The Intrinsic Dimension of Importance Sampling

Sergios Agapiou
(University of Cyprus)

With D. Sanz-Alonso (Brown), O. Papaspiliopoulos (UPF)
and A. Stuart (Caltech)

Summary

"*Our purpose in this paper is to overview various ways of measuring the computational complexity of importance sampling, to link them to one another through transparent mathematical reasoning, and to create cohesion in the vast published literature on this subject. In addressing these issues we will study importance sampling in a general abstract setting, and then in the particular cases of Bayesian inversion and filtering.*"

General Framework

$$\mu(du) \propto g(u)\pi(du)$$

μ, π in arbitrary $(\mathcal{X}, \mathcal{F})$

Autonormalized Importance Sampling (IS)

$$\mu(\phi) = \frac{\pi(\phi g)}{\pi(g)} \approx \frac{\frac{1}{N}\sum_{i=1}^{N} \phi(u^i) g(u^i)}{\frac{1}{N}\sum_{j=1}^{N} g(u^j)}, \quad u^i \overset{iid}{\sim} \pi,$$

$$= \sum_{i=1}^{N} w^i \phi(u^i) =: \mu^N(\phi),$$

where

$$\mu^N = \sum_{i=1}^{N} w^i \delta_{u^i}, \qquad w^i = \frac{g(u^i)}{\sum_j g(u^j)}, \qquad g \propto \frac{d\mu}{d\pi}.$$

Non-asymptotic Theorem; $\mu(du) \propto g(u)\pi(du)$

$$\rho := \frac{\pi(g^2)}{\pi(g)^2} \in [1, \infty].$$

Theorem (A., Papaspiliopoulos, Sanz-Alonso, Stuart '15)

$$d(\mu^N, \mu)^2 := \sup_{|\phi|\le 1} \mathbb{E}\Big[\big(\mu^N(\phi) - \mu(\phi)\big)^2\Big] \le \frac{4}{N}\rho = \frac{4}{N}\Big(1 + D_{\chi^2}(\mu||\pi)\Big).$$

Remarks

- Estimate is known to be asymptotically sharp (Slutsky's lemmas)

$$\sqrt{N}\big(\mu^N(\phi) - \mu(\phi)\big) \Rightarrow N\left(0, \frac{\pi(g^2\overline{\phi}^2)}{\pi(g)^2}\right), \quad \text{where } \overline{\phi} := \phi - \mu(\phi)$$

  This leads to the same bound if $|\phi| \leq 1$.

- Result relates to existing ones by Del Moral (covering numbers vs ρ), and by Crisan and collaborators

- We also show a non-asymptotic result on the MSE with less assumptions on test functions and more on g; based on ideas for calculating moments of ratios of estimators.

  Doukhan and Lang, Evaluation of moments of a ratio with application to regression estimation, 2009.

Weight Collapse: Unbounded Degrees of Freedom

$$\pi_d(du) = \prod_{i=1}^{d} \pi_1(du(i)), \;\; \mu_d(du) = \prod_{i=1}^{d} \mu_1(du(i)), \;\; \frac{d\mu_1}{d\pi_1} \propto g_1, \;\; \frac{d\mu_d}{d\pi_d} \propto g_d.$$

Then,

$$\rho_d := \frac{\pi_d(g_d^2)}{\pi_d(g_d)^2} = \left(\frac{\pi_1(g_1^2)}{\pi_1(g_1)^2}\right)^d.$$

In fact, it is known that $w^{(N)} \to 1$ as $d, N \to \infty$, unless N grows exponentially with d.

Bengtsson, Bickel, and Li, Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems, 2008.

Key: $\frac{1}{d}\sum_{i=1}^{d} u(i)$ has different a.s. limits under μ_∞ and π_∞, hence μ_∞ and π_∞ are singular.

Weight Collapse: Singular Limits

Suppose π fixed and let μ_γ

$$\frac{d\mu_\gamma}{d\pi} \propto g_\gamma(u) := \exp\bigl(\gamma^{-1}h(u)\bigr),$$

where h is smooth and has a unique minimum at u^*.

Laplace approximation yields, as $\gamma \to 0$,

$$\rho_\gamma = \frac{\pi(g_\gamma^2)}{\pi(g_\gamma)^2} \approx \sqrt{\frac{h''(u^*)}{4\pi\gamma}}.$$

Key: In this example $\mu = \mu_0$ and π are (formally) singular.

Bayesian Inverse Problems

$$\mu(du) \propto \underbrace{\exp\left(\left|y - \mathcal{G}(u)\right|_{\Gamma}^{2}\right)}_{g(u)} \pi(du)$$

Linear Bayesian Inverse Problems in Euclidean Space

Interested in recovering $u \in \mathbb{R}^{d_u}$ from data $y \in \mathbb{R}^{d_y}$.

Bayesian Inverse Problem

$$\left.\begin{aligned} &\text{Prior (proposal)}: u \sim \pi = N(0, \Sigma) \\ &\text{Data}: y = Ku + \eta, \quad \eta \sim N(0, \Gamma) \end{aligned}\right\} u|y \sim \mu = N(m, C).$$

Linear Bayesian Inverse Problems in Euclidean Space

Interested in recovering $u \in \mathbb{R}^{d_u}$ from data $y \in \mathbb{R}^{d_y}$.

Bayesian Inverse Problem

$$\left.\begin{aligned} &\text{Prior (proposal)}: u \sim \pi = N(0, \Sigma) \\ &\text{Data}: y = Ku + \eta, \quad \eta \sim N(0, \Gamma) \end{aligned}\right\} u|y \sim \mu = N(m, C).$$

First aim: relate two notions of intrinsic dimension that account for:

- the nominal dimension $d := \min\{d_u, d_y\}$.
- the size of the observational noise.
- the regularity of the prior and forward map relative to observation noise.

Intrinsic Dimension

Key: Eigenvalues of $A := \Sigma^{1/2} K^* \Gamma^{-1} K \Sigma^{1/2}$.

$$\text{efd} \overset{(*)}{:=} \text{Tr}\left((I + A)^{-1} A \right) \qquad \tau \overset{(**)}{:=} \text{Tr}\,(A)$$

Lemma 1 (A., Papaspiliopoulos, Sanz-Alonso, Stuart '15)

$$\frac{1}{\|I + A\|}\tau \le \text{efd} \le \tau.$$

(*) Spiegelhalter, Best, Carlin, Van der Linde, Bayesian measures of model complexity and fit, 2002.

(**) Bengtsson, Bickel, and Li, Curse-of-dimensionality revisited: Collapse of the particle filter in very large systems, 2008.

Linear Inverse Problems in Hilbert Space

Second aim: relate intrinsic dimension and ρ in settings with infinite nominal dimension.

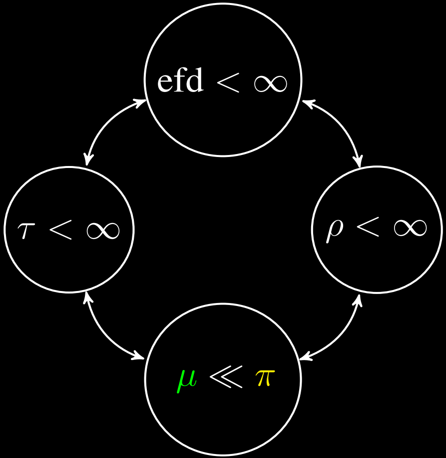
Linear Inverse Problems in Hilbert Space

Second aim: relate intrinsic dimension and ρ in settings with infinite nominal dimension.

$$y = Ku + \eta.$$

- Underlying separable Hilbert space $\mathcal{H} \longrightarrow$ KL expansion.
- $K, \Gamma, \Sigma : \mathcal{H} \to \mathcal{H}$ positive, self-adjoint.
- Realizations might be on larger spaces than $\mathcal{H}$.
- Intrinsic dimensions efd and τ defined in the same way.

Theorem: Link Between τ, efd, ρ and $\mu \ll \pi$



Diagonal Inverse Problems

$$y = Ku + \eta, \quad \eta \sim N(0, \Gamma), \quad u \sim N(0, \Sigma).$$

Assumption

- $\Gamma = \gamma I$.
- Eigenvalues of $A = \Sigma^{\frac{1}{2}} K^* K \Sigma^{\frac{1}{2}} : \left\{ \frac{j^{-\beta}}{\gamma} \right\}_{j=1}^{d}, \quad \beta \geq 0.$

$$\tau = \tau(d, \gamma, \beta), \quad \text{efd} = \text{efd}(d, \gamma, \beta), \quad \rho = \rho(d, \gamma, \beta).$$

$$\tau(\infty, \gamma, \beta) = \frac{1}{\gamma} \sum_{j=1}^{\infty} j^{-\beta} < \infty \iff \underline{\underline{\beta > 1}} \iff \mu_\infty \ll \pi_\infty.$$

Absolute continuity is also lost in the formal limit $\underline{\underline{\gamma = 0}}$.

Scalings

- ρ grows **algebraically** in the small noise limit ($\gamma \to 0$) if the nominal dimension d is finite.
- ρ grows **exponentially** in τ or efd as the nominal dimension grows ($d \to \infty$), or as the prior becomes rougher ($\beta \searrow 1$).

  Bengtsson, Bickel, and Li, Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems, 2008.
- ρ grows **factorially** in the small noise limit ($\gamma \to 0$) if $d = \infty$. The exponent in the rates relates naturally to efd.

Highlights

1. GENERAL FRAMEWORK
   - Non-asymptotic concentration inequalities for importance sampling. Highlighted importance of ρ.

2. INVERSE PROBLEMS
   - Established connection between notions of dimension: efd and τ.
   - Linked intrinsic dimensions to importance sampling through ρ.
   - Link with absolute continuity in Hilbert space setting.

3. FILTERING
   - Theory from inverse problems used to compare proposals.

References

- S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, A. M. Stuart, *Importance sampling: intrinsic dimension and computational cost*, to appear in Statistical Science.

- T. Bengtsson, P. Bickel, B. Li, *Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems*, Probability and Statistics: essays, 2008.

- A.M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numerica, 2010.

- D. Sanz-Alonso, *Importance sampling and necessary sample size: an information theory approach*, arXiv:1608.08814.