

## Stability of doubly-intractable distributions

Michael Habeck\*     Daniel Rudolf†     Björn Sprungk‡

### Abstract

Doubly-intractable distributions appear naturally as posterior distributions in Bayesian inference frameworks whenever the likelihood contains a normalizing function  $Z$ . Having two such functions  $Z$  and  $\tilde{Z}$  we provide estimates of the total variation and 1-Wasserstein distance of the resulting posterior probability measures. As a consequence this leads to local Lipschitz continuity w.r.t.  $Z$ . In the more general framework of a random function  $\tilde{Z}$  we derive bounds on the expected total variation and expected 1-Wasserstein distance. The applicability of the estimates is illustrated within the setting of two representative Monte Carlo recovery scenarios.

**Keywords:** doubly-intractable; Wasserstein distance; stability.

**MSC2020 subject classifications:** 60B10; 62C10; 65C05.

Submitted to ECP on May 3, 2020, final version accepted on August 12, 2020.

Supersedes arXiv:2004.07310v3.

## 1 Introduction

Suppose that  $(\Theta, d)$  is a complete and separable metric space equipped with its Borel  $\sigma$ -algebra  $\mathcal{B}(\Theta)$  and a  $\sigma$ -finite reference measure  $\mu$ . Assume that there are measurable functions  $\Phi: \Theta \rightarrow (-\infty, \infty)$  and  $Z: \Theta \rightarrow (0, \infty)$  with

$$C_Z := \int_{\Theta} \frac{\exp(-\Phi(\theta))}{Z(\theta)} \mu(d\theta) \in (0, \infty),$$

such that

$$\pi_Z(A) := \frac{1}{C_Z} \int_A \frac{\exp(-\Phi(\theta))}{Z(\theta)} \mu(d\theta), \quad A \in \mathcal{B}(\Theta), \quad (1.1)$$

defines a probability measure on  $(\Theta, \mathcal{B}(\Theta))$ . We are interested in stability properties of  $\pi_Z$  w.r.t. the function  $Z$ . Given another measurable function  $\tilde{Z}: \Theta \rightarrow (0, \infty)$  which is somehow close to  $Z$ , we ask whether  $\pi_Z$  and  $\pi_{\tilde{Z}}$  are also close to each other.

This question is motivated by applications within Bayesian inference, where such type of distributions appear, see for example [4, 5] as well as [17] and the references therein. The interpretation is as follows: Think of  $\theta \mapsto \exp(-\Phi(\theta))/Z(\theta)$  as a likelihood function which contains an unknown normalizing function  $Z$ . One is interested on

---

\*Friedrich-Schiller-Universität Jena, Max-Planck-Institut für biophysikalische Chemie, Arbeitsgruppe Mikroskopische Bildanalyse, Kollegiengasse 10, 07743 Jena, Germany. E-mail: michael.habeck@uni-jena.de

†Institute for Mathematical Stochastics, University of Goettingen, Goldschmidtstraße 7, 37077 Göttingen, Germany. E-mail: daniel.rudolf@uni-goettingen.de

‡Faculty of Mathematics and Computer Science, Technische Universität Bergakademie Freiberg, 09596 Freiberg, Germany. E-mail: bjoern.sprungk@math.tu-freiberg.de

sampling w.r.t. a posterior distribution based on the partially unknown likelihood function. Unknown here in the sense that it is infeasible to evaluate  $Z$  exactly. This and the fact that (1.1) itself contains an unknown normalizing constant  $C_Z$  is the reason for calling  $\pi_Z$  doubly-intractable. A recent survey for approximate sampling of such doubly-intractable distributions is given in [17]. We provide a motivating example for such scenarios.

**Example 1.1** (Gibbs distribution as likelihood). A Gibbs distribution on a finite state space  $G$  is determined by a probability mass function  $\rho(\cdot | \beta)$  with inverse temperature parameter  $\beta > 0$  given by

$$\rho(x | \beta) = \exp(-\beta H(x))/Z(\beta), \quad x \in G,$$

where  $H: G \rightarrow [0, \infty)$  is called Hamiltonian and  $Z(\beta) = \sum_{x \in G} \exp(-\beta H(x))$  partition function. Suppose that there is observational data  $x_{\text{obs}} \in G$  available as a realization of the Gibbs distribution but with unknown  $\beta$ . We then aim to gain knowledge of  $\beta$  through the realization of  $x_{\text{obs}}$ . In a Bayesian framework this leads to a posterior distribution of the form (1.1) with  $\Theta = (0, \infty)$  and  $\Phi(\beta) = \beta H(x_{\text{obs}})$  and  $Z(\beta)$ . Note that the Ising model fits into this framework: Let  $(V, E)$  be a graph with (non-empty) vertex set  $E$ , edge set  $V \subset E \times E$  and  $G = \{-1, 1\}^E$  with  $H(x) = -\sum_{(e, e') \in V} x(e)x(e')$ .

Having such an example in mind it is reasonable to recover  $Z$  by an approximation  $\tilde{Z}$  and to gain knowledge of the posterior distribution by sampling w.r.t.  $\pi_{\tilde{Z}}$  (which is hopefully close to  $\pi_Z$ ). A theoretical justification of that approach requires a stability investigation of  $\pi_Z$  w.r.t.  $Z$ .

For quantifying stability properties of probability measures we need to introduce how we want to measure the difference of distributions. For this we use either the total variation or the 1-Wasserstein distance based on the metric  $d$  on  $(\Theta, \mathcal{B}(\Theta))$ . Let us mention here that we only consider the 1-Wasserstein distance and therefore will simply refer to it as the Wasserstein distance. The main results of this note, stated and proven in Section 2, are upper bounds on the difference of  $\pi_Z$  and  $\pi_{\tilde{Z}}$  in terms of  $Z$  and  $\tilde{Z}$  with respect to that distances. We use the following notation. For a measure  $\nu$  on  $(\Theta, \mathcal{B}(\Theta))$  we denote the  $L^p(\nu)$ -norm, for  $p \geq 1$ , by  $\|\cdot\|_{\nu, p}$ , that is, for measurable  $f: \Theta \rightarrow \mathbb{R}$  we have  $\|f\|_{\nu, p} := (\int_{\Theta} |f(\theta)|^p \nu(d\theta))^{1/p}$ . In Theorem 2.1 and Theorem 2.8 we prove that the total variation and the Wasserstein distance (w.r.t.  $d$ ) of  $\pi_Z$  and  $\pi_{\tilde{Z}}$  is smaller than a constant times

$$\left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_Z, 2}.$$

For the total variation distance the result holds also with the  $L^1(\pi_Z)$ -norm instead of the  $L^2(\pi_Z)$ -norm on the right-hand side. In addition to that we provide a number of consequences under some further regularity conditions. For example, if  $\|\exp(-\Phi)/Z\|_{\mu, p} \leq K$  and  $\inf_{\theta \in \Theta} \tilde{Z}(\theta) \geq \ell$  for some  $p \in [1, \infty]$ , some  $K < \infty$  and  $\ell > 0$ , then

$$\frac{2K}{\ell C_Z} \left\| Z - \tilde{Z} \right\|_{\mu, p/(p-1)} \tag{1.2}$$

is again an upper bound of the total variation distance. Under some additional moment conditions on  $\mu$  and  $\|\exp(-\Phi)/\tilde{Z}\|_{\mu, p} \leq K$  a similar estimate is verified for the Wasserstein distance, see Corollary 2.10. Note that from (1.2) one can conclude a local Lipschitz continuity of the mapping  $Z \mapsto \pi_Z$  from a subset of  $L_{\mu, p/(p-1)}$ -functions to the set of probability measures on  $(\Theta, \mathcal{B}(\Theta))$ .

If  $Z$  takes the role of a normalizing constant, as in Example 1.1 above, an approximation  $\tilde{Z}$  of  $Z$  by numerical integration is natural. For this purpose Monte Carlo integration or Monte Carlo estimators, respectively, of  $Z$  are a common choice. Since Monte Carlo methods yield random approximations  $\tilde{Z}$ , we also conduct a stability analysis allowing

for randomized recovery algorithms of  $Z$  leading to random probability measures  $\pi_{\tilde{Z}}$ . In that randomized scenario we provide estimates of the expected total variation, see Corollary 2.11, and the expected Wasserstein distance, see Theorem 2.13. The total variation result reads as follows

$$\mathbb{E}\|\pi_Z - \pi_{\tilde{Z}(\cdot)}\|_{\text{tv}} \leq 2 \int_{\Theta} \left( \mathbb{E} \left| \frac{\tilde{Z}(\cdot, \theta)}{Z(\theta)} - 1 \right|^2 \right)^{1/2} \left( \mathbb{E} \left[ \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} \right]^2 \right)^{1/2} \pi_Z(d\theta).$$

Thus, the upper estimate depends on an averaged relative second moment difference of  $\tilde{Z}$  and  $Z$ . We apply our randomized stability results to Monte Carlo approximations of  $Z$  in two particular examples including the Gibbs distribution of Example 1.1.

In the following we discuss how our results fit into the literature. The study of stability properties w.r.t. posterior distributions in Bayesian inference attracted in recent years considerable attention, see e.g. [3, 9, 19, 20]. In the work [19] local Lipschitz continuity for bounded likelihood functions  $\theta \mapsto \exp(-\Phi(\theta))/Z(\theta)$  has been investigated and in [9] continuity results of posterior distributions w.r.t. perturbations within the observed data are proven. In contrast to [19] a consequence of our main estimate is local Lipschitz continuity also for possibly unbounded likelihood functions and in contrast to the continuity study in [9] we focus on quantitative rather than qualitative results. In Bayesian statistics a number of Markov chain approaches have been developed for approximate sampling of  $\pi_Z$ , for a comprehensive review we refer to [17]. One can distinguish two different types of approaches. The exact one, see [13, 15], where a Markov chain with limit distribution  $\pi_Z$  is constructed and the inexact one, where whenever a function evaluation of  $Z$  is needed, an approximation of it is used. The inexact setting leads to Markov chains which not necessarily target  $\pi_Z$ , but another distribution that is (hopefully) close to  $\pi_Z$ . In particular, the noisy Markov chain approach, for example investigated in [1, 14, 18] falls into this category. In addition to that also adaptive Markov chain Monte Carlo approaches have been developed, see [2, 10]. The inexact setting is closely related to our work, since there for a given  $\theta \in \Theta$  Monte Carlo approximations of  $Z(\theta)$  are employed. A similar estimate as the one w.r.t. the expected total variation distance follows from [11, Theorem 3.2] and [12, Theorem 1] under slightly different boundedness assumptions on the (random) likelihood. However, the estimate of the expected Wasserstein distance seems to be new.

The outline of our work is as follows: In the next section we state and prove our stability results. We introduce the total variation and Wasserstein distance as well as defining related quantities which we need for the formulation of our results. Furthermore, for a random function  $\tilde{Z}$  we provide estimates on the expected total variation and Wasserstein distance. Finally, we illustrate our bounds in a simple Monte Carlo recovery scenario and in a Gibbs distribution posterior setting. In the latter we use a multiple importance sampling approach.

## 2 Stability results

First, we derive bounds of  $\pi_Z$  and  $\pi_{\tilde{Z}}$  in the total variation and the Wasserstein distance. After that we state stability results for the more general case of  $\tilde{Z}(\theta)$  being a random variable for any  $\theta \in \Theta$ .

### 2.1 Total variation distance

Given two probability measures  $\nu_1, \nu_2$  on  $(\Theta, \mathcal{B}(\Theta))$  we define the total variation distance of  $\nu_1$  and  $\nu_2$  by

$$\|\nu_1 - \nu_2\|_{\text{tv}} := \sup_{|f|_{\infty} \leq 1} |\mathbb{E}_{\nu_1}(f) - \mathbb{E}_{\nu_2}(f)|$$

where  $|f|_{\infty} := \sup_{\theta \in \Theta} |f(\theta)|$  and  $\mathbb{E}_{\nu_1}(f) := \int_{\Theta} f(\theta) \nu_1(d\theta)$  for measurable  $f: \Theta \rightarrow \mathbb{R}$ .

**Theorem 2.1.** Suppose that  $Z: \Theta \rightarrow (0, \infty)$  and  $\tilde{Z}: \Theta \rightarrow (0, \infty)$  are measurable functions. Then, for  $\pi_Z$  and  $\pi_{\tilde{Z}}$  as in (1.1), we have

$$\|\pi_Z - \pi_{\tilde{Z}}\|_{\text{tv}} \leq 2 \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_{Z,1}}. \tag{2.1}$$

*Proof.* For a measurable function  $f: \Theta \rightarrow \mathbb{R}$  one has

$$|\mathbb{E}_{\pi_Z}(f) - \mathbb{E}_{\pi_{\tilde{Z}}}(f)| = \left| \int_{\Theta} f(\theta) \exp(-\Phi(\theta)) \left[ \frac{1}{\tilde{Z}(\theta)C_{\tilde{Z}}} - \frac{1}{Z(\theta)C_Z} \right] \mu(d\theta) \right|. \tag{2.2}$$

In order to bound the right-hand side, we note

$$\left| \frac{1}{\tilde{Z}(\theta)C_{\tilde{Z}}} - \frac{1}{Z(\theta)C_Z} \right| \leq \frac{|C_Z - C_{\tilde{Z}}|}{C_{\tilde{Z}}C_Z\tilde{Z}(\theta)} + \frac{|Z(\theta) - \tilde{Z}(\theta)|}{C_Z\tilde{Z}(\theta)Z(\theta)}, \tag{2.3}$$

such that, with finite  $|f|_{\infty} := \sup_{\theta \in \Theta} |f(\theta)|$ , we obtain

$$|\mathbb{E}_{\pi_Z}(f) - \mathbb{E}_{\pi_{\tilde{Z}}}(f)| \leq |f|_{\infty} \frac{|C_Z - C_{\tilde{Z}}|}{C_Z} + |f|_{\infty} \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_{Z,1}}.$$

Furthermore

$$\frac{|C_{\tilde{Z}} - C_Z|}{C_Z} \leq \frac{1}{C_Z} \int_{\Theta} \exp(-\Phi(\theta)) \left| \frac{1}{\tilde{Z}(\theta)} - \frac{1}{Z(\theta)} \right| \mu(d\theta) = \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_{Z,1}}. \tag{2.4}$$

By  $\|\pi_Z - \pi_{\tilde{Z}}\|_{\text{tv}} = \sup_{|f|_{\infty} \leq 1} |\mathbb{E}_{\pi_Z}(f) - \mathbb{E}_{\pi_{\tilde{Z}}}(f)|$  the statement follows.  $\square$

Let us provide some remarks and consequences.

**Remark 2.2.** The previous estimate is not sharp in the following sense. Set  $\tilde{Z} = cZ$  for some constant  $c > 0$ , then the left-hand side is zero, but the right-hand side in general not. This deficiency can be easily repaired by using the fact that  $\pi_{\tilde{Z}} = \pi_{\tilde{c}\tilde{Z}}$  for any arbitrary constant  $\tilde{c} > 0$ . With this fact Theorem 2.1 implies readily that

$$\|\pi_Z - \pi_{\tilde{Z}}\|_{\text{tv}} \leq 2 \inf_{\tilde{c} > 0} \left\| \frac{Z}{\tilde{c}\tilde{Z}} - 1 \right\|_{\pi_{Z,1}}. \tag{2.5}$$

For the new bound if  $\tilde{Z} = cZ$  (with  $c > 0$ ), the left- and right-hand side are both zero. In particular, for the slightly more conservative upper bound using the  $L^2(\pi_Z)$ -norm instead of the  $L^1(\pi_Z)$ -norm on the right-hand side of (2.5), one can derive that

$$\|\pi_Z - \pi_{\tilde{Z}}\|_{\text{tv}} \leq 2 \left\| \frac{\|Z/\tilde{Z}\|_{\pi_{Z,1}}}{\|Z/\tilde{Z}\|_{\pi_{Z,2}}^2} \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_{Z,2}} = \inf_{\tilde{c} > 0} \left\| \frac{Z}{\tilde{c}\tilde{Z}} - 1 \right\|_{\pi_{Z,2}}.$$

**Remark 2.3.** By interchanging the roles of  $Z$  and  $\tilde{Z}$  in (2.1) one easily obtains

$$\|\pi_Z - \pi_{\tilde{Z}}\|_{\text{tv}} \leq 2 \min \left\{ \left\| \frac{\tilde{Z}}{Z} - 1 \right\|_{\pi_{\tilde{Z},1}}, \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_{Z,1}} \right\}.$$

In the light of [19] one might ask for local Lipschitz continuity of the mapping  $Z \mapsto \pi_Z$ .

**Corollary 2.4.** Suppose that  $\inf_{\theta \in \Theta} \tilde{Z}(\theta) \geq \ell$  for some  $\ell > 0$ , then

$$\|\pi_Z - \pi_{\tilde{Z}}\|_{\text{tv}} \leq \frac{2}{\ell} \left\| \tilde{Z} - Z \right\|_{\pi_{Z,1}}.$$

Note that in Theorem 2.1 and in the previous corollary on the right-hand side the norm itself already depends on  $Z$ . One might argue that this hides some  $Z$ -dependence. Under an additional requirement we can remove this dependence by applying Hölder’s inequality.

**Corollary 2.5.** *Suppose that for  $p \in [1, \infty]$  we have  $\|\exp(-\Phi)/Z\|_{\mu,p} \leq K$  for some  $K < \infty$ , then*

$$\|\pi_Z - \pi_{\tilde{Z}}\|_{\text{tv}} \leq \frac{2K}{C_Z} \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\mu,p/(p-1)}.$$

If additionally  $\inf_{\theta \in \Theta} \tilde{Z}(\theta) \geq \ell$  for some  $\ell > 0$  then

$$\|\pi_Z - \pi_{\tilde{Z}}\|_{\text{tv}} \leq \frac{2K}{\ell C_Z} \|Z - \tilde{Z}\|_{\mu,p/(p-1)}.$$

The last inequality provides a local Lipschitz continuity w.r.t. the  $L_{p/(p-1)}(\mu)$ -norm in contrast to the local Lipschitz continuity w.r.t the  $L_1(\pi_Z)$ -norm of Corollary 2.4. For  $p = \infty$  it is essentially the estimate of [19, Theorem 8] in our context.

**Remark 2.6.** In the light of the final estimate in the previous corollary let us explain in more detail what we mean with local Lipschitz continuity: Define the set  $\mathcal{Z}_{\ell,p,K}$  for  $\ell > 0$ ,  $K < \infty$  and  $1 \leq p \leq \infty$  by all measurable functions  $Z: \Theta \rightarrow [\ell, \infty)$  with  $Z \in L_{p/(p-1)}(\mu)$  and  $\|\exp(-\Phi)/Z\|_{\mu,p} \leq K$ . Then, for any  $Z, \tilde{Z} \in \mathcal{Z}_{\ell,p,K}$  there exists a constant  $R_Z < \infty$  such that

$$\|\pi_Z - \pi_{\tilde{Z}}\|_{\text{tv}} \leq R_Z \|Z - \tilde{Z}\|_{\mu,p/(p-1)}.$$

## 2.2 Wasserstein distance

In the recent years the Wasserstein distance has become a standard tool in applied probability and statistics, see for example [9, 16, 18, 19]. One advantage of this distance is that it takes topological properties of the metric space  $(\Theta, d)$  into account, which provides a certain flexibility. For instance, for  $\theta, \tilde{\theta} \in \Theta$  the Wasserstein distance of the Dirac measures  $\delta_\theta$  and  $\delta_{\tilde{\theta}}$  goes to zero when  $d(\theta, \tilde{\theta}) \rightarrow 0$ .

Let us briefly provide definitions and basic facts. For probability distributions  $\nu_1, \nu_2$  on  $(\Theta, \mathcal{B}(\Theta))$  the Wasserstein distance of  $\nu_1$  and  $\nu_2$  is given by

$$W(\nu_1, \nu_2) := \inf_{\eta \in C(\nu_1, \nu_2)} \int_{\Theta \times \Theta} d(\theta_1, \theta_2) \eta(d(\theta_1, \theta_2)),$$

where  $C(\nu_1, \nu_2)$  denotes the set of couplings of  $\nu_1$  and  $\nu_2$ , that is, a probability measure  $\eta$  on  $\Theta \times \Theta$  belongs to  $C(\nu_1, \nu_2)$  (by definition) if  $\eta(A \times \Theta) = \nu_1(A)$  and  $\eta(\Theta \times A) = \nu_2(A)$  for any  $A \in \mathcal{B}(\Theta)$ . For measurable  $f: \Theta \rightarrow \mathbb{R}$  define the Lipschitz semi-norm

$$\|f\|_{\text{Lip}} := \sup_{\theta_1, \theta_2 \in \Theta, \theta_1 \neq \theta_2} \frac{|f(\theta_1) - f(\theta_2)|}{d(\theta_1, \theta_2)}$$

and note that the Wasserstein distance allows a dual representation. For arbitrary  $\theta_0 \in \Theta$  the Kantorovich-Rubinstein duality, for details see [21], is given by

$$W(\nu_1, \nu_2) = \sup_{\|f\|_{\text{Lip}} \leq 1, f(\theta_0) = 0} |\mathbb{E}_{\nu_1}(f) - \mathbb{E}_{\nu_2}(f)|. \tag{2.6}$$

**Remark 2.7.** Note that we only consider the 1-Wasserstein distance in contrast to the general  $p$ -Wasserstein distance given by

$$W_p(\nu_1, \nu_2) := \inf_{\eta \in C(\nu_1, \nu_2)} \left[ \int_{\Theta \times \Theta} d(\theta_1, \theta_2)^p \eta(d(\theta_1, \theta_2)) \right]^{1/p}$$

for  $p \geq 1$ . The reason for that lies in the fact that we essentially use the Kantorovich-Rubinstein duality. For the  $p$ -Wasserstein distance there is also a dual characterization, see [21], which is more involved and does not readily allow to employ (2.3) for estimating  $W_p$  in the case of  $p > 1$ .

Furthermore, for a measure  $\nu$  on  $(\Theta, \mathcal{B}(\Theta))$  and  $p \geq 1$  define

$$|\nu|^{(p)} := \inf_{\theta_0 \in \Theta} \left( \int_{\Theta} d(\theta_0, \theta)^p \nu(d\theta) \right)^{1/p}.$$

Note that if the metric  $d$  is bounded, that is,  $\sup_{\theta_1, \theta_2 \in \Theta} d(\theta_1, \theta_2) \leq D$  for some  $D < \infty$ , then  $|\nu|^{(p)} \leq D$ . Now we are able to formulate the Wasserstein stability estimate.

**Theorem 2.8.** *Suppose that  $Z: \Theta \rightarrow (0, \infty)$  and  $\tilde{Z}: \Theta \rightarrow (0, \infty)$  are measurable functions. Then, for  $\pi_Z$  and  $\pi_{\tilde{Z}}$  as in (1.1), we have*

$$\begin{aligned} W(\pi_Z, \pi_{\tilde{Z}}) &\leq \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_{Z,1}} |\pi_{\tilde{Z}}|^{(1)} + \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_{Z,2}} |\pi_Z|^{(2)} \\ &\leq \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_{Z,2}} \left( |\pi_{\tilde{Z}}|^{(1)} + |\pi_Z|^{(2)} \right). \end{aligned}$$

Moreover, if  $Z$  and  $\tilde{Z}$  are sufficiently close to each other, that is,  $\left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_{Z,2}} \leq 1 - \varepsilon$ , for an  $\varepsilon \in (0, 1)$ , then

$$W(\pi_Z, \pi_{\tilde{Z}}) \leq \frac{2}{\varepsilon} |\pi_Z|^{(2)} \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_{Z,2}}. \tag{2.7}$$

*Proof.* For arbitrary  $\theta_0 \in \Theta$  by (2.6) it is sufficient to estimate the right-hand side of (2.2). For this we again use (2.3) and obtain

$$W(\pi_Z, \pi_{\tilde{Z}}) \leq I_1 + I_2, \tag{2.8}$$

where

$$\begin{aligned} I_1 &:= \frac{|C_Z - C_{\tilde{Z}}|}{C_Z} \sup_{\|f\|_{\text{Lip}} \leq 1, f(\theta_0)=0} \int_{\Theta} f(\theta) \pi_{\tilde{Z}}(d\theta) \\ I_2 &:= \sup_{\|f\|_{\text{Lip}} \leq 1, f(\theta_0)=0} \int_{\Theta} f(\theta) \frac{|Z(\theta) - \tilde{Z}(\theta)|}{\tilde{Z}(\theta)} \pi_Z(d\theta). \end{aligned}$$

By (2.4) and taking the infimum over  $\theta_0 \in \Theta$  we obtain  $I_1 \leq \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_{Z,1}} |\pi_{\tilde{Z}}|^{(1)}$  and by additionally using the Cauchy-Schwarz inequality we get  $I_2 \leq \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_{Z,2}} |\pi_Z|^{(2)}$  which proves the first assertion. The second statement follows easily from the first by

$$|\pi_{\tilde{Z}}|^{(1)} \leq |\pi_Z|^{(1)} + \left| |\pi_{\tilde{Z}}|^{(1)} - |\pi_Z|^{(1)} \right| \leq |\pi_Z|^{(2)} + W(\pi_Z, \pi_{\tilde{Z}})$$

and rearranging the terms accordingly. □

**Remark 2.9.** Having an approximation  $\tilde{Z}$  of  $Z$  in mind, (2.7) tells us that “asymptotically”, that is, with sufficient accuracy of the recovery algorithm  $\tilde{Z}$ , no explicit bound on  $|\pi_{\tilde{Z}}|^{(1)}$  is required.

As in the total variation distance consideration, by a boundedness assumption and Hölder’s inequality we can exchange the  $\pi_Z$ -dependence on the right-hand side, by an explicit  $Z$ -dependence.

**Corollary 2.10.** *Suppose for  $p \in [1, \infty]$  that*

$$\|\exp(-\Phi)/Z\|_{\mu,p} \leq K \quad \text{and} \quad \|\exp(-\Phi)/\tilde{Z}\|_{\mu,p} \leq K,$$

for some  $K < \infty$ , then

$$W(\pi_Z, \pi_{\tilde{Z}}) \leq \frac{K|\mu|^{(2p/(p-1))}}{C_Z} \left( \frac{1}{C_Z} + \frac{1}{C_{\tilde{Z}}} \right) \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\mu, 2p/(p-1)}.$$

In particular, if additionally  $\inf_{\theta \in \Theta} \tilde{Z}(\theta) \geq \ell$  for some  $\ell > 0$ , then

$$W(\pi_Z, \pi_{\tilde{Z}}) \leq \frac{K|\mu|^{(2p/(p-1))}}{\ell C_Z} \left( \frac{1}{C_Z} + \frac{1}{C_{\tilde{Z}}} \right) \|Z - \tilde{Z}\|_{\mu, 2p/(p-1)}.$$

*Proof.* Set  $q := p/(p - 1)$  and note that  $1/p + 1/q = 1$ . In the following we frequently apply Hölder’s inequality. We have

$$\left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\pi_Z, 2} = \left\| \left( \frac{Z}{\tilde{Z}} - 1 \right)^2 \cdot \frac{\exp(-\Phi)}{ZC_Z} \right\|_{\mu, 1}^{1/2} \leq \frac{\sqrt{K}}{C_Z} \left\| \frac{Z}{\tilde{Z}} - 1 \right\|_{\mu, 2q}.$$

Furthermore, note that  $|\pi_{\tilde{Z}}|^{(1)} \leq |\pi_Z|^{(2)}$  and

$$|\pi_{\tilde{Z}}|^{(2)} = \inf_{\theta_0 \in \Theta} \left\| d(\theta_0, \cdot)^2 \frac{\exp(-\Phi)}{\tilde{Z}C_{\tilde{Z}}} \right\|_{\mu, 1}^{1/2} \leq \frac{\sqrt{K}}{C_{\tilde{Z}}} |\mu|^{(2q)}.$$

By the same arguments holds  $|\pi_Z|^{(2)} \leq \frac{\sqrt{K}}{C_Z} |\mu|^{(2q)}$ . Thus, by taking the previous estimate of Theorem 2.8 into account the proof is concluded.  $\square$

### 2.3 Randomization

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and let  $\mathbb{E}$  denote the expectation w.r.t.  $\mathbb{P}$ . Furthermore, let  $\tilde{Z}: \Omega \times \Theta \rightarrow (0, \infty)$  be a jointly measurable function, thus  $\tilde{Z}(\cdot, \theta)$  is a random variable for all  $\theta \in \Theta$ . Then, we call  $\tilde{Z}$  random function. By standard arguments the mapping  $\omega \mapsto \|\pi_Z - \pi_{\tilde{Z}(\omega, \cdot)}\|_{\text{tv}}$  from  $\Omega$  to  $\mathbb{R}$  is measurable. Similarly, if  $|\pi_Z|^{(1)} < \infty$  and  $|\pi_{\tilde{Z}(\omega, \cdot)}|^{(1)} < \infty$  for any  $\omega \in \Omega$ , then also the mapping  $\omega \mapsto W(\pi_Z, \pi_{\tilde{Z}(\omega, \cdot)})$  is measurable.

In this context Theorem 2.1 combined with a Fubini argument and Cauchy–Schwarz inequality lead to the following consequence:

**Corollary 2.11.** *Suppose  $Z: \Theta \rightarrow (0, \infty)$  is measurable and let  $\tilde{Z}: \Omega \times \Theta \rightarrow (0, \infty)$  be a random function. Then, for  $\pi_Z$  and  $\pi_{\tilde{Z}(\omega, \cdot)}$ ,  $\omega \in \Omega$ , given according to (1.1), we have*

$$\begin{aligned} \mathbb{E} \|\pi_Z - \pi_{\tilde{Z}(\cdot, \cdot)}\|_{\text{tv}} &\leq 2 \int_{\Theta} \mathbb{E} \left| \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} - 1 \right| \pi_Z(d\theta) \\ &\leq 2 \int_{\Theta} \left( \mathbb{E} \left| \frac{\tilde{Z}(\cdot, \theta)}{Z(\theta)} - 1 \right|^2 \right)^{1/2} \left( \mathbb{E} \left[ \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} \right]^2 \right)^{1/2} \pi_Z(d\theta). \end{aligned} \tag{2.9}$$

**Remark 2.12.** The final estimate (2.9) follows immediately by the Cauchy–Schwarz inequality. For Monte Carlo recovery approximation  $\tilde{Z}(\cdot, \theta)$  of  $Z$  it is usually convenient to bound the right-hand side of the second inequality in Corollary 2.11. The reason behind is that the term  $\mathbb{E} \left| \frac{\tilde{Z}(\cdot, \theta)}{Z(\theta)} - 1 \right|^2$  relates readily to the relative mean squared error of the method, whereas the “reversed” relative mean absolute error  $\mathbb{E} \left| \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} - 1 \right|$  is harder to bound directly.

Using similar arguments as in the proof of Theorem 2.8 we obtain the following result w.r.t. the expected Wasserstein distance.

**Theorem 2.13.** Suppose that  $Z: \Theta \rightarrow (0, \infty)$ , that  $\tilde{Z}: \Omega \times \Theta \rightarrow (0, \infty)$  is a random function and that for any  $\omega \in \Omega$  the probability measure  $\pi_{\tilde{Z}(\omega)}$  and  $\pi_Z$  are given through (1.1).

(i) Assume that for  $\mathbb{P}$ -almost any  $\omega \in \Omega$  we have  $|\pi_{\tilde{Z}(\omega)}|^{(1)} \leq R$  for some  $R < \infty$  and  $|\pi_Z|^{(1)} < \infty$ . Then

$$\mathbb{E}W(\pi_Z, \pi_{\tilde{Z}(\cdot)}) \leq \int_{\Theta} \mathbb{E} \left| \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} - 1 \right| (d(\theta, \theta_0) + R) \pi_Z(d\theta), \tag{2.10}$$

for any  $\theta_0 \in \Theta$ . In particular,

$$\mathbb{E}W(\pi_Z, \pi_{\tilde{Z}(\cdot)}) \leq (|\pi_Z|^{(2)} + R) \left( \int_{\Theta} \mathbb{E} \left| \frac{\tilde{Z}(\cdot, \theta)}{Z(\theta)} - 1 \right|^2 \mathbb{E} \left[ \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} \right]^2 \pi_Z(d\theta) \right)^{1/2}.$$

(ii) Assume that for  $\mathbb{P}$ -almost any  $\omega \in \Omega$  we have  $\left\| \frac{Z}{\tilde{Z}(\omega)} - 1 \right\|_{\pi_Z, 2} \leq 1 - \varepsilon$  for a number  $\varepsilon \in (0, 1)$ . Then

$$\mathbb{E}W(\pi_Z, \pi_{\tilde{Z}(\cdot)}) \leq \frac{2}{\varepsilon} |\pi_Z|^{(2)} \left( \int_{\Theta} \mathbb{E} \left| \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} - 1 \right|^2 \pi_Z(d\theta) \right)^{1/2}.$$

*Proof.* First we verify the statement of (2.13). From (2.8) we obtain for arbitrary  $\theta_0 \in \Theta$  that

$$\begin{aligned} \mathbb{E}W(\pi_Z, \pi_{\tilde{Z}(\cdot)}) &\leq R \mathbb{E} \left( \frac{|C_Z - C_{\tilde{Z}(\cdot)}|}{C_Z} \right) + \mathbb{E} \left( \int_{\Theta} d(\theta, \theta_0) \left| \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} - 1 \right| \pi_Z(d\theta) \right) \\ &\stackrel{(2.4)}{\leq} R \mathbb{E} \left\| \frac{Z}{\tilde{Z}(\cdot)} - 1 \right\|_{\pi_Z, 1} + \int_{\Theta} d(\theta, \theta_0) \mathbb{E} \left| \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} - 1 \right| \pi_Z(d\theta) \\ &= \int_{\Theta} (d(\theta, \theta_0) + R) \mathbb{E} \left| \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} - 1 \right| \pi_Z(d\theta), \end{aligned}$$

which gives (2.10). Further estimating (by using the Cauchy–Schwarz inequality several times) leads to

$$\begin{aligned} \mathbb{E}W(\pi_Z, \pi_{\tilde{Z}(\cdot)}) &\leq (|\pi_Z|^{(2)} + R) \left( \int_{\Theta} \left[ \mathbb{E} \left| \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} - 1 \right|^2 \right] \pi_Z(d\theta) \right)^{1/2} \\ &\leq (|\pi_Z|^{(2)} + R) \left( \int_{\Theta} \mathbb{E} \left| \frac{\tilde{Z}(\cdot, \theta)}{Z(\theta)} - 1 \right|^2 \mathbb{E} \left[ \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} \right]^2 \pi_Z(d\theta) \right)^{1/2}. \end{aligned}$$

Now we turn to the proof of the statement of (2.13). By (2.7) it is sufficient to estimate  $\mathbb{E} \left\| \frac{Z}{\tilde{Z}(\cdot)} - 1 \right\|_{\pi_Z, 2}$ . With Jensen’s inequality as well as a change of integration we have

$$\mathbb{E} \left\| \frac{Z}{\tilde{Z}(\cdot)} - 1 \right\|_{\pi_Z, 2} \leq \left( \int_{\Theta} \mathbb{E} \left| \frac{Z(\theta)}{\tilde{Z}(\cdot, \theta)} - 1 \right|^2 \pi_Z(d\theta) \right)^{1/2}$$

which concludes the proof. □

**Remark 2.14.** The condition in the first statement of Theorem 2.13 that for some  $R < \infty$  we assume  $|\pi_{\tilde{Z}(\omega)}|^{(1)} \leq R$  for almost any  $\omega \in \Omega$  seems to be restrictive. We would like to comment on that:

1. If  $(\Theta, d)$  is a bounded metric space, then  $|\pi_{\tilde{Z}(\omega)}|^{(1)} \leq R$  holds for  $R$  being the upper bound for  $d$ . Further note, that by  $\tilde{d}(\theta_1, \theta_2) := \min\{R, d(\theta_1, \theta_2)\}$  we can turn any metric space  $(\Theta, d)$  into a topologically equivalent bounded metric space  $(\Theta, \tilde{d})$ .
2. If  $\tilde{Z}$  is a Monte Carlo approximation of  $Z$ , then we might be able to derive functions  $\tilde{\ell}, \tilde{u}: \Theta \rightarrow \mathbb{R}$  satisfying  $\tilde{\ell}(\theta) \leq \tilde{Z}(\omega, \theta) \leq \tilde{u}(\theta)$  for all  $(\omega, \theta) \in \Omega \times \Theta$ , cf. Section 3. Setting

$$R := \frac{\inf_{\theta_0 \in \Theta} \int_{\Theta} d(\theta_0, \theta) \frac{\exp(-\Phi(\theta))}{\tilde{\ell}(\theta)} \mu(d\theta)}{\int_{\Theta} \frac{\exp(-\Phi(\theta))}{\tilde{u}(\theta)} \mu(d\theta)}$$

and assuming its finiteness then leads to  $|\pi_{\tilde{Z}(\omega)}|^{(1)} \leq R$ .

3. In addition to that, the second statement of Theorem 2.13 tells us, within the setting of a Monte Carlo approximation  $\tilde{Z}$ , that “asymptotically”, if at some accuracy level  $\tilde{Z}$  is sufficiently close to  $Z$  with probability one, then no explicit upper bound on  $|\pi_{\tilde{Z}(\omega)}|^{(1)}$  is required.

**Remark 2.15.** Related to results in this subsection are the recent publications [11, 12]. There, the authors studied well-posedness properties of Bayesian inverse problems with random likelihoods. In particular, they bounded the expected squared Hellinger distance between a random approximate posterior distribution and a desired one.

### 3 Monte Carlo recovery - illustrative examples

Given the probability measure  $\pi_Z$  on  $(\Theta, \mathcal{B}(\Theta))$  one might ask how to get  $\tilde{Z}$  which is “close” to  $Z$  in the sense of our stability results. Thus, we are looking for an approximation  $\tilde{Z}$  of  $Z$ . Observe that, for an arbitrary, maybe even unknown, constant  $c > 0$  it is actually sufficient to approximate the function  $c \cdot Z$ , since  $\pi_{cZ} = \pi_Z$ . However, the crucial difficulty lies in the fact that we do not have access to function evaluations of  $c \cdot Z$ . Nevertheless, there exist scenarios where Monte Carlo estimators can successfully be applied, see [1, 2, 4, 18]. This motivates the consideration of Monte Carlo recovery methods for  $Z$ . A random function  $\tilde{Z}_N$ , as defined in Section 2.3, is a Monte Carlo approximation of  $Z$  with information parameter  $N$  if  $\tilde{Z}_N$  uses at most  $N$  pieces of available information, e.g. function evaluations or samples w.r.t. a certain distribution. We consider two illustrating scenarios for which we provide explicit Monte Carlo approximations of  $Z$ .

#### 3.1 Simple Monte Carlo recovery

We consider the same framework as in [14, Section 4.1]. Let  $(G, \mathcal{G})$  be a measurable space and  $\rho: G \times \Theta \rightarrow [0, \infty)$  be a measurable function, such that

$$Z(\theta) := \int_G \rho(x, \theta) \nu_{\theta}(dx) \in (0, \infty),$$

where  $(\nu_{\theta})_{\theta \in \Theta}$  is a family of probability distribution on  $G$ . With another measurable function  $\Phi: \Theta \rightarrow (-\infty, \infty)$  this defines the distribution of interest  $\pi_Z$  as given in (1.1).

We assume that for any  $\theta \in \Theta$  we can sample w.r.t.  $\nu_{\theta}$ . Then  $\tilde{Z}_N(\theta) := \frac{1}{N} \sum_{j=1}^N \rho(X_j^{(\theta)}, \theta)$ , with  $X_1^{(\theta)}, \dots, X_N^{(\theta)}$  being an iid sample w.r.t.  $\nu_{\theta}$ , provides a Monte Carlo approximation of  $Z$ . Define  $Q_N(\theta) := \tilde{Z}_N(\theta)/Z(\theta)$  and note that  $(\mathbb{E}|Q_N(\theta) - 1|^2)^{1/2} = \frac{(\mathbb{E}|Q_1(\theta) - 1|^2)^{1/2}}{\sqrt{N}}$ , as well as by [14, Lemma 23] we have for the second inverse moment of  $Q_N(\theta)$  that  $(\mathbb{E}Q_N(\theta)^{-2})^{1/2} \leq (\mathbb{E}Q_1(\theta)^{-2})^{1/2}$ . Hence by Corollary 2.11 we obtain

$$\mathbb{E} \|\pi_Z - \pi_{\tilde{Z}_N(\cdot)}\|_{\text{tv}} \leq \frac{2}{\sqrt{N}} \int_{\Theta} (\mathbb{E}|Q_1(\theta) - 1|^2 \cdot \mathbb{E}Q_1(\theta)^{-2})^{1/2} \pi_Z(d\theta). \quad (3.1)$$

To further estimate the former inequality we impose the following regularity assumption throughout the rest of this section.

**Assumption 3.1.** For  $\rho: G \times \Theta \rightarrow [0, \infty)$  assume that there are measurable functions  $\ell, u: \Theta \rightarrow (0, \infty)$  satisfying  $\ell(\theta) \leq \rho(x, \theta) \leq u(\theta)$ , for all  $(x, \theta) \in G \times \Theta$ .

Then, we have  $\mathbb{E}|Q_1(\theta) - 1|^2 \leq \mathbb{E}Q_1(\theta)^2 \leq u(\theta)^2/Z(\theta)^2$  as well as  $\mathbb{E}Q_1(\theta)^{-2} \leq Z(\theta)^2/\ell(\theta)^2$ . Thus, with (3.1) we obtain

$$\mathbb{E}\|\pi_Z - \pi_{\tilde{Z}_N(\cdot)}\|_{\text{tv}} \leq \frac{2}{\sqrt{N}} \left\| \frac{u}{\ell} \right\|_{\pi_{Z,1}}. \tag{3.2}$$

Furthermore, setting

$$R := \frac{\inf_{\theta_0 \in \Theta} \int_{\Theta} d(\theta_0, \theta) \frac{\exp(-\Phi(\theta))}{\ell(\theta)} \mu(d\theta)}{\int_{\Theta} \frac{\exp(-\Phi(\theta))}{u(\theta)} \mu(d\theta)}$$

and assuming that it is finite easily gives  $|\pi_{\tilde{Z}_N(\omega, \cdot)}|^{(1)} \leq R$  for any  $\omega \in \Omega$ , see also Remark 2.14. Then, by Theorem 2.13 we have

$$\mathbb{E}W(\pi_Z, \pi_{\tilde{Z}_N(\cdot)}) \leq \frac{(R + |\pi_Z|^{(2)})}{\sqrt{N}} \left\| \frac{u}{\ell} \right\|_{\pi_{Z,2}}. \tag{3.3}$$

Summarized, one can say that if the integrals on the right-hand sides of (3.1), (3.2) and (3.3) are finite, then the difference of  $\pi_Z$  and  $\pi_{\tilde{Z}_N(\cdot)}$  measured either in the total variation or Wasserstein distances decreases with the classical Monte Carlo rate of convergence.

### 3.2 Gibbs distribution

Let  $G$  be a finite set and suppose that there is a measurable function  $h: G \times \Theta \rightarrow (-\infty, \infty)$ , such that

$$\rho(x | \theta) := \frac{\exp(-h(x, \theta))}{Z(\theta)}, \quad x \in G,$$

is a probability mass function on  $G$ , where  $Z(\theta) := \sum_{x \in G} \exp(-h(x, \theta))$  denotes the normalizing constant of  $\exp(-h(x, \theta))$  given  $\theta \in \Theta$ . In that setting, for a given  $x_{\text{obs}} \in G$  we have  $\Phi(\theta) = h(x_{\text{obs}}, \theta)$ . Note that this framework contains Example 1.1 as a special case. We impose the following condition.

**Assumption 3.2.** We assume that we can sample on  $G$  w.r.t. the distribution determined by  $\rho(\cdot | \theta)$  for any  $\theta \in \Theta$ .

Usually there are two arguments why this assumption is not too restrictive. The first is that one might rely on perfect sampling and the second, that one can, at least approximately, sample from such distributions by using Markov chains.

**Remark 3.3.** In the scenario of Example 1.1 for a given  $\theta \in \Theta$  (or rather  $\beta \in (0, \infty)$ ) in a series of papers a Monte Carlo product estimator for  $Z(\theta)$  (or better  $Z(\beta)$ ) has been analyzed, see [6] and the references therein. As an alternative also multiple importance sampling approaches have been applied for the approximation of  $Z(\theta)$ , see [2, 4, 10].

Motivated by that we consider a simple multiple importance sampling method. Assume that  $\theta_1, \dots, \theta_J \in \Theta$  and note that for any  $j \in \{1, \dots, J\}$  holds

$$\frac{Z(\theta)}{Z(\theta_j)} = \sum_{x \in G} \frac{\exp(-h(x, \theta))}{\exp(-h(x, \theta_j))} \rho(x | \theta_j).$$

For  $N \in \mathbb{N}$  let  $X_{1,j}, \dots, X_{N,j}$  be an iid sample w.r.t.  $\rho(\cdot | \theta_j)$  and for any  $i \in \{1, \dots, N\}$  let  $X_{i,1}, \dots, X_{i,J}$  be independent. Given  $p_1, \dots, p_J \in (0, 1)$  with  $\sum_{j=1}^J p_j = 1$  let

$$S^{(i)}(\theta) := \sum_{j=1}^J p_j \frac{\exp(-h(X_{i,j}, \theta))}{\exp(-h(X_{i,j}, \theta_j))}$$

and with  $S(\theta) := Z(\theta) \sum_{j=1}^J p_j / Z(\theta_j)$  note that  $\mathbb{E}S^{(i)}(\theta) = S(\theta)$ . Intuitively  $p_j$  weights the importance of the sample derived by  $\rho(\cdot | \theta_j)$ . Then, an unbiased estimator of  $S(\theta)$  is given by the multiple importance sampler

$$\tilde{S}_N(\theta) := \frac{1}{N} \sum_{i=1}^N S^{(i)}(\theta).$$

Note that  $S^{(1)}(\theta), \dots, S^{(N)}(\theta)$  is a sequence of iid random variables. With  $Q_N(\theta) := \tilde{S}_N(\theta) / S(\theta)$  a simple calculation reveals that  $(\mathbb{E}|Q_N(\theta) - 1|^2)^{1/2} = \frac{(\mathbb{E}|Q_1(\theta) - 1|^2)^{1/2}}{\sqrt{N}}$ . For the inverse second moment we have by [14, Lemma 2.3(i)] that  $\mathbb{E}[Q_N(\theta)^{-2}]^{1/2} \leq \mathbb{E}[Q_1(\theta)^{-2}]^{1/2}$ . Hence, by the fact that  $\pi_Z = \pi_S$  and by Corollary 2.4 we obtain

$$\begin{aligned} \mathbb{E}\|\pi_Z - \pi_{\tilde{S}_N(\cdot)}\|_{\text{tv}} &= \mathbb{E}\|\pi_S - \pi_{\tilde{S}_N(\cdot)}\|_{\text{tv}} \\ &\leq \frac{2}{\sqrt{N}} \int_{\Theta} (\mathbb{E}|Q_1(\theta) - 1|^2 \cdot \mathbb{E}Q_1(\theta)^{-2})^{1/2} \pi_Z(d\theta). \end{aligned} \tag{3.4}$$

To elaborate on this we impose the following assumption.

**Assumption 3.4.** For  $h: G \times \Theta \rightarrow (-\infty, \infty)$  assume that there are measurable functions  $\ell, u: \Theta \rightarrow [0, \infty)$  satisfying  $\ell(\theta) \leq \exp(-h(x, \theta)) \leq u(\theta)$  for all  $(x, \theta) \in G \times \Theta$ .

Under this condition we have with the identity of Bieanymé that

$$\begin{aligned} \mathbb{E}|Q_1(\theta) - 1|^2 &= \frac{1}{S(\theta)^2} \mathbb{E} \left[ \sum_{j=1}^J p_j \left( \frac{\exp(-h(X_{1,j}, \theta))}{\exp(-h(X_{1,j}, \theta_j))} - \frac{Z(\theta)}{Z(\theta_j)} \right) \right]^2 \\ &= \frac{1}{S(\theta)^2} \sum_{j=1}^J p_j^2 \mathbb{E} \left[ \frac{\exp(-h(X_{1,j}, \theta))}{\exp(-h(X_{1,j}, \theta_j))} - \frac{Z(\theta)}{Z(\theta_j)} \right]^2 \\ &\leq \frac{1}{S(\theta)^2} \sum_{j=1}^J p_j^2 \mathbb{E} \left( \frac{\exp(-2h(X_{1,j}, \theta))}{\exp(-2h(X_{1,j}, \theta_j))} \right) \leq \frac{u(\theta)^2}{S(\theta)^2} \sum_{j=1}^J \frac{p_j^2}{\ell(\theta_j)^2} \end{aligned}$$

and

$$\mathbb{E}Q_1(\theta)^{-2} = S(\theta)^2 \mathbb{E} \left( \sum_{k=1}^J p_k \frac{\exp(-h(X_{1,k}, \theta))}{\exp(-h(X_{1,k}, \theta_k))} \right)^{-2} \leq \frac{S(\theta)^2}{\ell(\theta)^2} \left( \sum_{k=1}^J \frac{p_k}{u(\theta_k)} \right)^{-2}.$$

Therefore, (3.4) implies

$$\mathbb{E}\|\pi_Z - \pi_{\tilde{S}_N(\cdot)}\|_{\text{tv}} \leq \frac{2}{\sqrt{N}} \frac{\left( \sum_{j=1}^J \frac{p_j^2}{\ell(\theta_j)^2} \right)^{1/2}}{\sum_{k=1}^J \frac{p_k}{u(\theta_k)}} \left\| \frac{u}{\ell} \right\|_{\pi_Z, 1}.$$

Under Assumption 3.4 we can also derive an upper bound for the expected Wasserstein distance of  $\pi_Z$  and  $\pi_{\tilde{S}_N(\cdot)}$ . For this observe that  $\ell(\theta) \sum_{j=1}^J \frac{p_j}{u(\theta_j)} \leq \tilde{S}_N(\theta) \leq u(\theta) \sum_{k=1}^J \frac{p_k}{\ell(\theta_k)}$ , such that, if

$$R := \frac{\sum_{k=1}^J \frac{p_k}{\ell(\theta_k)}}{\sum_{j=1}^J \frac{p_j}{u(\theta_j)}} \cdot \frac{\inf_{\theta_0 \in \Theta} \int_{\Theta} d(\theta, \theta_0) \frac{u(\theta)}{\ell(\theta)} \mu(d\theta)}{\int_{\Theta} \frac{u(\theta)}{\ell(\theta)} \mu(d\theta)}$$

is finite, then  $|\pi_{\tilde{S}_N(\omega)}|^{(1)} \leq R < \infty$ , compare Remark 2.14. Now, we obtain with Theorem 2.13, in particular (2.10), that

$$\mathbb{E}W(\pi_Z, \pi_{\tilde{S}_N(\cdot)}) \leq \frac{(|\pi_Z|^{(2)} + R)}{\sqrt{N}} \frac{\left(\sum_{j=1}^J \frac{p_j^2}{\ell(\theta_j)^2}\right)^{1/2}}{\sum_{k=1}^J \frac{p_k}{u(\theta_k)}} \left\| \frac{u}{\ell} \right\|_{\pi_Z, 2}.$$

Summarized, we observe again that the expected total variation as well as expected Wasserstein distance decays with (at least) the usual Monte Carlo rate of convergence.

## 4 Conclusion

We conducted a stability analysis of doubly-intractable distributions. In particular, given two functions  $Z, \tilde{Z}: \Theta \rightarrow (0, \infty)$  we derived estimates of the total variation and Wasserstein distance of  $\pi_Z$  and  $\tilde{\pi}_{\tilde{Z}}$ . Essentially it turns out that if a relative difference between  $Z$  and  $\tilde{Z}$ , measured in a certain  $L^p$ -sense, is small, then also  $\pi_Z$  and  $\pi_{\tilde{Z}}$  are close to each other. We also consider a randomization of  $\tilde{Z}$ , that is, for any  $\theta \in \Theta$  we have a random variable  $\tilde{Z}(\theta)$ . In this context we provide estimates on the expected total variation and Wasserstein distance of  $\pi_Z$  to the random measure  $\pi_{\tilde{Z}}$ . In addition to that we illustrate our bounds in two simple Monte Carlo recovery settings.

Finally let us comment on further aspects. In the stability analysis we focused on the total variation and Wasserstein distance, but of course also other quantities for measuring the difference of distributions, such as the Hellinger distance or Kullback-Leibler divergence, are reasonable to investigate. Furthermore, in Section 2.3 we only considered the expected difference of distributions. In the light of [6] and also [7, 8] statements of the type “small error with high probability” are desirable. In particular, an investigation concerning the approximation of functions by Monte Carlo recovery algorithms seems to be a challenging and very interesting task.

## References

- [1] P. Alquier, N. Friel, R. Everitt, and A. Boland, *Noisy Monte Carlo: Convergence of Markov chains with approximate transition kernels*, *Statistics and Computing* **26** (2016), no. 1-2, 29–47. MR-3439357
- [2] Yves F. Atchadé, Nicolas Lartillot, Christian Robert, et al., *Bayesian computation for statistical models with intractable normalizing constants*, *Brazilian Journal of Probability and Statistics* **27** (2013), no. 4, 416–436. MR-3105037
- [3] M. Dashti and A. Stuart, *The Bayesian approach to inverse problems*, *Handbook of Uncertainty Quantification* (R. Ghanem, D. Higdon, and H. Owhadi, eds.), Springer, 2017, pp. 311–428. MR-3839555
- [4] M. Habeck, *Bayesian approach to inverse statistical mechanics*, *Physical Review E* **89** (2014), no. 5, 052113.
- [5] D. Hunter and M. Handcock, *Inference in curved exponential family models for networks*, *Journal of Computational and Graphical Statistics* **15** (2006), no. 3, 565–583. MR-2291264
- [6] V. Kolmogorov, *A faster approximation algorithm for the Gibbs partition function*, *Proceedings of the 31st Conference On Learning Theory* (Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet, eds.), *Proceedings of Machine Learning Research*, vol. 75, PMLR, 2018, pp. 228–249.
- [7] R. J. Kunsch, E. Novak, and D. Rudolf, *Solvable integration problems and optimal sample size selection*, *Journal of Complexity* **53** (2019), 40–67. MR-3953086
- [8] R. J. Kunsch and D. Rudolf, *Optimal confidence for Monte Carlo integration of smooth functions*, *Advances in Computational Mathematics* (2019), 1–28. MR-4047026
- [9] J. Latz, *On the well-posedness of Bayesian inverse problems*, arXiv preprint arXiv:1902.10257 (2019). MR-4075344

- [10] F. Liang, I. Jin, Q. Song, and J. Liu, *An adaptive exchange algorithm for sampling from distributions with intractable normalizing constants*, Journal of the American Statistical Association **111** (2016), no. 513, 377–393. MR-3494666
- [11] H. Ch. Lie, T. Sullivan, and A. Teckentrup, *Random forward models and log-likelihoods in Bayesian inverse problems*, SIAM/ASA Journal on Uncertainty Quantification **6** (2018), no. 4, 1600–1629. MR-3876869
- [12] H. Ch. Lie, T. Sullivan, and A. Teckentrup, *Error bounds for some approximate posterior measures in Bayesian inference*, arXiv preprint arXiv:1911.05669 (2019).
- [13] D. MacKay, I. Murray, and Z. Ghahramani, *MCMC for doubly-intractable distributions*, Uncertainty in Artificial Intelligence, vol. 22, 2006.
- [14] F. Medina-Aguayo, D. Rudolf and N. Schweizer, *Perturbation bounds for Monte Carlo within Metropolis via restricted approximations*, Stochastic Processes and their Applications (2019). MR-4074698
- [15] J. Møller, A. Pettitt, R. Reeves, and K. Berthelsen, *An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants*, Biometrika **93** (2006), no. 2, 451–458. MR-2278096
- [16] V. Panaretos and Y. Zemel, *Statistical aspects of Wasserstein distances*, Annual review of statistics and its application **6** (2019), 405–431. MR-3939527
- [17] J. Park and M. Haran, *Bayesian inference in the presence of intractable normalizing functions*, Journal of the American Statistical Association **113** (2018), no. 523, 1372–1390. MR-3862364
- [18] D. Rudolf and N. Schweizer, *Perturbation theory for Markov chains via Wasserstein distance*, Bernoulli (2018), no. 4A, 2610–2639. MR-3779696
- [19] B. Sprungk, *On the local Lipschitz stability of Bayesian inverse problems*, Inverse Problems **36** (2020). MR-4105328
- [20] A. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numerica **19** (2010), 451–559. MR-2652785
- [21] C. Villani, *Optimal transport: Old and new*, Springer Verlag, Berlin, 2009, Grundlehren der Mathematischen Wissenschaften, no. 338. MR-2459454

---

# Electronic Journal of Probability

## Electronic Communications in Probability

---

### Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS<sup>1</sup>)
- Easy interface (EJMS<sup>2</sup>)

### Economical model of EJP-ECP

- Non profit, sponsored by IMS<sup>3</sup>, BS<sup>4</sup>, ProjectEuclid<sup>5</sup>
- Purely electronic

### Help keep the journal free and vigorous

- Donate to the IMS open access fund<sup>6</sup> (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

---

<sup>1</sup>LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

<sup>2</sup>EJMS: Electronic Journal Management System <http://www.vtex.lt/en/ejms.html>

<sup>3</sup>IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

<sup>4</sup>BS: Bernoulli Society <http://www.bernoulli-society.org/>

<sup>5</sup>Project Euclid: <https://projecteuclid.org/>

<sup>6</sup>IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>