Log-concave density estimation: high dimensions

Richard J. Samworth

University of Cambridge

'Geometry of Big Data' workshop
IPAM

30 Apr 2019

The original problem

Let $X_1, \ldots, X_n$ be a random sample from a density f_0 on $\mathbb{R}^d$.

$\boxed{\text{How should we estimate } f_0\text{?}}$

The original problem

Let $X_1, \ldots, X_n$ be a random sample from a density f_0 on $\mathbb{R}^d$.

$\boxed{\text{How should we estimate } f_0\text{?}}$

Two main alternatives:

- Parametric models: use e.g. MLE. Assumptions often too restrictive.
- Nonparametric models: use e.g. kernel density estimate. Choice of bandwidth difficult, particularly for $d > 1$.

Shape-constrained estimation

Nonparametric shape constraints are becoming increasingly popular (Groeneboom et al. 2001, Walther 2002, Pal et al. 2007, Dümbgen and Rufibach 2009, Schuhmacher et al. 2011, Koenker and Mizera 2010, Groeneboom and Jongbloed, 2014, Baraud and Birgé, 2016, Doss and Wellner, 2016, Han and Wellner, 2016, Robeva et al., 2018, Carpenter et al., 2018, Dagan and Kur, 2019).

Shape-constrained estimation

Nonparametric shape constraints are becoming increasingly popular (Groeneboom et al. 2001, Walther 2002, Pal et al. 2007, Dümbgen and Rufibach 2009, Schuhmacher et al. 2011, Koenker and Mizera 2010, Groeneboom and Jongbloed, 2014, Baraud and Birgé, 2016, Doss and Wellner, 2016, Han and Wellner, 2016, Robeva et al., 2018, Carpenter et al., 2018, Dagan and Kur, 2019).

A density f is log-concave if $\log f$ is concave.

Shape-constrained estimation

Nonparametric shape constraints are becoming increasingly popular (Groeneboom et al. 2001, Walther 2002, Pal et al. 2007, Dümbgen and Rufibach 2009, Schuhmacher et al. 2011, Koenker and Mizera 2010, Groeneboom and Jongbloed, 2014, Baraud and Birgé, 2016, Doss and Wellner, 2016, Han and Wellner, 2016, Robeva et al., 2018, Carpenter et al., 2018, Dagan and Kur, 2019).

A density f is log-concave if $\log f$ is concave.

- Univariate examples: normal, logistic, Gumbel densities, as well as Weibull, Gamma, Beta densities for certain parameter values.

Shape-constrained estimation

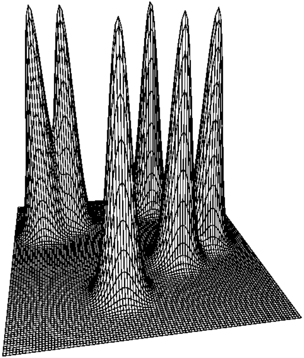
Nonparametric shape constraints are becoming increasingly popular (Groeneboom et al. 2001, Walther 2002, Pal et al. 2007, Dümbgen and Rufibach 2009, Schuhmacher et al. 2011, Koenker and Mizera 2010, Groeneboom and Jongbloed, 2014, Baraud and Birgé, 2016, Doss and Wellner, 2016, Han and Wellner, 2016, Robeva et al., 2018, Carpenter et al., 2018, Dagan and Kur, 2019).

A density f is log-concave if $\log f$ is concave.

- Univariate examples: normal, logistic, Gumbel densities, as well as Weibull, Gamma, Beta densities for certain parameter values.
- The class is closed under marginalisation, conditioning, convolution and linear transformations.

Unbounded likelihood!

Consider maximising the likelihood $L(f) := \prod_{i=1}^{n} f(X_i)$ over all densities f.



Log-concave projections (Dümbgen, S. and Schuhmacher, 2011)

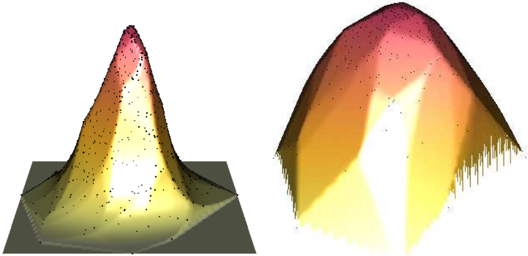
Let $\mathcal{P}_d$ be the set of probability distributions P on $\mathbb{R}^d$ with $\int_{\mathbb{R}^d} \|x\| \, dP(x) < \infty$ and $P(H) < 1$ for all hyperplanes H.

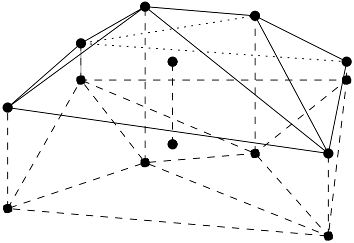
Let $\mathcal{F}_d$ be the set of upper semi-continuous log-concave densities on $\mathbb{R}^d$.

There exists a well-defined projection $\psi^* : \mathcal{P}_d \to \mathcal{F}_d$ given by

$$\psi^*(P) := \operatorname*{argmax}_{f \in \mathcal{F}_d} \int_{\mathbb{R}^d} \log f \, dP.$$

The log-concave MLE





Theoretical properties (Cule and S., 2010; Dümbgen, S. and Schuhmacher, 2011)

The log-concave projection is continuous with respect to the Wasserstein (Mallows-1) distance.

In particular, let $X_1, \ldots, X_n \overset{\mathrm{iid}}{\sim} P_0 \in \mathcal{P}_d$, and let $f^* := \psi^*(P_0)$. Taking $a_0 > 0$ and $b_0 \in \mathbb{R}$ such that $f^*(x) \leq e^{-a_0\|x\|+b_0}$, we have for any $a < a_0$ that

$$\int_{\mathbb{R}^d} e^{a\|x\|}|\hat{f}_n(x) - f^*(x)|\,dx \overset{\mathrm{a.s.}}{\to} 0.$$

Moreover, if f^* is continuous, then

$$\sup_{x \in \mathbb{R}^d} e^{a\|x\|}|\hat{f}_n(x) - f^*(x)| \overset{\mathrm{a.s.}}{\to} 0.$$

Minimax lower bounds (Kim and S., 2016)

Let $d_{\mathrm{H}}^2(f,g) := \int_{\mathbb{R}^d} (f^{1/2} - g^{1/2})^2$ denote the squared Hellinger distance between f and g. There exist positive constants $c_1, c_2, \ldots$ such that

$$\inf_{\tilde{f}_n} \sup_{f_0 \in \mathcal{F}_d} \mathbb{E} d_{\mathrm{H}}^2(\tilde{f}_n, f_0) \geq \begin{cases} c_1 n^{-4/5} & \text{if } d = 1 \\ c_d n^{-2/(d+1)} & \text{if } d \geq 2. \end{cases}$$

Thus, when $d \geq 3$, the problem is *fundamentally harder* than estimating a density of smoothness 2.

Minimax lower bounds (Kim and S., 2016)

Let $d_{\mathrm{H}}^2(f,g) := \int_{\mathbb{R}^d}(f^{1/2} - g^{1/2})^2$ denote the squared Hellinger distance between f and g. There exist positive constants $c_1, c_2, \ldots$ such that

$$\inf_{\tilde{f}_n} \sup_{f_0 \in \mathcal{F}_d} \mathbb{E} d_{\mathrm{H}}^2(\tilde{f}_n, f_0) \geq \begin{cases} c_1 n^{-4/5} & \text{if } d = 1 \\ c_d n^{-2/(d+1)} & \text{if } d \geq 2. \end{cases}$$

Thus, when $d \geq 3$, the problem is *fundamentally harder* than estimating a density of smoothness 2.

In fact, Dagan and Kur (2019) recently showed that c_d may be chosen independent of d.

MLE supremum risk (Kim and S., 2016)

Let $d_X^2(\hat{f}_n, f_0) := n^{-1} \sum_{i=1}^n \log \frac{\hat{f}_n(X_i)}{f_0(X_i)}$. Then

$$d_{\mathrm{H}}^2(\hat{f}_n, f_0) \leq d_{\mathrm{KL}}^2(\hat{f}_n, f_0) \leq d_X^2(\hat{f}_n, f_0).$$

Moreover, the log-concave MLE $\hat{f}_n$ satisfies

$$\sup_{f_0 \in \mathcal{F}_d} \mathbb{E} d_X^2(\hat{f}_n, f_0) = \begin{cases} O(n^{-4/5}) & \text{if } d = 1 \\ O(n^{-2/3} \log n) & \text{if } d = 2 \\ O(n^{-1/2} \log n) & \text{if } d = 3. \end{cases}$$

MLE supremum risk (Kim and S., 2016)

Let $d_X^2(\hat{f}_n, f_0) := n^{-1} \sum_{i=1}^n \log \frac{\hat{f}_n(X_i)}{f_0(X_i)}$. Then

$$d_{\mathrm{H}}^2(\hat{f}_n, f_0) \leq d_{\mathrm{KL}}^2(\hat{f}_n, f_0) \leq d_X^2(\hat{f}_n, f_0).$$

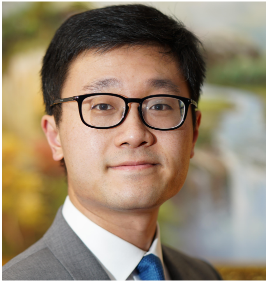
Moreover, the log-concave MLE $\hat{f}_n$ satisfies

$$\sup_{f_0 \in \mathcal{F}_d} \mathbb{E} d_X^2(\hat{f}_n, f_0) = \begin{cases} O(n^{-4/5}) & \text{if } d = 1 \\ O(n^{-2/3} \log n) & \text{if } d = 2 \\ O(n^{-1/2} \log n) & \text{if } d = 3. \end{cases}$$

Dagan and Kur (2019) showed that

$$\sup_{f_0 \in \mathcal{F}_d} \mathbb{E} d_{\mathrm{H}}^2(\hat{f}_n, f_0) = O(n^{-2/(d+1)} \log n) \quad \text{if } d \geq 4.$$

High dimensions



Joint work with Min Xu

Minkowski functionals

Let $\mathcal{K}$ be the set of closed, convex subsets K of $\mathbb{R}^d$ with $0 \in \mathrm{int}(K)$.

For $K \in \mathcal{K}$, the *Minkowski functional* $\|\cdot\|_K : \mathbb{R}^d \to [0,\infty)$ is given by

$$\|x\|_K := \inf\{t \in [0,\infty) : x \in tK\}.$$

Let $K \in \mathcal{K}$ and $x, y \in \mathbb{R}^d$. Then

(i) $\|x\|_K < \infty$;

(ii) $x \in K$ iff $\|x\|_K \leq 1$;

(iii) $x \in \partial K$ iff $\|x\|_K = 1$, where $\partial K := K \setminus \mathrm{int}(K)$;

(iv) $\|x+y\|_K \leq \|x\|_K + \|y\|_K$ and, if $\alpha \geq 0$, then $\|\alpha x\|_K = \alpha\|x\|_K$.

K-homothetic log-concave densities

Say that a density f on $\mathbb{R}^d$ is *homothetic* if there exist a decreasing function $r : (0, \|f\|_\infty) \to [0, \infty)$, a set $A \in \mathcal{B}(\mathbb{R}^d)$ with $0 \in \mathrm{int}(A)$ and $\mu \in \mathbb{R}^d$ such that $\{x : f(x) \geq t\} = r(t)A + \mu$ for every $t \in (0, \|f\|_\infty)$.

K-homothetic log-concave densities

Say that a density f on $\mathbb{R}^d$ is *homothetic* if there exist a decreasing function $r : (0, \|f\|_\infty) \to [0, \infty)$, a set $A \in \mathcal{B}(\mathbb{R}^d)$ with $0 \in \mathrm{int}(A)$ and $\mu \in \mathbb{R}^d$ such that $\{x : f(x) \geq t\} = r(t)A + \mu$ for every $t \in (0, \|f\|_\infty)$.

Let Φ be the set of all upper semi-continuous, concave, decreasing functions $\phi : [0, \infty) \to [-\infty, \infty)$.

Let f be a u.s.c. density on $\mathbb{R}^d$. Then f is homothetic and log-concave iff there exist $K \in \mathcal{K}$, $\mu \in \mathbb{R}^d$ and $\phi \in \Phi$ such that $f(\cdot) = e^{\phi(\|\cdot - \mu\|_K)}$.

K-homothetic log-concave densities

Say that a density f on $\mathbb{R}^d$ is *homothetic* if there exist a decreasing function $r : (0, \|f\|_\infty) \to [0, \infty)$, a set $A \in \mathcal{B}(\mathbb{R}^d)$ with $0 \in \mathrm{int}(A)$ and $\mu \in \mathbb{R}^d$ such that $\{x : f(x) \geq t\} = r(t)A + \mu$ for every $t \in (0, \|f\|_\infty)$.

Let Φ be the set of all upper semi-continuous, concave, decreasing functions $\phi : [0, \infty) \to [-\infty, \infty)$.

Let f be a u.s.c. density on $\mathbb{R}^d$. Then f is homothetic and log-concave iff there exist $K \in \mathcal{K}$, $\mu \in \mathbb{R}^d$ and $\phi \in \Phi$ such that $f(\cdot) = e^{\phi(\|\cdot - \mu\|_K)}$.

If also $f(\cdot) = e^{\tilde{\phi}(\|\cdot - \tilde{\mu}\|_{\tilde{K}})}$, then there exist $\sigma, \sigma' > 0$ such that $\tilde{\phi}(\cdot) = \phi(\sigma \cdot)$ and $\tilde{K} = \sigma K + \sigma'(\mu - \tilde{\mu})$; moreover, if f is not a uniform density, then $\tilde{\mu} = \mu$.

K-homothetic log-concave projections

Let $\mathcal{P}_d$ be the set of probability distributions P on $\mathbb{R}^d$ with finite mean and with $P(\{0\}) < 1$. Let $\mathcal{F}_d^K$ be the set of usc, K-homothetic, log-concave densities with centering vector 0.

There exists a well-defined projection $f^* : \mathcal{P}_d \to \mathcal{F}_d^K$ given by

$$f^*(P) := \operatorname*{argmax}_{f \in \mathcal{F}_d^K} \int_{\mathbb{R}^d} \log f(x) \, dP(x).$$

K-homothetic log-concave projections

Let $\mathcal{P}_d$ be the set of probability distributions P on $\mathbb{R}^d$ with finite mean and with $P(\{0\}) < 1$. Let $\mathcal{F}_d^K$ be the set of usc, K-homothetic, log-concave densities with centering vector 0.

There exists a well-defined projection $f^* : \mathcal{P}_d \to \mathcal{F}_d^K$ given by

$$f^*(P) := \operatorname*{argmax}_{f \in \mathcal{F}_d^K} \int_{\mathbb{R}^d} \log f(x)\, dP(x).$$

Now let $\mathcal{Q}$ denote the set of probability measures Q on $[0, \infty)$ with finite mean and $Q(\{0\}) < 1$. Then we can also define $\phi^* : \mathcal{Q} \to \Phi$ by

$$\phi^*(Q) := \operatorname*{argmax}_{\phi \in \Phi} \int_{\mathbb{R}^d} \phi\, dQ.$$

Risk bounds when K is known

Let $f_0(\cdot) = e^{\phi_0(\|\cdot\|_K)} \in \mathcal{F}_d^K$, and suppose that $X_1, \ldots, X_n \stackrel{\mathrm{iid}}{\sim} f_0$ with empirical distribution $\mathbb{P}_n$. Let $\hat{f}_n := f^*(\mathbb{P}_n)$*

There exists a universal constant $C > 0$ such that for $n \geq 8$,

$$\mathbb{E}d_X^2(\hat{f}_n, f_0) \leq \frac{C}{n^{4/5}}.$$

Thus, the risk bound does not depend on d.

*This may be computed by setting $Z_i = \|X_i\|_K$ for $i \in [n]$ and then, writing $\mathbb{Q}_n$ for the empirical distribution of $Z_1, \ldots, Z_n$, computing $\hat{f}_n(\cdot) := e^{\phi^*(\mathbb{Q}_n)(\|\cdot\|_K)}$.

Adaptive bounds

Let $\phi^{(k)}$ be the set of $\phi \in \Phi$ such that there exist $r_0 > 0$ and intervals $I_1, \ldots, I_k$ with $I_j = [a_{j-1}, a_j]$ for which ϕ is affine on each I_j, and $\phi(r) = -\infty$ for $r > r_0 = a_k$.

Let $\mathcal{H}$ denote the set of densities of the form $r \mapsto d\lambda_d(K) r^{d-1} e^{\phi(r)}$ for $\phi \in \Phi$, and let $\mathcal{H}^{(k)}$ be the set of $h \in \mathcal{H}$ for which the corresponding ϕ belongs to $\Phi^{(k)}$.

Define $h_0 \in \mathcal{H}$ by $h_0(r) := d\lambda_d(K) r^{d-1} e^{\phi_0(r)}$. Then, writing $\nu_k := 2^{1/2} \wedge \inf_{h \in \mathcal{H}^{(k)}} d_{\mathrm{KL}}(h_0, h)$, there exists a universal constant $C > 0$ such that for $n \geq 8$,

$$\mathbb{E} d_X^2(\hat{f}_n, f_0) \leq \min_{k \in [n]} \biggl(\nu_k^2 + C \frac{k}{n} \log^{5/4} \frac{en}{k} \biggr).$$

Estimating K when partially known

Let $K_0 \in \mathcal{K}$ be balanced and in isotropic position, so that $\int_{K_0} \frac{xx^\top}{\lambda_d(K_0)}\,dx = I_d$.

Let $r_1, r_2 > 0$ be such that $r_1 B_d(0,1) \subseteq K_0 \subseteq r_2 B_d(0,1)$ and let $r_0 := r_2/r_1$.

Let $\Sigma_0 \succ 0$, $K = \Sigma_0^{1/2} K_0$, and let $f_0 \in \mathcal{F}_d^{K,\mu}$ be such that $f_0(\cdot) = e^{\phi_0(\|\cdot - \mu\|_K)}$. Assume that K_0 is known but that Σ_0 is unknown.

Assume that $X_1, \ldots, X_{2n} \overset{\mathrm{iid}}{\sim} f_0$, and denote the sample covariance matrix by $\hat{\Sigma} := n^{-1} \sum_{i=n+1}^{2n} (X_i - \hat{\mu})(X_i - \hat{\mu})^\top$, where $\hat{\mu} := n^{-1} \sum_{i=n+1}^{2n} X_i$. We let $\hat{K} := \hat{\Sigma}^{1/2} K_0$ and $\hat{f}_n := f^*_{\hat{K},\hat{\mu}}(\mathbb{P}_n)$, where $\mathbb{P}_n$ is based on $X_1, \ldots, X_n$.

Risk bounds when K partially known

We have

$$\mathbb{E}d_{\mathrm{H}}^2(\hat{f}_n, f_0) \lesssim r_0 \frac{d^{3/2}}{n^{1/2}} \log^3(en).$$

Moreover, if ϕ_0 is such that ϕ_0' is absolutely continuous and that $\inf_{r\in[0,\infty)} \phi_0''(r) \geq -D_0$ for some $D_0 > 0$, then

$$\mathbb{E}d_{\mathrm{H}}^2(\hat{f}_n, f_0) \lesssim \frac{1}{n^{4/5}} + r_0^2(D_0^2 + 1)\frac{d^3}{n} \log^6(en).$$

Finally, if $f_0(\cdot) = e^{-a\|\cdot-\mu\|_K+b}$ for some $a > 0$ and $b \in \mathbb{R}$, then

$$\mathbb{E}d_{\mathrm{H}}^2(\hat{f}_n, f_0) \lesssim \frac{1}{n} \log^{5/4}(en) + r_0^2 \frac{d^3}{n} \log^6(en).$$

Remark: In this setting $r_0 \leq d^{1/2}$.

Estimating K nonparametrically

Data: $M \in \mathbb{N}$ and $X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+M} \in \mathbb{R}^d$.

Result: $\hat{K} \in \mathcal{K}$.

Set $k \leftarrow \lfloor \log n \rfloor$;

For $m \in [M]$, let $\theta_m = X_{n+m}/\|X_{n+m}\|_2$;

for $m \in [M]$ **do**

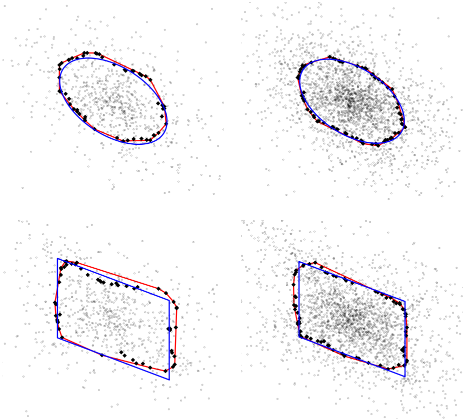
$\quad \mathcal{I}_m^k \leftarrow \bigl\{i \in [n] : X_i^\top \theta_m \geq \ k\text{-th}\max\{X_i^\top \theta_{m'} \ : \ m' \in [M]\}\bigr\}$;

$\quad t_m \leftarrow |\mathcal{I}_m^k|^{-1} \sum_{i \in \mathcal{I}_m^k} \|X_i\|_2$;

end

Set $\hat{K} \leftarrow \mathrm{conv}\{t_1\theta_1, \ldots, t_m\theta_m\}$.

Algorithm in action



Let $\mu = 0$. Let $\hat{K}$ be the output of the algorithm with $M := \lceil n^{\frac{d-1}{d+1}} \rceil$, and let $\hat{f}_n := f^*_{\hat{K}}(\mathbb{P}_n)$. If $r_0^2 n^{-\frac{1}{d+1}} \log^3(en) \leq 1/64$, then

$$\mathbb{E} d_{\mathrm{H}}^2(\hat{f}_n, f_0) \lesssim_{d,r_0} n^{-\frac{1}{d+1}} \log^3(en).$$

Moreover, if in addition, ϕ_0' is absolutely continuous and $\inf_{r \in [0,\infty)} \phi_0''(r) \geq -D_0$ for some $D_0 > 0$, then

$$\mathbb{E} d_{\mathrm{H}}^2(\hat{f}_n, f_0) \lesssim_{d,r_0} n^{-\frac{2}{d+1}} \log^6(en).$$

Remark: In this general setting, $r_0 \leq d$.

Summary

- In low dimensions, the log-concave MLE provides an attractive, fully automatic nonparametric approach to density estimation.

Summary

- In low dimensions, the log-concave MLE provides an attractive, fully automatic nonparametric approach to density estimation.
- In higher dimensions, the incorporation of additional symmetry constraints facilitates feasible algorithms that may mitigate the curse of dimensionality.

Summary

- In low dimensions, the log-concave MLE provides an attractive, fully automatic nonparametric approach to density estimation.
- In higher dimensions, the incorporation of additional symmetry constraints facilitates feasible algorithms that may mitigate the curse of dimensionality.

Main reference:

- Xu, M. and Samworth, R. J. (2019) High-dimensional nonparametric density estimation via symmetry and shape constraints. Preprint. `https://arxiv.org/abs/1903.06092`.

Other references

- Carpenter, T., Diakonikolas, I., Sidiropoulos, A. and Stewart A. (2018) Near-optimal sample complexity bounds for maximum likelihood estimation of multivariate log-concave densities. *Proc. Mach. Learn. Res.*, **75**, 1–29.
- Cule, M. and Samworth, R. (2010) Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Statist.*, **4**, 254–270.
- Cule, M., Samworth, R. and Stewart, M. (2010) Maximum likelihood estimation of a multi-dimensional log-concave density. *J. Roy. Statist. Soc., Ser. B. (with discussion)*, **72**, 545–607.
- Dagan, Y. and Kur, G. (2019) The log-concave maximum likelihood estimator is optimal in high dimensions. `https://arxiv.org/abs/1903.05315`.
- Dümbgen, L. and Rufibach, K. (2009) Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli*, **15**, 40–68.
- Dümbgen, L., Samworth, R. and Schuhmacher, D. (2011) Approximation by log-concave distributions with applications to regression. *Ann. Statist.*, **39**, 702–730.
- Groeneboom, P. and Jongbloed, G. (2014) *Nonparametric Inference under Shape Constraints*, Cambridge University Press, Cambridge.
- Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001) Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.*, **29**, 1653–1698.
- Kim, A. K. H., Guntuboyina, A. and Samworth, R. J. (2018) Adaptation in log-concave density estimation. *Ann. Statist.*, **46**, 2279–2306.
- Kim, A. K. H. and Samworth, R. J. (2016) Global rates of convergence in log-concave density estimation. *Ann. Statist.*, **44**, 2756–2779.
- Koenker, R. and Mizera, I. (2010) Quasi-concave density estimation. *Ann. Statist.*, **38**, 2998–3027.
- Pal, J., Woodroofe, M. and Meyer, M. (2007) Estimating a Polya frequency function. In *Complex datasets and Inverse problems, Networks and Beyond Tomography*, vol. **54** of *Lecture Notes - Monograph Series*, 239–249. IMS.
- Robeva, E., Sturmfels, B., Tran, N. and Uhler, C. (2018) Maximum likelihood estimation for totally positive log-concave densities. `https://arxiv.org/abs/1806.10120`.
- Schuhmacher, D., Hüsler, A. and Dümbgen, L. (2011) Multivariate log-concave distributions as a nearly parametric model. *Statistics & Risk Modeling*, **28**, 277–295.
- Seregin, A. and Wellner, J. A. (2010) Nonparametric estimation of convex-transformed densities. *Ann. Statist.*, **38**, 3751–3781.
- Walther, G. (2002) Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.*, **97**, 508–513.