# Inverse Problems

## Sommersemester 2023

Vesa Kaarnioja
vesa.kaarnioja@fu-berlin.de

FU Berlin, FB Mathematik und Informatik

Ninth lecture, June 12, 2023

# Recap: Bayes' formula for inverse problems

We are interested in the inverse problem of solving $x \in \mathbb{R}^d$ from

$$y = F(x) + \eta,$$

where $y \in \mathbb{R}^k$ is the measurement vector, $F\colon \mathbb{R}^d \to \mathbb{R}^k$ the forward mapping, and $\eta \in \mathbb{R}^k$ is noise. We model $x$, $y$, and $\eta$ as random variables. Then we have:

**Theorem (Bayes' theorem)**

*We assume:*

- *The noise $\eta$ has the probability density $\nu$ on $\mathbb{R}^k$.*
- *The parameter $x$ has the probability density $\pi$ on $\mathbb{R}^d$.*
- *The random variables $x$ and $\eta$ are independent.*

*Then the likelihood is $\mathbb{P}(y|x) = \nu(y - F(x))$ and we can write*

$$\pi^y(x) := \mathbb{P}(x|y) = \frac{\mathbb{P}(y|x)\mathbb{P}(x)}{\mathbb{P}(y)} =: \frac{\nu(y - F(x))\pi(x)}{Z(y)},$$

*provided that $Z(y) := \int_{\mathbb{R}^d} \nu(y - F(x))\pi(x)\mathrm{d}x > 0$.*

Bayes' formula:

$$\pi^y(x) = \frac{\nu(y - F(x))\pi(x)}{Z(y)}.$$

- The *prior model* $\pi(x)$ describes *a priori information*. It should assign high probability to objects $x$ which are typical in light of *a priori information*, and low probability to unexpected $x$.
- The *likelihood model* $\mathbb{P}(y|x) = \nu(y - F(x))$ processes measurement information. It gives low probability to objects that produce simulated data which is very different from the measured data.
- The number $Z(y)$ can be seen as a normalization constant.
- The *posterior distribution* $\pi^y(x) = \mathbb{P}(x|y)$ represents the updated knowledge about the parameter of interest $x$, given the evidence $y$.

Since the normalization constant $Z(y)$ is often not of interest in our considerations, we frequently write the Bayes' formula as

$$\pi^y(x) \propto \nu(y - F(x))\pi(x),$$

where the symbol $\propto$ means equality up to a constant factor.

# Case study: one-dimensional deconvolution

As motivation[†], suppose that we are interested in estimating a signal $f : [0,1] \to \mathbb{R}$ from noisy, blurred observations modeled as

$$y_i = y(s_i) = \int_0^1 K(s_i, t) f(t)\, \mathrm{d}t + \eta_i, \quad i \in \{1, \ldots, k\},$$

where the blurring kernel is

$$K(s,t) = \exp\left( -\frac{1}{2\omega^2}(s-t)^2 \right), \quad \omega = 0.5,$$

and $\eta \in \mathbb{R}^k$ is measurement noise.

[†]We will consider the so-called "linear-Gaussian setting" as well as computational techniques for sampling posterior densities in more detail in a couple of weeks. Specifically, we will not consider the question of *how to draw samples from the posterior density* today. We will revisit this question in more detail at a later time.

# Discrete model

Midpoint rule:

$$y_i = \int_0^1 K(s_i, t) f(t)\,\mathrm{d}t + \eta_i \approx \frac{1}{d}\sum_{j=1}^{d} K(s_i, t_j) x_j + \eta_i,$$

where $t_j = \frac{j}{d} - \frac{1}{2d}$ and $x_j = f(t_j)$ for $j \in \{1, \dots, d\}$.

If we have $s_i = \frac{i}{k} - \frac{1}{2k}$ for $i \in \{1, \dots, k\}$, then we have the discrete linear model

$$y = Ax + \eta, \quad \text{where } A_{i,j} = \frac{1}{d} K(s_i, t_j).$$

To employ the Bayesian approach, we treat $y$, $\eta$, and $x$ as random variables. We assume that $\eta$ is Gaussian noise with variance $\sigma^2 I$,

$$\eta \sim \mathcal{N}(0, \sigma^2 I), \qquad \nu(\eta) \propto \exp\Big(-\frac{1}{2\sigma^2}\|\eta\|^2\Big).$$

The likelihood is then given by

$$\mathbb{P}(y|x) = \nu(y - Ax) \propto \exp\Big(-\frac{1}{2\sigma^2}\|y - Ax\|^2\Big).$$

Next, we have to choose a prior distribution for the unknown. Assume that we know that $x(0) = x(1) = 0$ and that $x$ is quite smooth, that is, the value of $x(t)$ in a point is more or less the same as in its neighbor. We will then model the unknown as

$$x_j = \frac{1}{2}(x_{j-1} + x_{j+1}) + W_j, \quad j = 1, \ldots, k, \tag{1}$$

where the term $W_j$ follows a Gaussian distribution $\mathcal{N}(0, \gamma^2)$.
The variance $\gamma^2$ determines how much the reconstructed function $x$ departs from the smoothness model $x_j = \frac{1}{2}(x_{j-1} + x_{j+1})$. We can write (1) as

$$Lx = W, \quad \text{where} \quad L := \frac{1}{2}\begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}.$$

This leads to the so-called *smoothness prior*

$$\pi(x) \propto \exp\Big(-\frac{1}{2\gamma^2}\|Lx\|^2\Big).$$

Using Bayes' formula, we get the posterior distribution

$$\pi^y(x) \propto \exp\Big( - \frac{1}{2\sigma^2}\|y - Ax\|^2 - \frac{1}{2\gamma^2}\|Lx\|^2\Big).$$

For the numerical experiment, we simulate measurements using the (smooth) ground truth signal

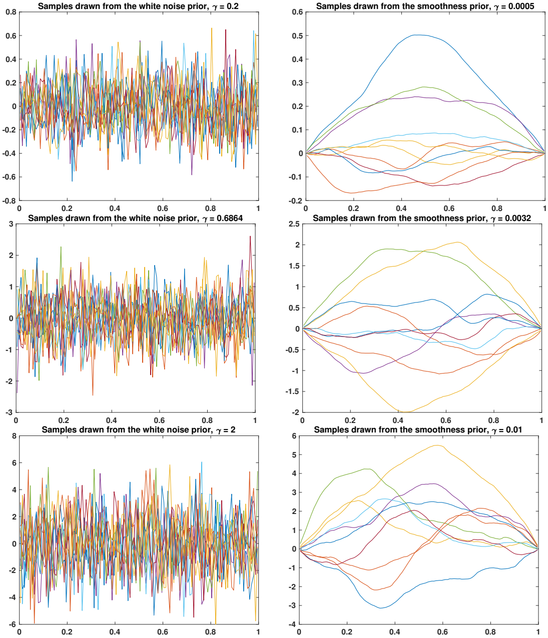
$$f(t) = 8t^3 - 16t^2 + 8t,$$

which satisfies $f(0) = f(1) = 0$. The measurements are contaminated with 10% *relative* noise ($\sigma \approx 0.0618$) and we set $d = k = 120$.

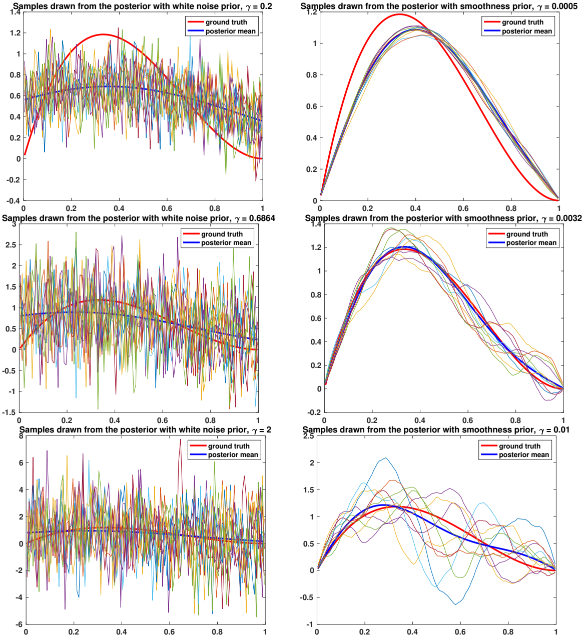
Let us draw samples from the prior and posterior. As comparison, we also consider a posterior obtained using the white noise prior, i.e.,

$$\pi_0^y(x) \propto \Bigg( - \frac{1}{2\sigma^2}\|y - Ax\|^2\Bigg)\pi_{\mathrm{pr},0}(x), \quad \pi_{\mathrm{pr},0}(x) \propto \exp\Bigg( - \frac{1}{2\gamma^2}\|x\|^2\Bigg).$$

**Remark:** Let us discuss the *implementational* details (sampling from Gaussian posterior distributions, formulae for the posterior means and variances of Gaussian posterior distributions) for this example in more detail *next week*.




Samples drawn from the white noise prior and the smoothness prior for several values of $\gamma$.




Samples drawn from the posterior corresponding to both the white noise prior and the smoothness prior for several values of $\gamma$. We also plot the ground truth solution and the posterior mean. The solutions in the middle row *roughly* satisfy the Morozov discrepancy principle.

As the previous example illustrates, many practical problems tend to be high-dimensional. The measurement model for the discretized deconvolution example

$$y = Ax + \eta,$$

with $A \in \mathbb{R}^{k \times d}$, $x \in \mathbb{R}^d$, and $y, \eta \in \mathbb{R}^k$, where $k$ corresponds to the number of points $s_1, \ldots, s_k$ where we observe the signal and $d$ corresponds to the number of quadrature points $t_1, \ldots, t_d$ discretizing the unknown quantity $x$.

A grid with only $k = d = 120$ points already corresponds to a 120-dimensional posterior, so visualization of the posterior density is highly nontrivial.

In practice, we are often interested in various point estimates, statistics, samples, or the spread of the posterior distribution.

# Bayesian estimators

The posterior distribution can be used to define estimators for the conditional random variable $x|y \sim \pi^y(x)$. In general, an estimator $\hat{x}$ is any function of the data $y$. The estimate $\hat{x}(y)$ is itself an $\mathbb{R}^d$-valued random variable whose properties give information about the usefulness and quality of the estimator.

Bayesian estimators are those defined via the posterior distribution $\pi^y$. We present the two most prominent ones. The *conditional mean (CM) estimator*, which is defined as the mean

$$\hat{x}_{\mathrm{CM}}(y) = \mathbb{E}[x|y] = \int_{\mathbb{R}^d} u\pi^y(u)\mathrm{d}u$$

of the posterior distribution.

The *maximum a posteriori (MAP) estimator*, which is defined as the mode

$$\hat{x}_{\mathrm{MAP}}(y) = \arg\max_{u\in\mathbb{R}^d} \pi^y(u)$$

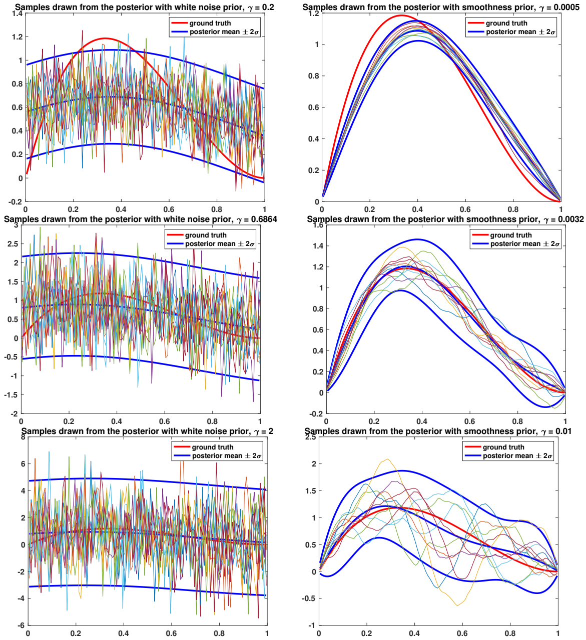
of the posterior distribution (if a unique mode exists).

One way to estimate spread are Bayesian *credible sets*. A level $1-\alpha$ credible set $\mathcal{C}_\alpha$ with $\alpha \in (0,1)$ satisfies

$$\mathbb{P}(x \in \mathcal{C}_\alpha | y) = \int_{\mathcal{C}_\alpha} \pi^y(u)\mathrm{d}u = 1-\alpha.$$

For small $\alpha$, it is a region that contains a large fraction of the posterior mass.

# **Deconvolution example:** posteriors with $2\sigma$ credibility envelopes.



**Example.** Assume that $x \in \mathbb{R}$ and that the posterior density is given by

$$\pi^y(u) = \frac{c}{\sigma_1}\phi\left(\frac{u}{\sigma_1}\right) + \frac{1-c}{\sigma_2}\phi\left(\frac{u-1}{\sigma_2}\right),$$

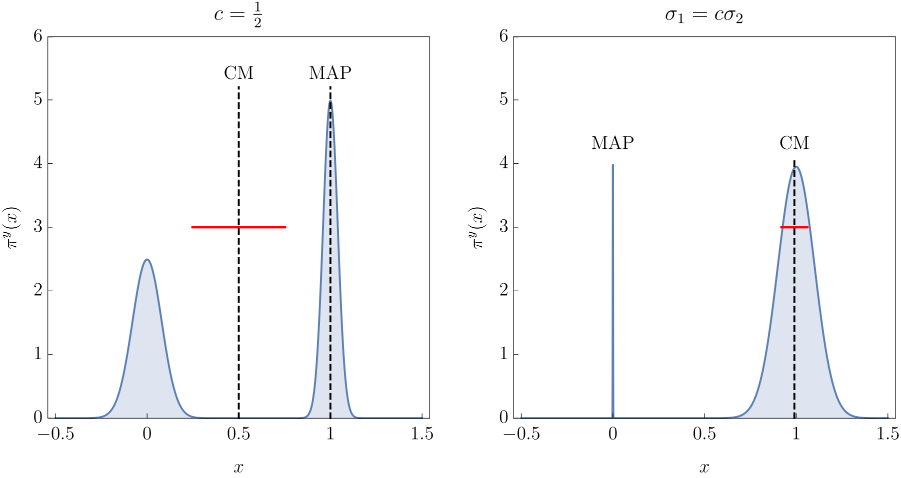
where $c \in (0,1)$, $\sigma_1, \sigma_2 > 0$, and $\phi$ is the density of the standard normal distribution, $\phi(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)$. In this case,

$$\hat{x}_{\text{CM}} = 1-c \quad \text{and} \quad \hat{x}_{\text{MAP}} = \begin{cases} 0 & \text{if } c/\sigma_1 > (1-c)/\sigma_2, \\ 1 & \text{if } c/\sigma_1 < (1-c)/\sigma_2. \end{cases}$$

If $c = \frac{1}{2}$ and $\sigma_1, \sigma_2$ are small, the probability that $x$ takes values near $\hat{x}_{\text{CM}}$ is small. On the other hand, if $\sigma_1 = c\sigma_2$, then $c/\sigma_1 = 1/\sigma_2 > (1-c)/\sigma_2$, so that $\hat{x}_{\text{MAP}} = 0$. If $c$ is small, this is, however, a bad estimate for $x$, since the probability for $x$ to take values near 0 is small. Last of all, we notice that when the conditional mean gives a poor estimate, this is reflected in a larger posterior variance

$$\sigma^2 = \int_{-\infty}^{\infty} (u - \hat{x}_{\text{CM}})^2 \pi^y(u)\mathrm{d}u.$$

We cannot say that one estimator is better than the other in all applications.




Left: the density with $\sigma_1 = 0.08$, $\sigma = 0.04$, and $c = \frac{1}{2}$. The CM estimate represents the distribution poorly. Notice that when the CM gives a poor estimate, this is reflected in wider variance (1 standard deviation is depicted as a red line). Right: the density with $\sigma_1 = 0.001$, $\sigma_2 = 0.1$, and $c = 0.01$. The MAP gives a poor estimate since it is in an unlikely part of the computational domain.

The maximum likelihood estimate

$$\hat{x}_{\mathrm{ML}}(y) = \arg\max_{u \in \mathbb{R}^d} \mathbb{P}(y|u)$$

answers the question: "which value of the unknown is most likely to produce the measured data?"

The ML estimate is a non-Bayesian estimate, and in the case of ill-posed inverse problems, often not useful. It is analogous to solving a classical inverse problem without regularization.

# Well-posedness

Assume that the posterior density is given by

$$\pi^y(x) = \frac{1}{Z} g(x)\pi(x)$$

with likelihood $g(x)$ and prior density $\pi(x)$. Now consider an approximation

$$\pi^y_\delta(x) = \frac{1}{Z_\delta} g_\delta(x)\pi(x)$$

resulting from an approximated likelihood $g_\delta(x)$. Such an approximation can result, for example, from an approximation $F_\delta$ of the forward operator $F$ or from perturbed data $y_\delta$.

The question is therefore:

$$\text{does} \quad |g - g_\delta| = \mathcal{O}(\delta) \quad \text{imply} \quad d(\pi^y, \pi^y_\delta) = \mathcal{O}(\delta)$$

for small enough $\delta > 0$ and some metric $d(\cdot,\cdot)$ on probability densities?

- Well-posedness refers to the continuity of the method of obtaining the posterior distribution with respect to different perturbations in the parameters. In practice, this could mean for example the following: If we have two measurements close to each other, does this mean the corresponding posterior distributions are close in some metric? Recall that ill-posed problems generally are discontinuous in this regard, i.e., without regularization, small difference in measurements can induce arbitrarily large difference in reconstruction. Does the Bayesian approach then regularize the problem? The answer is yes under certain assumptions on the modeling.
- We will proceed to show that, under certain conditions, $\pi^y$ and $\pi_\delta^y$ satisfy

  $$d(\pi^y, \pi_\delta^y) \leq c\delta$$

  for $\delta$ small enough, some $c > 0$, and some metric $d(\cdot, \cdot)$ on probability densities.
- To this end, we define two metrics for probability densities: the total variation distance and the Hellinger distance.

# Metrics for probability densities

We introduce the total variation distance and the Hellinger distance, both of which have been used to show well-posedness results. Here, we will use the Hellinger distance to establish the well-posedness of Bayesian inverse problems.

Let $\pi$ and $\pi'$ be the probability densities of two random variables with values in $\mathbb{R}^d$. We define the *total variation distance* between $\pi$ and $\pi'$ as

$$d_{\mathsf{TV}}(\pi, \pi') = \frac{1}{2} \int_{\mathbb{R}^d} \left|\pi(x) - \pi'(x)\right| \mathrm{d}x = \frac{1}{2} \|\pi - \pi'\|_{L^1},$$

and the *Hellinger distance* between $\pi$ and $\pi'$ as

$$d_{\mathsf{H}}(\pi, \pi') = \left(\frac{1}{2} \int_{\mathbb{R}^d} \left|\sqrt{\pi(x)} - \sqrt{\pi'(x)}\right|^2 \mathrm{d}x\right)^{\frac{1}{2}} = \frac{1}{\sqrt{2}} \left\|\sqrt{\pi} - \sqrt{\pi'}\right\|_{L^2}.$$

The normalization constants are chosen in such a way that the largest possible distance between two densities is one, as can be seen in the following lemma.

Lemma

*For any two probability densites $\pi$ and $\pi'$,*

$$0 \leq d_{\mathsf{TV}}(\pi, \pi') \leq 1 \quad \textit{and} \quad 0 \leq d_{\mathsf{H}}(\pi, \pi') \leq 1.$$

*Proof.* The lower bounds follow immediately from the definition of $d_{\mathsf{TV}}$ and $d_{\mathsf{H}}$. It remains to prove the upper bounds. To this end, we estimate

$$d_{\mathsf{TV}}(\pi, \pi') = \frac{1}{2}\int_{\mathbb{R}^d} |\pi(x) - \pi'(x)| \mathrm{d}x \leq \frac{1}{2}\int_{\mathbb{R}^d} \pi(x)\mathrm{d}x + \frac{1}{2}\int_{\mathbb{R}^d} \pi'(x)\mathrm{d}x = 1$$

and

$$\begin{aligned} d_{\mathsf{H}}(\pi, \pi')^2 &= \frac{1}{2}\int_{\mathbb{R}^d} \left|\sqrt{\pi(x)} - \sqrt{\pi'(x)}\right|^2 \mathrm{d}x \\ &= \frac{1}{2}\int_{\mathbb{R}^d} \left(\pi(x) + \pi'(x) - 2\sqrt{\pi(x)\pi'(x)}\right) \mathrm{d}x \\ &\leq \frac{1}{2}\int_{\mathbb{R}^d} \left(\pi(x) + \pi'(x)\right) \mathrm{d}x = 1. \quad \square \end{aligned}$$

In what follows, we will establish bounds between Hellinger and total variation distance and show how both distances can be used to bound the difference of expected values with respect to two different densities; these results will be useful in subsequent lectures.

### Lemma

*For any two probability densities $\pi$ and $\pi'$, the total variation and Hellinger distance are related by the inequalities*

$$\frac{1}{\sqrt{2}} d_{\mathsf{TV}}(\pi, \pi') \le d_{\mathsf{H}}(\pi, \pi') \le \sqrt{d_{\mathsf{TV}}(\pi, \pi')}.$$

*Proof.* Using the Cauchy–Schwarz inequality and $(a+b)^2 \le 2a^2 + 2b^2$ leads to

$$
\begin{aligned}
d_{\mathsf{TV}}(\pi, \pi') &= \frac{1}{2}\int_{\mathbb{R}^d} \left|\sqrt{\pi(x)} - \sqrt{\pi'(x)}\right| \cdot \left|\sqrt{\pi(x)} + \sqrt{\pi'(x)}\right| \mathrm{d}x \\
&\le \left(\frac{1}{2}\int_{\mathbb{R}^d} \left|\sqrt{\pi(x)} - \sqrt{\pi'(x)}\right|^2 \mathrm{d}x\right)^{\frac{1}{2}} \left(\frac{1}{2}\int_{\mathbb{R}^d} \left|\sqrt{\pi(x)} + \sqrt{\pi'(x)}\right|^2 \mathrm{d}x\right)^{\frac{1}{2}} \\
&\le d_{\mathsf{H}}(\pi, \pi') \left(\frac{1}{2}\int_{\mathbb{R}^d} \left(2\pi(x) + 2\pi'(x)\right) \mathrm{d}x\right)^{\frac{1}{2}} = \sqrt{2} d_{\mathsf{H}}(\pi, \pi').
\end{aligned}
$$

Notice that $|\sqrt{\pi(x)} - \sqrt{\pi'(x)}| \le |\sqrt{\pi(x)} + \sqrt{\pi'(x)}|$, since $\sqrt{\pi(x)}, \sqrt{\pi'(x)} \ge 0$. Thus, we have

$$
\begin{aligned}
d_{\mathsf{H}}(\pi, \pi')^2 &= \frac{1}{2}\int_{\mathbb{R}^d} \left|\sqrt{\pi(x)} - \sqrt{\pi'(x)}\right|^2 \mathrm{d}x \\
&\le \frac{1}{2}\int_{\mathbb{R}^d} \left|\sqrt{\pi(x)} - \sqrt{\pi'(x)}\right| \cdot \left|\sqrt{\pi(x)} + \sqrt{\pi'(x)}\right| \mathrm{d}x \\
&= \frac{1}{2}\int_{\mathbb{R}^d} \left|\pi(x) - \pi'(x)\right| \mathrm{d}x = d_{\mathsf{TV}}(\pi, \pi'). \quad \square
\end{aligned}
$$

The following lemmata show that if two densities are close in total variation or Hellinger distance, expectations computed with respect to both densities are also close.

## Lemma

*Let $f$ be a real valued function on $\mathbb{R}^d$ such that $\mathbb{E}^{\pi}[f^2] + \mathbb{E}^{\pi'}[f^2] =: f_2^2 < \infty$, then*

$$\left|\mathbb{E}^{\pi}[f] - \mathbb{E}^{\pi'}[f]\right| \leq 2 f_2 d_{\mathsf{H}}(\pi, \pi'). \tag{2}$$

*Proof.* We estimate

$$\left|\mathbb{E}^{\pi}[f] - \mathbb{E}^{\pi'}[f]\right| = \left|\int_{\mathbb{R}^d} f(x)\left(\pi(x) - \pi'(x)\right) \mathrm{d}x\right|$$

$$= \left|\int_{\mathbb{R}^d} f(x)\left(\sqrt{\pi(x)} - \sqrt{\pi'(x)}\right)\left(\sqrt{\pi(x)} + \sqrt{\pi'(x)}\right) \mathrm{d}x\right|$$

$$\leq \left(\frac{1}{2}\int_{\mathbb{R}^d}\left|\sqrt{\pi(x)} - \sqrt{\pi'(x)}\right|^2 \mathrm{d}x\right)^{\frac{1}{2}} \left(2\int_{\mathbb{R}^d} |f(x)|^2 \left|\sqrt{\pi(x)} + \sqrt{\pi'(x)}\right|^2 \mathrm{d}x\right)^{\frac{1}{2}}$$

$$\leq d_{\mathsf{H}}(\pi, \pi')\left(4\int_{\mathbb{R}^d} |f(x)|^2 \left(\pi(x) + \pi'(x)\right) \mathrm{d}x\right)^{\frac{1}{2}} = 2 f_2 d_{\mathsf{H}}(\pi, \pi'). \qquad \square$$

### Lemma

*Let $f$ be a real valued function on $\mathbb{R}^d$ such that $\sup_{x \in \mathbb{R}^d} |f(x)| =: \|f\|_\infty < \infty$, then*

$$\left| \mathbb{E}^{\pi}[f] - \mathbb{E}^{\pi'}[f] \right| \leq 2\|f\|_\infty d_{\mathrm{TV}}(\pi, \pi').$$

*Moreover, the following variational characterization of the total variation distance holds:*

$$d_{\mathrm{TV}}(\pi, \pi') = \frac{1}{2} \sup_{\|f\|_\infty \leq 1} \left| \mathbb{E}^{\pi}[f] - \mathbb{E}^{\pi'}[f] \right|.$$

**Remark:** Note that the result for the Hellinger distance only assumes that $f$ is square integrable with respect to $\pi$ and $\pi'$, whereas the result for the total variation distance requires that $f$ is bounded.

*Proof.* For the first part of the lemma, note that

$$\begin{aligned}\left|\mathbb{E}^{\pi}[f]-\mathbb{E}^{\pi'}[f]\right| &= \left|\int_{\mathbb{R}^d} f(x)\big(\pi(x)-\pi'(x)\big)\mathrm{d}x\right| \\ &\le 2\|f\|_\infty \cdot \frac{1}{2}\int_{\mathbb{R}^d} |\pi(x)-\pi'(x)|\mathrm{d}x = 2\|f\|_\infty d_{\mathsf{TV}}(\pi,\pi').\end{aligned}$$

This in particular shows that, for any $f$ with $\|f\|_\infty = 1$,

$$d_{\mathsf{TV}}(\pi,\pi') \ge \frac{1}{2}\left|\mathbb{E}^{\pi}[f]-\mathbb{E}^{\pi'}[f]\right|.$$

Our goal now is to show a choice of $f$ with $\|f\|_\infty = 1$ that achieves equality. Define $f(x) := \mathrm{sign}\Big(\pi(x)-\pi'(x)\Big)$, so that $f(x)\Big(\pi(x)-\pi'(x)\Big) = |\pi(x)-\pi'(x)|$. Then, $\|f\|_\infty = 1$ and

$$\begin{aligned}d_{\mathsf{TV}}(\pi,\pi') &= \frac{1}{2}\int_{\mathbb{R}^d} |\pi(x)-\pi'(x)|\,\mathrm{d}x = \frac{1}{2}\int_{\mathbb{R}^d} f(x)\Big(\pi(x)-\pi'(x)\Big)\,\mathrm{d}x \\ &= \frac{1}{2}\left|\mathbb{E}^{\pi}[f]-\mathbb{E}^{\pi'}[f]\right|.\end{aligned}$$

This completes the proof of the variational characterization. □

# Approximation theorem

We denote by

$$g(x) = \nu(y - F(x)) \quad \text{and} \quad g_\delta(x) = \nu(y - F_\delta(x))$$

the likelihoods associated with $F$ and $F_\delta$, so that

$$\pi^y(x) = \frac{1}{Z} g(x)\pi(x) \quad \text{and} \quad \pi^y_\delta(x) = \frac{1}{Z_\delta} g_\delta(x)\pi(x)$$

with corresponding normalizing constants $Z, Z_\delta > 0$. We make the following assumptions on $g$ and $g_\delta$.

**Assumption 1.** There exist $\delta^+ > 0$, constants $K_1, K_2 > 0$, and a function $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ such that $\mathbb{E}^\pi[\varphi^2] \leq K_1$ and for all $\delta \in (0, \delta^+)$,

1. $\left|\sqrt{g(x)} - \sqrt{g_\delta(x)}\right| \leq \varphi(x)\delta$ for all $x \in \mathbb{R}^d$,
2. $\left|\sqrt{g(x)}\right| + \left|\sqrt{g_\delta(x)}\right| \leq K_2$ for all $x \in \mathbb{R}^d$.

### Lemma

*Under* **Assumption 1** *there exist* $\tilde{\delta}^+ > 0$, $c_1, c_2 \in (0, +\infty)$ *such that*

$$|Z - Z_\delta| \leq c_1 \delta \quad \textit{and} \quad Z, Z_\delta > c_2, \quad \textit{for } \delta \in (0, \tilde{\delta}^+).$$

*Proof.* Since $Z = \int_{\mathbb{R}^d} g(x)\pi(x)\mathrm{d}x$ and $Z_\delta = \int_{\mathbb{R}^d} g_\delta(x)\pi(x)\mathrm{d}x$ we have

$$|Z - Z_\delta| = \left| \int_{\mathbb{R}^d} (g(x) - g_\delta(x))\pi(x)\mathrm{d}x \right|$$

$$\leq \left( \int_{\mathbb{R}^d} \left| \sqrt{g(x)} - \sqrt{g_\delta(x)} \right|^2 \pi(x)\mathrm{d}x \right)^{\frac{1}{2}} \left( \int_{\mathbb{R}^d} \left| \sqrt{g(x)} + \sqrt{g_\delta(x)} \right|^2 \pi(x)\mathrm{d}x \right)^{\frac{1}{2}}$$

$$\leq \left( \int_{\mathbb{R}^d} \delta^2 \phi(x)^2 \pi(x)\mathrm{d}x \right)^{\frac{1}{2}} \left( \int_{\mathbb{R}^d} K_2^2 \pi(x)\mathrm{d}x \right)^{\frac{1}{2}}$$

$$\leq \sqrt{K_1} K_2 \delta, \quad \delta \in (0, \delta^+).$$

And when $\delta \leq \tilde{\delta}^+ := \min\{\frac{Z}{2\sqrt{K_1}K_2}, \delta^+\}$, we have

$$Z_\delta \geq Z - |Z - Z_\delta| \geq \frac{1}{2} Z.$$

The lemma follows by taking $c_1 = \sqrt{K_1} K_2$ and $c_2 = \frac{1}{2} Z$. □

### Theorem (Well-posedness)

*Under* **Assumption 1**, *there exist* $\widetilde{\delta}^+ > 0$ *and* $c > 0$ *such that*

$$d_{\mathrm{H}}(\pi^y, \pi^y_\delta) \leq c\delta \quad \textit{for all } \delta \in (0, \widetilde{\delta}^+).$$

*Proof.* We break the distance into two error parts, one caused by the difference between $Z$ and $Z_\delta$, the other caused by the difference between $g$ and $g_\delta$:

$$\begin{aligned} d_{\mathrm{H}}(\pi^y, \pi^y_\delta) &= \frac{1}{\sqrt{2}} \left\| \sqrt{\pi^y} - \sqrt{\pi^y_\delta} \right\|_{L^2} \\ &= \frac{1}{\sqrt{2}} \left\| \sqrt{\frac{g\pi}{Z}} - \sqrt{\frac{g\pi}{Z_\delta}} + \sqrt{\frac{g\pi}{Z_\delta}} - \sqrt{\frac{g_\delta \pi}{Z_\delta}} \right\|_{L^2} \\ &\leq \frac{1}{\sqrt{2}} \left\| \sqrt{\frac{g\pi}{Z}} - \sqrt{\frac{g\pi}{Z_\delta}} \right\|_{L^2} + \frac{1}{\sqrt{2}} \left\| \sqrt{\frac{g\pi}{Z_\delta}} - \sqrt{\frac{g_\delta \pi}{Z_\delta}} \right\|_{L^2}. \end{aligned}$$

On the previous slide, we obtained

$$d_{\mathrm{H}}(\pi^y, \pi^y_\delta) \leq \frac{1}{\sqrt{2}}\Big\|\sqrt{\frac{g\pi}{Z}} - \sqrt{\frac{g\pi}{Z_\delta}}\Big\|_{L^2} + \frac{1}{\sqrt{2}}\Big\|\sqrt{\frac{g\pi}{Z_\delta}} - \sqrt{\frac{g_\delta\pi}{Z_\delta}}\Big\|_{L^2}.$$

Using the previous Lemma, for $\delta \in (0, \tilde{\delta}^+)$, we have for the first term

$$\Big\|\sqrt{\frac{g\pi}{Z}} - \sqrt{\frac{g\pi}{Z_\delta}}\Big\|_{L^2} = \Big|\frac{1}{\sqrt{Z}} - \frac{1}{\sqrt{Z_\delta}}\Big|\underbrace{\Big(\int_{\mathbb{R}^d} g(x)\pi(x)\mathrm{d}x\Big)^{\frac{1}{2}}}_{=\sqrt{Z}}$$

$$= \Big|1 - \frac{\sqrt{Z}}{\sqrt{Z_\delta}}\Big| = \Big|\frac{\sqrt{Z_\delta} - \sqrt{Z}}{\sqrt{Z_\delta}}\Big| = \frac{|Z - Z_\delta|}{(\sqrt{Z} + \sqrt{Z_\delta})\sqrt{Z_\delta}} \leq \frac{c_1}{2c_2}\delta.$$

For the second term, we obtain

$$\Big\|\sqrt{\frac{g\pi}{Z_\delta}} - \sqrt{\frac{g_\delta\pi}{Z_\delta}}\Big\|_{L^2} = \frac{1}{\sqrt{Z_\delta}}\Big(\int_{\mathbb{R}^d}\Big|\sqrt{g(x)} - \sqrt{g_\delta(x)}\Big|^2\pi(x)\mathrm{d}x\Big)^{\frac{1}{2}} \leq \sqrt{\frac{K_1}{c_2}}\delta.$$

Therefore

$$d_{\mathrm{H}}(\pi^y, \pi^y_\delta) \leq \frac{1}{\sqrt{2}}\frac{c_1}{2c_2}\delta + \frac{1}{\sqrt{2}}\sqrt{\frac{K_1}{c_2}}\delta = c\delta,$$

with $c = \frac{1}{\sqrt{2}}\frac{c_1}{2c_2} + \frac{1}{\sqrt{2}}\sqrt{\frac{K_1}{c_2}}$ independent of $\delta$. □

Notice that, together with (2), i.e., the inequality

$$\left|\mathbb{E}^{\pi}[f] - \mathbb{E}^{\pi'}[f]\right| \leq 2f_2 d_{\mathsf{H}}(\pi, \pi'), \quad f_2^2 := \mathbb{E}^{\pi}[f^2] + \mathbb{E}^{\pi'}[f^2],$$

this theorem guarantees that expectations computed with respect to $\pi^y$ and $\pi^y_\delta$ are in the order of $\delta$ apart.