# Finite-Sample Analysis of Fixed-k Nearest Neighbor Density Functional Estimators

Shashank Singh and Barnabás Póczos

Machine Learning Department, Carnegie Mellon University

## Introduction

- Many important statistical quantities are *functionals of a probability density*:
$$F(P) = \mathbb{E}_{X \sim P}\left[f(p(X))\right],$$
where $P$ is a probability measure with density function $p$.
  - Key examples include entropies, divergences, and mutual information.
  - Lack of practical and theoretically justified nonparametric estimators.
- Bias-corrected **k**NN estimators $\hat{F}_k$ perform very well empirically but **lack known convergence rates. We provide these.**

## Background: kNN Density Estimation

- Given IID samples $X_1, \dots, X_n \sim P$, $\hat{p}_k(x)$ estimates $p(x)$ via local linear approximation.

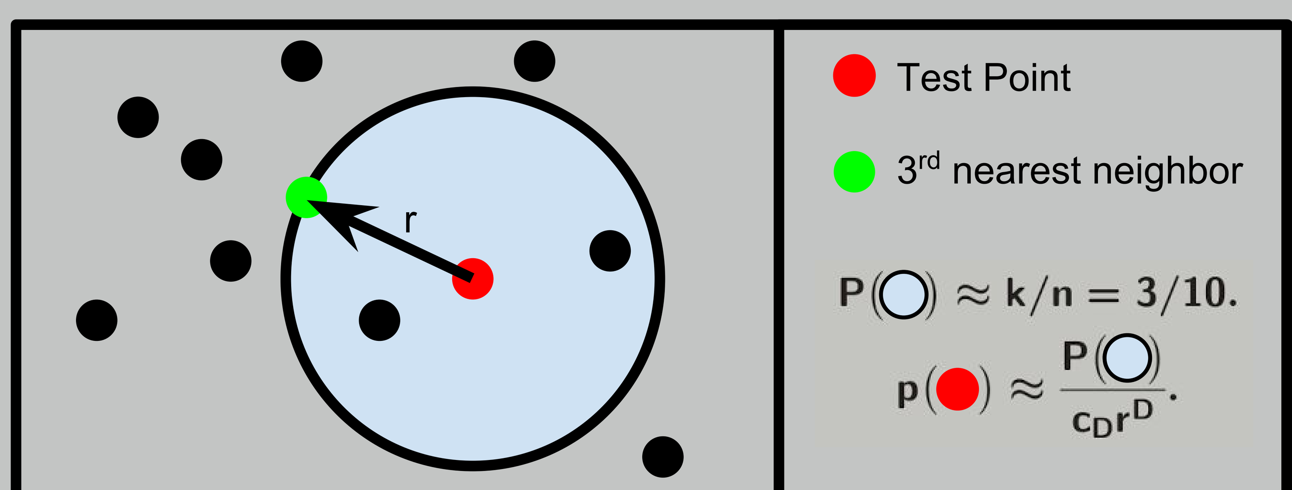


**Figure:** Illustration of **k**NN density estimation with $k = 3$, $n = 10$, $D = 2$.

- As $n \to \infty$, $k \to \infty$ is needed for $\mathbb{V}\left[\hat{p}_k(x)\right] \to 0$.
  - increases smoothing bias, slowing convergence of plug in estimate
- If we fix $k$, since $f$ is nonlinear, plug-in functional estimate is asymptotically biased
- However, we can analytically correct for this bias

## Bias-Corrected kNN Entropy Estimation

- Consider, e.g., estimating Shannon entropy [3, 2]:
$$F(P) = \mathbb{E}_{X \sim P}\left[\log p(X)\right].$$
- **Key Observation: Distribution of $P(\bigcirc)$ is independent of $p$.**
- Specifically, $\mathbb{E}\left[\log P(\bigcirc)\right] = \psi(k) - \psi(n)$ is known, and so
$$\begin{aligned} F(P) = \mathbb{E}\left[\log p(x)\right] &\approx \mathbb{E}\left[\log P(\bigcirc) - \log c_D - D \log r\right] \\ &= \psi(k) - \psi(n) - \log c_D - D\, \mathbb{E}\left[\log \varepsilon_k(X)\right] \end{aligned}$$
- Replacing $\mathbb{E}\left[\log \varepsilon_k(X)\right]$ with $\frac{1}{n}\sum_{i=1}^n \log \varepsilon_k(X_i)$ gives the estimate:
$$\hat{F}(P) = \psi(n) - \psi(k) + \log c_D + \frac{D}{n}\sum_{i=1}^{n} \log \varepsilon_k(X_i).$$

## Finite-sample behavior of kNN distances

- More general (asymptotic) bias corrections arise from the fact that $p(x)\varepsilon_k^D(x)n/k$ has an Erlang asymptotic distribution. [5]
  - This can be used to construct a bias correction $\mathcal{B}$ such that
$$\mathbb{E}\left[\mathcal{B}\left(f\left(\hat{p}_k(X)\right)\right)\right] = \mathbb{E}\left[f\left(\frac{P(B(X, \varepsilon_k(X)))}{\mu(B(X, \varepsilon_k(X)))}\right)\right].$$
- We provide two finite-sample versions of this fact:
  - We show $\varepsilon_k(x)$ is **tightly concentrated** about $\left(\frac{k}{p(x)n}\right)^{1/D}$.
    - As $r \to \infty$, $\mathbb{P}\left[\varepsilon_k(x) > r\right] \asymp r^{Dk} e^{-r^D}$.
    - As $r \to 0$, $\mathbb{P}\left[\varepsilon_k(x) < r\right] \asymp r^{Dk}$.
  - If $f : (0, \infty) \to \mathbb{R}$ is continuous and non-decreasing, then
$$\mathbb{E}\left[f\left(\varepsilon_k(x)\right)\right] \asymp f\left(\left(\frac{k}{np(x)}\right)^{1/D}\right).$$

## Main Results: Bias Bound

- Assume that
  - $p$ is $\beta$-Hölder continuous for some $\beta \in (0, 2]$
  - $p$ does not approach $0$ too quickly at the boundary of its support
    - Entropy and **k**NN distances are sensitive to low probability regions

  Then, the bias of $\hat{F}_k(P)$ is of order at most
$$\left|\mathbb{E}\left[\hat{F}_k(P)\right] - F(P)\right| \asymp \left(\frac{k}{n}\right)^{\beta/D} \tag{1}$$

## Main Results: Variance Bound

- Under very mild assumptions, we show that
$$\mathbb{V}\left[\hat{F}_k(P)\right] \asymp n^{-1}. \tag{2}$$
- Proof tricky because sum of **dependent** terms
- **Key Observation**: there exists constant $N_{k,D} \in \mathcal{Z}_+$ such that any $x \in \mathcal{X}$ can be within the **k**NN of at most $N_{k,D}$ other points

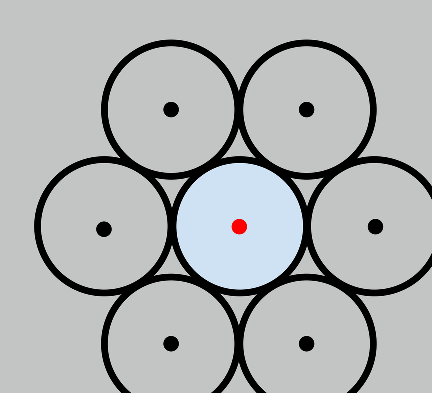

**Figure:** Illustration of $N_{1,2} = 6$.

- Hence, at most $N_{k,D}$ nearest neighbor distances depend on any sample.

## Conclusions

- Combining bias and variance bounds gives a mean squared error bound:
$$\mathbb{E}\left[\left(\hat{F}_{\mathcal{B}}(P) - F(P)\right)^2\right] \asymp \left(\frac{k}{n}\right)^{2\beta/D} + n^{-1}.$$
  - This suggests **fixed k** gives best MSE convergence rate
  - Gives rate $\asymp n^{-\min\left\{\frac{2\beta}{D}, 1\right\}}$, **competitive with best practical estimators**
  - Modifications are needed to leverage more smoothness (i.e., $\beta > 2$).
- Fixed-**k** estimators **automatically adapt** to unknown smoothness of **p**

## kNN Functional Estimators: Examples

- Our results apply to other examples of **k**NN functional estimators:

| Functional Name | Functional Form | Bias Correction | Ref. |
|---|---|---|---|
| Shannon Entropy | $\mathbb{E}\left[\log p(X)\right]$ | Additive constant: $\psi(n) - \psi(k) + \log(k/n)$ | [3][2] |
| Rényi-$\alpha$ Entropy | $\mathbb{E}\left[p^{\alpha-1}(X)\right]$ | Multiplicative constant: $\frac{\Gamma(k)}{\Gamma(k+1-\alpha)}$ | [5, 4] |
| KL Divergence | $\mathbb{E}\left[\log \frac{p(X)}{q(X)}\right]$ | None* | [7] |
| $\alpha$-Divergence | $\mathbb{E}\left[\left(\frac{p(X)}{q(X)}\right)^{\alpha-1}\right]$ | Multiplicative constant: $\frac{\Gamma^2(k)}{\Gamma(k-\alpha+1)\Gamma(k+\alpha-1)}$ | [6] |

## References

- Thomas B Berrett, Richard J Samworth, and Ming Yuan. Efficient multivariate entropy estimation via **k**-nearest neighbour distances. *arXiv preprint arXiv:1606.00304*, 2016.
- M. N. Goria, N. N. Leonenko, V. V. Mergel, and P. L. Novi Inverardi. A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *J. Nonparametric Stat.*, 17:277–297, 2005.
- L. F. Kozachenko and N. N. Leonenko. A statistical estimate for the entropy of a random vector. *Problems of Information Transmission*, 23:9–16, 1987.
- N. Leonenko and L. Pronzato. Correction of 'a class of Rényi information estimators for mulitidimensional densities' Ann. Statist., 36(2008) 2153-2182, 2010.
- N. Leonenko, L. Pronzato, and V. Savani. A class of Rényi information estimators for multidimensional densities. *Annals of Statistics*, 36(5):2153–2182, 2008.
- B. Poczos and J. Schneider. On the estimation of alpha-divergences. In *International Conference on AI and Statistics (AISTATS)*, volume 15 of *JMLR Workshop and Conference Proceedings*, pages 609–617, 2011.
- Q. Wang, S.R. Kulkarni, and S. Verdú. Divergence estimation for multidimensional densities via **k**-nearest-neighbor distances. *IEEE Transactions on Information Theory*, 55(5), 2009.

 sss1@andrew.cmu.edu