{b} \sum_{k=1}^{b} \frac{(G_\theta(v_k; x_i) - y_i)}{\sqrt{2\pi}\gamma^{3/2}} \exp\left(-\frac{(G_\theta(v_k; x_i) - y_i)^2}{2\gamma}\right) x_{ij}\right. \\
&+ \frac{1}{b(b-1)} \sum_{k=1}^{b} \sum_{k' \neq k} \left(\frac{(G_\theta(v_k; x_i) - G_\theta(v_{k'}; x_i))}{\sqrt{2\pi}\gamma^{3/2}} \exp\left(-\frac{(G_\theta(v_k; x_i) - G_\theta(v_{k'}; x_i))^2}{2\gamma}\right) x_{ij}\right. \\
&\left.\left. - \frac{(G_\theta(v_k; x_i) - G_\theta(v_{k'}; x_i))}{\sqrt{2\pi}\gamma^{3/2}} \exp\left(-\frac{(G_\theta(v_k; x_i) - G_\theta(v_{k'}; x_i))^2}{2\gamma}\right) x_{ij}\right)\right\} \\
&= \frac{1}{n} \sum_{i=1}^{n} \left\{2\frac{1}{b} \sum_{k=1}^{b} \frac{(G_\theta(v_k; x_i) - y_i)}{\sqrt{2\pi}\gamma^{3/2}} \exp\left(-\frac{(G_\theta(v_k; x_i) - y_i)^2}{2\gamma}\right) x_{ij}\right\}.
\end{aligned}$$

with $v_1, \ldots, v_N \sim p_v$. Above uses the fact that

$$
\begin{aligned}
&\frac{\partial}{\partial z}\mathcal{K}_\gamma(z,x) = -\frac{(z-x)}{\sqrt{2\pi}\gamma^{3/2}} \exp\left(-\frac{(z-x)^2}{2\gamma}\right) \\
&\frac{\partial}{\partial \beta_j} G_\theta(v;x) = x_j
\end{aligned}
$$

and the cancellation of the second MMD term happens as $\frac{\partial}{\partial \beta_j} G_\theta(v;x)$ doesn't depend on v.

We can then use the fact that $\left|\frac{\partial}{\partial y}\mathcal{K}_\gamma(y,y')\right| \leq \frac{1}{\sqrt{2\pi\gamma}} \exp\left(-\frac{1}{2}\right)$ for any value of y or y' and bound

$$
\begin{aligned}
\hat{\lambda}_j(\theta + \nu \cdot t, \nu) &= \max\left(0, \nu\left(-\frac{\partial}{\partial \theta_j}\log(\pi(\theta + \nu \cdot t)) + \omega[\varphi_{b,n}(\theta + \nu \cdot t)]_j\right)\right) \\
&\leq \frac{|\beta_j - \mu_0|}{s_0^2} + \frac{2w\frac{1}{n}\sum_{i=1}^n |x_{ij}|}{\sqrt{2\pi\gamma}} \exp\left(-\frac{1}{2}\right) + \frac{1}{s_0^2}t \\
&= \hat{\Lambda}_j(\theta + \nu \cdot t, \nu)
\end{aligned}
$$

For the residual variance:

$$
\begin{aligned}
&[\varphi_{b,n}(\theta)]_{d+1} := \\
&-\frac{1}{n}\sum_{i=1}^n \left\{-2\frac{1}{b}\sum_{k=1}^b \frac{(G_\theta(v_k;x_i) - y_i)}{\sqrt{2\pi}\gamma^{3/2}} \exp\left(-\frac{(G_\theta(v_k;x_i) - y_i)^2}{2\gamma}\right) \exp(\log(\sigma))\sqrt{-2\log(v_{k1})} cos(2\pi v_{k2})\right. \\
&+\frac{1}{b(b-1)}\sum_{k=1}^b \sum_{k' \neq k} \left(\frac{(G_\theta(v_k;x_i) - G_\theta(v_{k'};x_i))}{\sqrt{2\pi}\gamma^{3/2}} \exp\left(-\frac{(G_\theta(v_k;x_i) - G_\theta(v_{k'};x_i))^2}{2\gamma}\right) \exp(\log(\sigma))\sqrt{-2\log(v_{k1}} \right. \\
&\left.\left.-\frac{(G_\theta(v_k;x_i) - G_\theta(v_{k'};x_i))}{\sqrt{2\pi}\gamma^{3/2}} \exp\left(-\frac{(G_\theta(v_k;x_i) - G_\theta(v_{k'};x_i))^2}{2\gamma}\right) \exp(\log(\sigma))\sqrt{-2\log(u_{k'1})} cos(2\pi u_{k'2})\right)\right\} \\
&= -\frac{1}{n}\sum_{i=1}^n \left\{-2\frac{1}{b}\sum_{k=1}^b \frac{(G_\theta(v_k;x_i) - y_i)}{\sqrt{2\pi}\gamma^{3/2}} \exp\left(-\frac{(G_\theta(v_k;x_i) - y_i)^2}{2\gamma}\right) \exp(\log(\sigma))\sqrt{-2\log(v_{k1})} cos(2\pi v_{k2})\right. \\
&+\frac{1}{b(b-1)}\sum_{k=1}^b \sum_{k' \neq k} \exp(\log(\sigma))\left(\sqrt{-2\log(v_{k1})} cos(2\pi v_{k2}) - \sqrt{-2\log(u_{k'1})} cos(2\pi u_{k'2})\right) \\
&\left.\times\frac{(G_\theta(v_k;x_i) - G_\theta(v_{k'};x_i))}{\sqrt{2\pi}\gamma^{3/2}} \exp\left(-\frac{(G_\theta(v_k;x_i) - G_\theta(v_{k'};x_i))^2}{2\gamma}\right)\right\}
\end{aligned}
$$

as

$$\frac{\partial}{\partial \log(\sigma)} G_\theta(v;x) = \exp(\log(\sigma))\sqrt{-2\log(u_1)} cos(2\pi u_2).$$

While Corollary 4 proves that a finite bound on this exists, the bounds derived in Corollary 4 are intractable to use in practice. Instead, we use the following *approximate* bound

$$\begin{aligned}
&\hat{\lambda}_{m,p+1}(\theta+\nu\cdot t,\nu) := \max\left(0, \nu\left(-\frac{\partial}{\partial\theta_{d+1}}\log(\pi(\theta+\nu\cdot t)) + \omega[\varphi_{b,n}(\theta+\nu\cdot t)]_{d+1}\right)\right)\\
&\le a_0 + 2b_0\exp(-2(\log(\sigma)+\nu_{p+1}t)) + \frac{2\omega}{\sqrt{2\pi}\sigma^2}\exp\left(-\frac{1}{2}\right)\exp(\log(\sigma)+\nu_{p+1}t)R\\
&= \Lambda_{p+1}(\theta+\nu\cdot t,\nu)
\end{aligned}$$

where R is an assumed upper bound on $\max_{v_1}\{\sqrt{-2\log(v_1)}\}$. □

Corollary 2. *Sampling stopping times according to the computational bound in Lemma 2 is achieved for $j=1,\dots,d$ by sampling $u_j \sim U[0,1]$ and setting*

$$\tau_j^* := \frac{\sqrt{a_j^2 - 2\log(u_j)b_j} - a_j}{b_j}$$

with

$$\begin{aligned}
a_j &:= \frac{|\beta_j-\mu_0|}{s_0^2} + \frac{2w\frac{1}{n}\sum_{i=1}^n |x_{ij}|}{\sqrt{2\pi}\gamma}\exp\left(-\frac{1}{2}\right)\\
b_j &:= \frac{1}{s_0^2}
\end{aligned}$$

and

$$\tau_{p+1}^* = \min\{\tau_{p+1}^{(1)}, \tau_{p+1}^{(2)}, \tau_{p+1}^{(3)}\}$$

with

$$
\tau^{(1)}_{p+1} = \frac{-\log(u^{(1)}_{p+1})}{a_0}
$$

$$
\tau^{(2)}_{p+1} = \begin{cases} -\frac{1}{2\nu_{p+1}} \log\left(1 + \frac{\log(u^{(2)}_{p+1})\nu_{p+1}\exp(2\log(\sigma))}{b_0}\right) & \mathit{if}\ -\log(u^{(2)}_{p+1}) < \frac{b_0}{\nu_{p+1}\exp(2\log(\sigma))} \\ \infty & \mathit{otherwise} \end{cases}
$$

$$
\tau^{(3)}_{p+1} = \begin{cases} \frac{1}{\nu_{p+1}} \log\left(1 - \frac{\log(u^{(3)}_{p+1})\nu_{p+1}\exp(-\log(\sigma))}{\left(\frac{2n}{\sqrt{2\pi}\gamma}\exp\left(-\frac{1}{2}\right)R\right)}\right) & \mathit{if}\ -\log(u^{(3)}_{p+1}) > -\frac{\left(\frac{2\omega}{\sqrt{2\pi}\gamma}\exp\left(-\frac{1}{2}\right)R\right)}{\nu_{p+1}\exp(-\log(\sigma))} \\ \infty & \mathit{otherwise} \end{cases}
$$

$u^{(1)}_{p+1}, u^{(2)}_{p+1}, u^{(3)}_{p+1} \sim Unif(0,1)$.

Proof. Sampling stopping times according to the computational bounds in Lemma 2 requires, for the regression coefficients, $j = 1, \dots, p$, following the procedure outlined in Corollary 1 with

$$
a_j := \frac{|\beta_j - \mu_0|}{s_0^2} + \frac{2w\frac{1}{n}\sum_{i=1}^{n}|x_{ij}|}{\sqrt{2\pi}\gamma}\exp\left(-\frac{1}{2}\right)
$$
$$
b_j := \frac{1}{s_0^2}
$$

For the residual variance with $\theta_{p+1} = \log(\sigma)$ we write

$$
\hat{\Lambda}_{p+1}(\theta + \nu \cdot t, \nu) = \Lambda^{(1)}_{p+1}(\theta + \nu \cdot t, \nu) + \Lambda^{(2)}_{p+1}(\theta + \nu \cdot t, \nu) + \Lambda^{(3)}_{p+1}(\theta + \nu \cdot t, \nu)
$$

with $\Lambda^{(1)}_{p+1}(\theta + \nu \cdot t, \nu) = a_0$ and $\Lambda^{(2)}_{p+1}(\theta + \nu \cdot t, \nu) = 2b_0 \exp(-2(\log(\sigma) + \nu_{p+1}t))$ and $\Lambda^{(3)}_{p+1}(\theta + \nu \cdot t, \nu) := \frac{2\omega}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\right) \exp(\log(\sigma) + \nu_{p+1}t)R$, allows us to sample stopping times τ_{p+1} from a Poisson process with rate $\int_0^t \Lambda_{p+1}(\theta + \nu \cdot s, \nu)ds$ via Poisson superposition, first sampling $\tau^{(l)}_{p+1}$ according to a Poisson process with rate $\int_0^t \Lambda^{(l)}_{p+1}(\theta + \nu \cdot s, \nu)ds$, $l = 1, \dots, 3$ and setting $\tau_{p+1} = \min\{\tau^{(1)}_{p+1}, \tau^{(2)}_{p+1}, \tau^{(3)}_{p+1}\}$.

Sampling $\tau_{p+1}^{(1)}$ requires evaluating

$$\int_0^t \Lambda_{p+1}^{(1)}(\theta + \nu \cdot s, \nu) ds = \int_0^t a_0 ds = a_0 t$$

which by Cinlar's Algorithm (Cinlar, 1975) (Algorithm 1) means that for $u^{(1)} \sim Unif(0,1)$

$$\tau_{p+1}^{(1)} := \inf \{t : a_0 t \geq s\} = \frac{-\log(u^{(1)})}{a_0}$$

Sampling $\tau_{p+1}^{(2)}$ requires evaluating

$$\begin{aligned}
\int_0^t \Lambda_{p+1}^{(2)}(\theta + \nu \cdot s, \nu) ds &= \int_0^t 2b_0 \exp(-2(\log(\sigma) + \nu_{p+1} s)) ds \\
&= 2b_0 \left[-\frac{1}{2\nu_{p+1}} \exp(-2(\log(\sigma) + \nu_{p+1} s)) \right]_0^t \\
&= \frac{b_0}{\nu_{p+1}} [\exp(-2\log(\sigma)) - \exp(-2(\log(\sigma) + \nu_{p+1} t))]
\end{aligned}$$

which by Cinlar's Algorithm (Cinlar, 1975) (Algorithm 1) means that for $u^{(2)} \sim Unif(0,1)$

$$\begin{aligned}
\tau_{p+1}^{(2)} &:= \inf \left\{ t : \frac{b_0}{\nu_{p+1}} [\exp(-2\log(\sigma)) - \exp(-2(\log(\sigma) + \nu_{p+1} t))] \geq -\log(u^{(2)}) \right\} \\
&= \inf \left\{ t : [1 - \exp(-2\nu_{p+1} t)] \geq \frac{-\log(u^{(2)}) \nu_{p+1} \exp(2\log(\sigma))}{b_0} \right\} \\
&= -\frac{1}{2\nu_{p+1}} \log \left(1 + \frac{\log(u^{(2)}) \nu_{p+1} \exp(2\log(\sigma))}{b_0} \right)
\end{aligned}$$

for $-\log(u^{(2)}) < \frac{b_0}{\nu_{p+1} \exp(2\log(\sigma))}$ or else $\tau_{p+1}^{(2)} = \infty$.

Lastly, sampling $\tau_{p+1}^{(3)}$ requires evaluating

$$\int_0^t \Lambda_{p+1}^{(3)}(\theta + \nu \cdot s, \nu) ds = \int_0^t \exp(\log(\sigma) + \nu_{p+1} s) \left(\frac{2\omega}{\sqrt{2\pi\gamma}} \exp\left(-\frac{1}{2}\right) R \right) ds$$
$$= \frac{1}{\nu_{p+1}} \left(\frac{2\omega}{\sqrt{2\pi\gamma}} \exp\left(-\frac{1}{2}\right) R \right) [\exp(\log(\sigma) + \nu_{p+1} s)]_0^t$$
$$= \frac{1}{\nu_{p+1}} \left(\frac{2\omega}{\sqrt{2\pi\gamma}} \exp\left(-\frac{1}{2}\right) R \right) [\exp(\log(\sigma) + \nu_{p+1} t) - \exp(\log(\sigma))]$$

which by Cinlar's Algorithm (Cinlar, 1975) (Algorithm 1) means that for $u^{(3)} \sim Unif(0,1)$

$$\tau_{p+1}^{(3)} := \inf \left\{ t : \frac{1}{\nu_{p+1}} \left(\frac{2\omega}{\sqrt{2\pi\gamma}} \exp\left(-\frac{1}{2}\right) R \right) [\exp(\log(\sigma) + \nu_{p+1} t) - \exp(\log(\sigma))] \geq -\log(u^{(3)}) \right\}$$
$$= \inf \left\{ t : [\exp(\nu_{p+1} t) - 1] \geq \frac{-\log(u^{(3)}) \nu_{p+1} \exp(-\log(\sigma))}{\left(\frac{2\omega}{\sqrt{2\pi\gamma}} \exp\left(-\frac{1}{2}\right) R \right)} \right\}$$
$$= \frac{1}{\nu_{p+1}} \log \left(1 - \frac{\log(u^{(3)}) \nu_{p+1} \exp(-\log(\sigma))}{\left(\frac{2n}{\sqrt{2\pi\gamma}} \exp\left(-\frac{1}{2}\right) R \right)} \right)$$

for $-\log(u^{(3)}) > -\frac{\left(\frac{2\omega}{\sqrt{2\pi\gamma}} \exp\left(-\frac{1}{2}\right) R \right)}{\nu_{p+1} \exp(-\log(\sigma))}$ or else $\tau_{p+1}^{(3)} = \infty$. □

2.3 Robustified copula with the MMD

Here we provide implementation details for Section 5.1 of the paper.

We propose a Bayesian extension of Alquier and Gerber (2024) that uses the zig-zag to sample from the MMD-Bayes posterior for copula parameter ρ. We place a beta prior with hyperparameter $a_0 = 1$ and $b_0 = 1$ on $\frac{\rho+1}{2} \in [0,1]$ and repraræmetrise $\rho \mapsto \theta$ with $\rho(\theta) = \frac{2}{1+\exp(-\theta)} - 1$ so that the zig-zag can sample from the implied posterior of $\theta \in \mathbb{R}$. Lemma 3 provides computational bonds for this example and Corollary 3 shows how sampling from these is implemented.

Lemma 3 (Computational Bounds for MMD-Bayes for Gaussian Copula model)**.** *Consider the MMD-Bayes posterior for a bivariate Gaussian copula model with* $Beta(a_0, b_0)$ *prior on*

$\frac{\rho+1}{2}$ *where* ρ *is the unknown correlation. We repraramétrise* $\rho \mapsto \theta$ *with* $\rho(\theta) = \frac{2}{1+\exp(-\theta)} - 1$ *so that the zig-zag can sample from* $\theta \in \mathbb{R}^1$*. The zig-zag switching rates are given by*

$$
\begin{aligned}
&\hat{\lambda}(\theta, \nu) := \max\Bigg(0, \nu b_0 - \nu(a_0 + b_0)\frac{\exp(-\theta)}{(1+\exp(-\theta))} \\
&\Bigg\{-\frac{2w\nu}{mn}\sum_{i=1}^{n}\sum_{k=1}^{b}\frac{(\Phi^{-1}(\hat{u}_{2i}) - G_\theta^{(2)}(v_k))}{2\pi\gamma^2} \\
&\exp\left(-\frac{(\Phi^{-1}(\hat{u}_{1i}) - v_{k1})^2 + (\Phi^{-1}(\hat{u}_{2i}) - (\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2}))^2}{2\gamma}\right)\left(v_{k1} - \frac{\rho(\theta)}{\sqrt{1-\rho(\theta)^2}}v_{k2}\right) \\
&-\frac{\nu w}{m(m-1)}\sum_{k=1}^{b}\sum_{k'\neq k}^{m}\frac{((\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2}) - (\rho(\theta)v_{k'1} + \sqrt{1-\rho(\theta)^2}v_{k'2}))}{2\pi\gamma^2} \\
&\exp\left(-\frac{(v_{k1} - v_{k'1})^2 + ((\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2}) - (\rho(\theta)v_{k'1} + \sqrt{1-\rho(\theta)^2}v_{k'2}))^2}{2\gamma}\right) \\
&\cdot\left(\left(v_{k1} - \frac{\rho(\theta)}{\sqrt{1-\rho(\theta)^2}}v_{k2}\right) - \left(v_{k'1} - \frac{\rho(\theta)}{\sqrt{1-\rho(\theta)^2}}v_{k'2}\right)\right)\Bigg\}\frac{2\exp(-\theta)}{(1+\exp(-\theta))^2}\Bigg)
\end{aligned}
$$

where $v_1, \ldots, v_m \sim p_v$ *with* $p_v(v) = \mathcal{N}(v_1; 0, 1)\mathcal{N}(v_2; 0, 1)$.

Computational bounds are given by

$$
\hat{\Lambda}_j(\theta + \nu \cdot t, \nu) = \max\{a_0, b_0\} + \frac{4\omega}{(2\pi\gamma)^{3/2}}R\exp\left(-\frac{1}{2}\right)\left(2 + \exp\left(-\frac{1}{2}\theta - \frac{1}{2}\nu \cdot t\right)\right)
$$

where R *is an assumed upper bound on* $|v|$ *for* $v \sim \mathcal{N}(0, 1)$.

The bounds in Lemma 3 are slightly different from those in Corollary 4 of the main paper. While Corollary 4 proves that an upper bound exists, obtaining those bounds requires solving an intractable optimisation. Lemma 3 is therefore an approximation of Corollary 4. Although v has unbounded support Corollary 4 guarantees that as long as we take R big enough this will in fact be an upper bound, i.e. looser than the bounds in Corollary 4. We find however setting $R = 3$ provides a bound that is never exceeded. We keep track of the maximum value of $\hat{\lambda}(\theta + \nu \cdot t, \nu)/\Lambda(\theta + \nu \cdot t, \nu)$ throughout the sampling and check that this is always less than 1.

Proof. Firstly, we can indirectly sample $u_{1:b} \sim p_\theta(u)$ where $p_\theta(u)$ is the density of a bivariate Gaussian copula with parameter ρ by first $v_{1:b}$ with $v_k = \{v_{k1}, v_{k2}\} \sim p_v$ and $p_v(v) = \mathcal{N}(v_1; 0, 1)\mathcal{N}(v_2; 0, 1)$ and setting $u_k = \{v_{k1}, v_{k2}\} = \{G_\theta^{(1)}(v_k), G_\theta^{(2)}(v_k)\} = F(v_k; \theta)$ with

$$G_{\theta,1}(v_k) = \Phi(v_1)$$

$$G_{\theta,2}(v_k) = \Phi(\rho v_1 + \sqrt{1-\rho^2} v_2).$$

As a result, for the MMD-Bayes posterior

$$\Psi_n(\theta) := -\log \pi(\theta) + w\mathsf{L}_n^{\mathcal{K}_\gamma}(\theta)$$

$$= b_0\theta + (a_0 + b_0)\log\left(1 + \exp(-\theta)\right)$$

$$- \frac{\omega}{n}\sum_{i=1}^{n} 2\mathbb{E}_{v\sim p_0(\cdot)}\left[\mathcal{K}_\gamma(\hat{u}_i, G_\theta(v))\right] + \omega\mathbb{E}_{v,v'\sim p_0(\cdot)}\left[\mathcal{K}_\gamma(G_\theta(v), G_\theta(v'))\right]$$

$$= b_0\theta + (a_0 + b_0)\log\left(1 + \exp(-\theta)\right)$$

$$- \frac{\omega}{n}\sum_{i=1}^{n} 2\mathbb{E}_{v\sim p(\cdot)}\left[\frac{1}{2\pi\gamma}\exp\left(\frac{(\Phi^{-1}(\hat{u}_{i1}) - \Phi^{-1}(G_{\theta,1}(v)))^2 + (\Phi^{-1}(\hat{u}_{i2}) - \Phi^{-1}(G_{\theta,2}(v)))^2}{2\gamma}\right)\right]$$

$$+ \omega\mathbb{E}_{v,v'\sim p(\cdot)}\left[\frac{1}{2\pi\gamma}\exp\left(\frac{(\Phi^{-1}(G_{\theta,1}(v)) - \Phi^{-1}(G_{\theta,1}(v')))^2 + (\Phi^{-1}(G_{\theta,2}(v)) - \Phi^{-1}(G_{\theta,2}(v')))^2}{2\gamma}\right)\right]$$

$$= b_0\theta + (a_0 + b_0)\log\left(1 + \exp(-\theta)\right)$$

$$- \frac{\omega}{n}\sum_{i=1}^{n} 2\mathbb{E}_{v\sim p(\cdot)}\left[\frac{1}{2\pi\gamma}\exp\left(\frac{(\Phi^{-1}(\hat{u}_{i1}) - v_1)^2 + (\Phi^{-1}(\hat{u}_{i2}) - v_2)^2}{2\gamma}\right)\right]$$

$$+ \omega\mathbb{E}_{v,v'\sim p(\cdot)}\left[\frac{1}{2\pi\gamma}\exp\left(\frac{(\Phi^{-1}(v_1 - v_1')^2 + (v_2 - v_2')^2}{2\gamma}\right)\right]$$

The zig-zag requires the evaluation of

$$\hat{\lambda}(\theta, \nu) = \max\left(0, \nu\left(-\frac{\partial}{\partial\theta}\log(\pi(\theta)) + \omega\varphi_{b,n}(\theta)\right)\right)$$

and bounding of $\hat{\lambda}(\mu + \nu \cdot t, \nu)$, where

$$\varphi_{b,n}(\theta) = -\frac{1}{n}\sum_{i=1}^{n} 2\frac{1}{b}\sum_{k=1}^{b}\frac{1}{2\pi\gamma^2}\exp\left(-\frac{(\Phi^{-1}(\hat{u}_{1i}) - v_{k1})^2 + (\Phi^{-1}(\hat{u}_{2i}) - (\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2}))^2}{2\gamma}\right)$$

$$\cdot \{(\Phi^{-1}(\hat{u}_{1i}) - v_{k1})\frac{\partial}{\partial\theta}v_{k1} + (\Phi^{-1}(\hat{u}_{2i}) - (\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2}))\frac{\partial}{\partial\theta}(\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2}))\}$$

$$+ \frac{1}{m(m-1)}\sum_{k=1}^{b}\sum_{k'\neq k}^{m} -\frac{1}{2\pi\gamma^2}\exp\left(-\frac{(v_{k1} - v_{k'1})^2 + ((\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2}) - (\rho(\theta)v_{k'1} + \sqrt{1-\rho(\theta)^2}v_{}}{2\gamma}\right.$$

$$\cdot \left\{(v_{k1} - v_{k'1})(\frac{\partial}{\partial\theta}v_{k1} - \frac{\partial}{\partial\theta}v_{k'1})\right.$$

$$+((\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2}) - (\rho(\theta)v_{k'1} + \sqrt{1-\rho(\theta)^2}v_{k'2}))(\frac{\partial}{\partial\theta}(\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2})) - \frac{\partial}{\partial\theta}(\rho(\theta)v_{k'1}$$

$$= -\frac{1}{n}\sum_{i=1}^{n} 2\frac{1}{b}\sum_{k=1}^{b}\frac{1}{2\pi\gamma^2}\exp\left(-\frac{(\Phi^{-1}(\hat{u}_{1i}) - v_{k1})^2 + (\Phi^{-1}(\hat{u}_{2i}) - (\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2}))^2}{2\gamma}\right)$$

$$\cdot \{(\Phi^{-1}(\hat{u}_{2i}) - (\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2}))\}\left(v_{k1} - \frac{\rho(\theta)}{\sqrt{1-\rho(\theta)^2}}v_{k2})\right)\frac{2\exp(-\theta)}{(1+\exp(-\theta))^2}$$

$$- \frac{1}{m(m-1)}\frac{\omega}{n}\sum_{k=1}^{b}\sum_{k'\neq k}^{m}\frac{1}{2\pi\gamma^2}\exp\left(-\frac{(v_{k1} - v_{k'1})^2 + ((\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2}) - (\rho(\theta)v_{k'1} + \sqrt{1-\rho(\theta)^2}v_{}}{2\gamma}\right.$$

$$\cdot ((\rho(\theta)v_{k1} + \sqrt{1-\rho(\theta)^2}v_{k2}) - (\rho(\theta)v_{k'1} + \sqrt{1-\rho(\theta)^2}v_{k'2}))$$

$$\cdot \left(\left(v_{k1} - \frac{\rho(\theta)}{\sqrt{1-\rho(\theta)^2}}v_{k2})\right) - \left(v_{k'1} - \frac{\rho(\theta)}{\sqrt{1-\rho(\theta)^2}}v_{k'2})\right)\right)\frac{2\exp(-\theta)}{(1+\exp(-\theta))^2}$$

with $v_1, \ldots, v_b \sim p_v$. This uses the fact that

$$\rho(\theta) = \frac{2}{1+\exp(-\theta)} - 1$$
$$\partial\rho(\theta) = \frac{2\exp(-\theta)}{(1+\exp(-\theta))^2}$$

While Corollary 4 of the main paper proves that a finite bound on this exists, the bounds derived in Corollary 4 are intractable to use in practice. Instead, we use the fact that

$$\left|\frac{(y-y')}{\sqrt{2\pi}\gamma^{3/2}}\exp\left(-\frac{(y-y')^2}{2\gamma}\right)\right| \leq \frac{1}{\sqrt{2\pi\gamma}}\exp\left(-\frac{1}{2}\right)$$
$$\left|\frac{1}{\sqrt{2\pi\gamma}}\exp\left(-\frac{(y-y')^2}{2\gamma}\right)\right| \leq \frac{1}{\sqrt{2\pi\gamma}}$$

for any value of y or y', and

$$\left|\left(v_1 - \frac{\rho(\theta)}{\sqrt{1-\rho(\theta)^2}}v_2)\right)\frac{2\exp(-\theta)}{(1+\exp(-\theta))^2}\right| \leq \frac{2}{\sqrt{2\pi}}v_1 + \frac{1}{\sqrt{2\pi}}v_2$$

for all v, $\theta \in \Theta$, and use the following *approximate* bound

$$\begin{aligned}
\hat{\lambda}(\theta+\nu\cdot t,\nu) :&= \max\left(0, \nu\left(-\frac{\partial}{\partial\theta}\log(\pi(\theta+\nu\cdot t)) + \omega\varphi_{b,n}(\theta+\nu\cdot t)\right)\right) \\
&\leq \max\{a_0, b_0\} + \frac{12\omega}{(2\pi\gamma)^{3/2}}R\exp\left(-\frac{1}{2}\right) \\
&= \Lambda(\theta+\nu\cdot t,\nu)
\end{aligned}$$

where R is an assumed upper bound on $|v|$ for $v \sim \mathcal{N}(0,1)$. □

Corollary 3. *Sampling stopping times according to the computational bound in Lemma 3 is achieved by setting*

$$\tau^* = \frac{-\log(u)}{a}$$

with $a = \max\{a_0, b_0\} + \frac{12\omega}{(2\pi\gamma)^{3/2}}R\exp\left(-\frac{1}{2}\right)$ *and* $u \sim U[0,1]$.

Proof. Sampling τ^* requires evaluating

$$\int_0^t \hat{\Lambda}(\theta+\nu\cdot s,\nu)ds = \int_0^t a\,ds = at$$

where $a = \max\{a_0, b_0\} + \frac{12\omega}{(2\pi\gamma)^{3/2}} R \exp\left(-\frac{1}{2}\right)$, which by Cinlar's Algorithm (Cinlar, 1975) (Algorithm 1) means that for $u^{(1)} \sim Unif(0,1)$

$$\tau^* := \inf\{t : at \geq s\} = \frac{-\log(u)}{a}.$$

$\square$

2.4 Robustified Poisson regression with the β-Divergence

Here we provide implementation details for Section 5.2.2 of the paper.

In Poisson regression the mass function of count data y, conditional on predictor variables x is given by $p_\theta(y; x) = e^{yx^\top\theta} e^{-e^{x^\top\theta}}/y!$. The βD-loss for such a model involves the infinite sum $\sum_{u=0}^{\infty} p_\theta(u; x)^{\beta+1}$ which for $\beta > 0$ is not available in closed form. We therefore use the zig-zag to sample from the βD-Bayes posterior for θ. We place Gaussian priors with mean $\mu_0 = 0$ and variance $s_0^2 = 1$ on each θ_j, $j = 1, \ldots, d$. Lemma 4 provides computational bonds for this example and Corollary 4 shows how sampling from these is implemented.

Lemma 4 (Computational Bounds for βD-Bayes for Poisson regression model). *Consider the βD-Bayes posterior for a Poisson regression model with $\mathcal{N}(\mu_0, s_0^2)$ priors on unknown regression coefficients $\theta = \{\theta_1, \ldots, \theta_d\}$. The zig-zag switching rates are given by*

$$\hat{\lambda}_j(\theta, \nu) := \max\left(0, \nu_j \frac{\theta_j - \mu_0}{s_0^2} + \nu_j \frac{\omega(\beta+1)}{n} \sum_{i=1}^{n} \left\{ \frac{1}{b} \sum_{k=1}^{b} p_\theta(u_{ik}; x_i)^\beta (u_{ik} - e^{x_i^\top(\theta)}) x_{ij} \right.\right.$$
$$\left.\left. - p_\theta(y_i; x_i)^\beta \left(y_i - e^{x_i^\top(\theta + \nu \cdot t)}\right) x_{ij} \right\}\right),$$

$j = 1, \ldots, d$ where $u_{i1}, \ldots, u_{ik} \sim p_\theta(\cdot; x_i)$ $i = 1, \ldots, n$ with $p_\theta(y; x) = \frac{e^{yx^\top\theta} e^{-e^{x^\top\theta}}}{y!}$.

Computational bounds are given by

$$\hat{\Lambda}_j(\theta+\nu\cdot t,\nu) := \left(\frac{|\theta_j-\mu_0|}{s_0^2}+\frac{\omega(\beta+1)}{n}\sum_{i=1}^n |x_{ij}|\left(y_i+\frac{1}{\beta}\right)\right)+\frac{1}{s_0^2}t$$
$$+\, e^{\max_{i=1,\ldots,n}\{\sum_{j=1}^p |x_{ij}|\}t}\frac{\omega(\beta+1)^2}{n\beta}\sum_{i=1}^n |x_{ij}|e^{x_i^\top\theta}.$$

Proof. As the βD-loss requires evaluation of models density, in any case, we *directly* sample from the poison regression model in order to estimate $\mathsf{L}_n^\beta(\theta)$.

As a result, for the βD-Bayes posterior

$$\begin{aligned}\Psi_n(\theta) &:= -\sum_{j=1}^p \log\pi(\theta_j)+w\mathsf{L}_n^\beta(\theta)\\ &= -\sum_{j=1}^p \log\pi(\theta_j)+\frac{\omega}{n}\sum_{i=1}^n\left\{\mathbb{E}_{u\sim p_\theta(\cdot;x_i)}\left[p_\theta(u;x_i)^\beta\right]-\frac{\beta+1}{\beta}p_\theta(y_i;x_i)^\beta\right\}\end{aligned}$$

The zig-zag requires the evaluation of

$$\hat{\lambda}_j(\theta,\nu) := \max\left(0,\nu_j\left(-\frac{\partial}{\partial\theta_j}\log(\pi(\theta))+\omega[\varphi_{b,n}(\theta+\nu\cdot t)]_j\right)\right),$$

and bounding of $\hat{\lambda}_j(\theta+\nu\cdot t,\nu)$ for $j=1,\ldots p$. The log-derivative trick provides that

$$[\varphi_{b,n}(\theta)]_j := \frac{(\beta+1)}{n}\sum_{i=1}^n\left\{\frac{1}{b}\sum_{k=1}^b p_{\theta+\nu\cdot t}(u_{ik};x_i)^\beta\nabla_{\theta_j}\log p_{\theta+\nu\cdot t}(u_{ik};x_i)\right.$$
$$\left.-p_{\theta+\nu\cdot t}(y_i;x_i)^{\beta-1}\nabla_{\theta_j}p_{\theta+\nu\cdot t}(y_i;x_i)\right\}$$

for $u_i = \{u_{i1}, \ldots, u_{ib}\}$ with $u_{ik} \sim p_\theta(\cdot; x_i)$, $i = 1, \ldots, n$, $k = 1, \ldots, b$. Next, we can use that

$$\nabla_{\theta_j} p_\theta(y; x) = \left(y - e^{x^\top \theta}\right) x_j p_\theta(y; x)$$
$$\nabla_{\theta_j} \log p_\theta(y; x) = (y - e^{x^\top \theta}) x_j$$

to rewrite

$$\varphi_{b,n}(\theta) = \frac{(\beta+1)}{n} \sum_{i=1}^{n} \left\{ \frac{1}{b} \sum_{k=1}^{b} p_{\theta+\nu\cdot t}(u_{ik}; x_i)^\beta (u_{ik} - e^{x_i^\top(\theta)}) x_{ij} \right.$$
$$\left. - p_{\theta+\nu\cdot t}(y_i; x_i)^\beta \left(y_i - e^{x_i^\top(\theta)}\right) x_{ij} \right\}$$

Then, using the fact that $p_{\theta+\nu\cdot t}(u_{ik}; x_i)^\beta u_{ik} < \frac{\exp(x_i^\top \theta)+1}{\beta}$ and that p is a probability mass function, so $p_\theta(\cdot, x) \leq 1$, and only has support on the positive integers, we can write

$$\begin{aligned}
\nu_j[\varphi_{b,n}(\theta + \nu \cdot t)]_j &\leq \frac{(\beta+1)}{n} \sum_{i=1}^{n} |x_{ij}| \left(\frac{\exp(x_i^\top(\theta + \nu \cdot t)) + 1}{\beta} \right. \\
&\quad \left. + e^{x_i^\top(\theta+\nu\cdot t)} + e^{x_i^\top(\theta+\nu\cdot t)} + y_i \right) \\
&\leq \frac{(\beta+1)}{n} \sum_{i=1}^{n} |x_{ij}| \left(y_i + \frac{1}{\beta}\right) + \frac{(\beta+1)}{n} \sum_{i=1}^{n} |x_{ij}| \frac{\beta+1}{\beta} e^{x_i^\top \theta} e^{x_i^\top \nu\cdot t} \\
&\leq \frac{(\beta+1)}{n} \sum_{i=1}^{n} |x_{ij}| \left(y_i + \frac{1}{\beta}\right) + e^{\max_{i=1,\ldots,n}\{\sum_{j=1}^{p} |x_{ij}|\} t} \frac{(\beta+1)^2}{\beta} \frac{1}{n} \sum_{i=1}^{n} |x_{ij}| e^{x_i^\top \theta}.
\end{aligned}$$

and therefore

$$\hat{\lambda}_{m,j}(\theta + \nu \cdot t, \nu) := \max\left(0, \nu_j \left(-\frac{\partial}{\partial \theta_j} \log(\pi(\theta + \nu \cdot t)) + \omega[\varphi_{m,n}(\theta + \nu \cdot t)]_j\right)\right)$$
$$\leq \left(\frac{|\theta_j - \mu_0|}{s_0^2} + \frac{\omega(\beta+1)}{n} \sum_{i=1}^{n} |x_{ij}| \left(y_i + \frac{1}{\beta}\right)\right) + \frac{1}{s_0^2} t + e^{\max_i\{\sum_{j=1}^{p} |x_{ij}|\} t} \frac{(\beta+1)^2}{\beta} \frac{\omega}{n} \sum_{i=1}^{n} |x_{ij}| e^{x_i^\top \theta}$$
$$= \hat{\Lambda}_j(\theta + \nu \cdot t, \nu).$$

$\square$

Corollary 4. *Sampling stopping times according to the computational bound in Lemma 4 is achieved for $j = 1, \dots, d$ by setting*

$$\tau_j^* = \min\{\tau_j^{(1)}, \tau_j^{(2)}\}$$

with

$$\begin{aligned}
\tau_j^{(1)} &= \frac{\sqrt{a_j^2 - 2\log(u_j^{(1)})b_j} - a_j}{b_j} \\
\tau_j^{(2)} &= \frac{1}{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}} \log\left(1 - \log(u_j^{(2)}) \frac{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}}{\frac{(\beta+1)^2}{\beta}\frac{\omega}{n}\sum_{i=1}^n |x_{ij}|e^{x_i^\top\theta}}\right) \\
a_j &= \left(\frac{|\theta_j - \mu_0|}{s_0^2} + \frac{\omega(\beta+1)}{n}\sum_{i=1}^n |x_{ij}|\left(y_i + \frac{1}{\beta}\right)\right) \\
b_j &= \frac{1}{s_0^2}.
\end{aligned}$$

with $u_j^{(1)}, u_j^{(2)} \sim Unif(0,1)$.

Proof. Writing

$$\hat{\Lambda}_j(\theta + \nu \cdot t, \nu) = \Lambda_j^{(1)}(\theta + \nu \cdot t, \nu) + \Lambda_j^{(2)}(\theta + \nu \cdot t, \nu)$$

with $\Lambda_j^{(1)}(\theta + \nu \cdot t, \nu) = \left(\frac{|\theta_j - \mu_0|}{s_0^2} + \frac{\omega(\beta+1)}{n}\sum_{i=1}^n |x_{ij}|\left(y_i + \frac{1}{\beta}\right)\right) + \frac{1}{s_0^2}t$ and $\Lambda_j^{(2)}(\theta + \nu \cdot t, \nu) = e^{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}t}\frac{(\beta+1)^2}{\beta}\frac{\omega}{n}\sum_{i=1}^n |x_{ij}|e^{x_i^\top\theta}$, allows us to sample stopping times τ_j from a Poisson process with rate $\int_0^t \hat{\Lambda}_j(\theta + \nu \cdot s, \nu)ds$ as required by Lemma 4 via Poisson superposition, first sampling $\tau_j^{(l)}$ according to a Poisson process with rate $\int_0^t \Lambda_j^{(l)}(\theta + \nu \cdot s, \nu)ds$, $l = 1, \dots, 2$ and setting $\tau_j = \min\{\tau_j^{(1)}, \tau_j^{(2)}\}$.

Sampling $\tau_j^{(1)}$ for $j = 1, \dots, d$ can be done following the procedure outlined in Corollary

1 with

$$a_j := \left(\frac{\theta_j - \mu_0}{s_0^2} + \frac{\omega(\beta+1)}{n} \sum_{i=1}^{n} |x_{ij}| \left(y_i + \frac{1}{\beta} \right) \right)$$
$$b_j := \frac{1}{s_0^2}.$$

Sampling $\tau_j^{(2)}$ requires evaluating

$$\begin{aligned}
\int_0^t \Lambda_{p+1}^{(2)}(s;\theta,\nu)ds &= \int_0^t e^{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}s} \frac{(\beta+1)^2}{\beta} \frac{\omega}{n} \sum_{i=1}^{n} |x_{ij}| e^{x_i^\top \theta} ds \\
&= \frac{(\beta+1)^2}{\beta} \frac{\omega}{n} \sum_{i=1}^{n} |x_{ij}| e^{x_i^\top \theta} \left[\frac{}{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}} e^{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}s} \right]_0^t \\
&= \frac{(\beta+1)^2}{\beta} \frac{\omega}{n} \frac{\sum_{i=1}^{n} |x_{ij}| e^{x_i^\top \theta}}{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}} \left[e^{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}t} - 1 \right]
\end{aligned}$$

and therefore the stopping times can be simulated following Cinlar's Algorithm (Cinlar, 1975) (Algorithm 1) by sampling $u_j^{(2)} \sim Unif(0,1)$ and solving

$$\begin{aligned}
\tau_j^{(2)} &:= \inf \left\{ t : \frac{(\beta+1)^2}{\beta} \frac{\omega}{n} \frac{\sum_{i=1}^{n} |x_{ij}| e^{x_i^\top \theta}}{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}} \left[e^{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}t} - 1 \right] \geq -\log(u_j^{(2)}) \right\} \\
&= \inf \left\{ t : \left[e^{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}t} - 1 \right] \geq -\log(u_j^{(2)}) \frac{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}}{\frac{(\beta+1)^2}{\beta} \frac{\omega}{n} \sum_{i=1}^{n} |x_{ij}| e^{x_i^\top \theta}} \right\} \\
&= \frac{1}{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}} \log \left(1 - \log(u_j^{(2)}) \frac{\max_{i=1,\dots,n}\{\sum_{j=1}^p |x_{ij}|\}}{\frac{(\beta+1)^2}{\beta} \frac{\omega}{n} \sum_{i=1}^{n} |x_{ij}| e^{x_i^\top \theta}} \right).
\end{aligned}$$

□

References

Alquier, P. and M. Gerber (2024). Universal robust regression via maximum mean discrepancy. *Biometrika 111*(1), 71–92. 7, 15

Cinlar, E. (1975). *Introduction to stochastic processes*. Englewood Cliffs, NJ: Prentice-Hall.

1, 6, 14, 15, 20, 24