Sharper rates for estimating differential entropy under Gaussian convolutions

Jonathan Weed∗
Department of Mathematics, MIT
December 15, 2018

Abstract
In this short note, we show that, given access to n i.i.d. samples from a compactly supported d-dimensional distribution P, the differential entropy of P convolved with an isotropic Gaussian can be estimated at the rate $O(n^{1/2})$ by a plug-in estimator. This answers a question of Goldfeld et al. (2018).

We consider the following problem: given i.i.d. samples from a distribution $P$ on $[-1,1]^d$, how well can one estimate the differential entropy of $P$ convolved with an isotropic Gaussian? If we denote by $N_\sigma$ the distribution $N(0,\sigma^2 I_d)$ and by $\ast$ the convolution operator, Goldfeld et al. (2018) recently showed that there exists a simple estimator which converges to $h(P \ast N_\sigma)$ at nearly the parametric rate. Indeed, writing $P_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$, where $X_i \sim P$ i.i.d., they showed (Goldfeld et al., 2018, Theorem 2) that the plug-in estimator $h(P_n \ast N_\sigma)$ achieves:

$$E| h(P_n \ast N_\sigma) - h(P \ast N_\sigma) | \leq c_{\sigma,d} \frac{(\log n)^{\frac{d}{4}}}{\sqrt{n}}. \tag{1}$$

This rate is striking in that it is significantly better than what could be achieved by a generic estimator using samples from $P \ast N_\sigma$ alone (see, e.g., Han et al., 2017). In the interest of obtaining sharp rates, Goldfeld et al. (2018) posed the question of whether the logarithmic term $(\log n)^{\frac{d}{4}}$ could be improved to $(\log n)^c$ for some universal constant c.

In this note, we answer this question in the affirmative, showing in fact that the plug-in estimator $h(P_n \ast N_\sigma)$ achieves exactly the parametric rate, without logarithmic factors.

Theorem 1. For any distribution $P$ supported on $[-1,1]^d$, we have

$$E| h(P_n \ast N_\sigma) - h(P \ast N_\sigma) | \leq c_{\sigma,d} \frac{1}{\sqrt{n}},$$

for $c_{\sigma,d} := \frac{\sqrt{d} \cdot 2^{d+3}}{\min\{\sigma^2,\sigma^{d+1}\}}$.

The proof of Theorem 1 relies on the following proposition. Denote by $W_1(P,Q)$ the 1-Wasserstein distance between $P$ and $Q$, i.e., $W_1(P,Q) := \inf_\gamma \int \|x-y\| d\gamma(x,y)$, where the infimum is taken over all couplings of $P$ and $Q$.

Proposition 1. If $P$ is supported on $[-1,1]^d$, then

$$EW_1(P_n \ast N_\sigma, P \ast N_\sigma) \leq c'_{\sigma,d} \frac{1}{\sqrt{n}},$$

for $c'_{\sigma,d} := \frac{\sqrt{d} \cdot 2^{d+2}}{\min\{1,\sigma^d\}}$.

∗Email: jweed@mit.edu. Supported in part by the Josephine de Kármán Fellowship.
To connect Proposition 1 to the question of entropy estimation, we employ the following result due to Polyanskiy and Wu (2016).

**Proposition 2** [Polyanskiy and Wu 2016 Proposition 5]. Let $P$ and $Q$ be distributions supported on $[-1, 1]^d$, with $v_P := \mathbb{E}_{X \sim P} \|X\|^2$ and $v_Q := \mathbb{E}_{X \sim Q} \|X\|^2$. Then

$$|h(Q * N_\sigma) - h(P * N_\sigma)| \leq \frac{1}{2\sigma^2} (|v_Q - v_P| + 2\sqrt{d}W_1(Q * N_\sigma, P * N_\sigma)).$$

When $Q = P_n$, Jensen’s inequality implies $\mathbb{E}|v_{P_n} - v_P| \leq \frac{1}{\sqrt{n}} \text{var}_{X \sim P}([-1, 1]^d)^{1/2} \leq d/\sqrt{n}$. Hence, Theorem 1 follows directly from Propositions 1 and 2. It therefore suffices to give a proof of Proposition 1.

**Proof of Proposition 1**. Denote by $f$ the density of $P \ast N_\sigma$, and by $f_n$ the density of $P_n \ast N_\sigma$. We let $\phi_\sigma(x) := (2\pi\sigma^2)^{-d/2} \exp\left(-\frac{1}{2\sigma^2}\|x\|^2\right)$ be the density of $N_\sigma$. We use the following upper bound (Villani 2008 Theorem 6.15):

$$W_1(P_n \ast N_\sigma, P \ast N_\sigma) \leq \int_{\mathbb{R}^d} \|z\| |f_n(z) - f(z)| \, dz.$$

This yields

$$\mathbb{E}W_1(P_n \ast N_\sigma, P \ast N_\sigma) \leq \int_{\mathbb{R}^d} \|z\| \mathbb{E}|f_n(z) - f(z)| \, dz$$

$$= \int_{\mathbb{R}^d} \|z\| \left| \frac{1}{n} \sum_{i=1}^n \phi_\sigma(z - X_i) - \mathbb{E}\phi_\sigma(z - X) \right| \, dz$$

$$\leq \frac{1}{\sqrt{n}} \int_{\mathbb{R}^d} \|z\| \left( \mathbb{E}(\phi_\sigma(z - X) - \mathbb{E}\phi_\sigma(z - X))^2 \right)^{1/2} \, dz, \quad X \sim P$$

$$\leq \frac{1}{\sqrt{n}} \int_{\mathbb{R}^d} \|z\| \left( \mathbb{E}\phi_\sigma(z - X)^2 \right)^{1/2} \, dz.$$

When $z \in [-2, 2]^d$, we use the bound $(\mathbb{E}\phi_\sigma(z - X)^2)^{1/2} \leq \max_{z \in \mathbb{R}^d} \phi_\sigma(z) = (2\pi\sigma^2)^{-d/2}$. For $z \notin [-2, 2]^d$, we have $\|z - X\|^2 \geq \|z/2\|^2$ almost surely, which yields $(\mathbb{E}\phi_\sigma(z - X)^2)^{1/2} \leq \phi_\sigma(z/2)$. We obtain

$$\mathbb{E}W_1(P_n \ast N_\sigma, P \ast N_\sigma) \leq \frac{(2\pi\sigma^2)^{-d/2}}{\sqrt{n}} \int_{z \in [-2, 2]^d} \|z\| \, dz + \frac{1}{\sqrt{n}} \int_{z \in \mathbb{R}^d} \|z\| \phi_\sigma(z/2) \, dz$$

$$\leq \left( (2\pi\sigma^2)^{-d/2} \cdot 4^d \cdot 2 + 2^{d+1} \right) \cdot \sqrt{d/n}$$

$$\leq \max\{1, \sigma^{-d}\} 2^{d+2} \sqrt{d/n}. \quad \square$$

**Acknowledgements** The author thanks P. Rigollet and G. Mena for helpful discussions.

**References**


