1 To Do

1.1 Major

1. Extend proposition 3. It currently describes projections onto elliptical log-concave densities only when Σ and μ are fixed. Σ and μ should be variables to be optimized by the projection.

2. Derive the rate of convergence of the MLE for spherically symmetric distributions. The steps to accomplish this follow that of KS 2016.

3. Strengthen the continuity of log-concave projection result from DSS to account for error of estimating Σ and μ.

1.2 Minor

1. Put the finishing touch on proposition 3. Use DSS 2011 to show existence.

2. Finish identifiability proof.

2 Introduction

2.1 Notation

For a vector $x \in \mathbb{R}^p$, $\|x\|, \|x\|_2$ both denotes the l_2 norm. For a matrix A, $\|A\|_2$ denotes the operator norm. We represent positive definiteness of a matrix A as $A \succ 0$, and semidefiniteness as $A \succeq 0$. Given a vector x and a matrix $A \succ 0$, $\|x\|_A = \sqrt{x^\top A^{-1} x}$ denotes the Mahalanobis distance.

2.2 Elliptical Density

A p-dimensional random vector has the Elliptical density if the pdf is of the form

$$f(x\,;\,\mu, \Sigma) = |\Sigma|^{-1/2} g_p\big(\|x-\mu\|_\Sigma\big)$$

where Ω is a positive definite matrix and $\|x\|_\Sigma$ is the Mahalanobis distance, $\|x\|_\Sigma^2 = x^\top \Sigma^{-1} x$.

where $g_p : \mathbb{R}^+ \to \mathbb{R}^+$ is a generator function with the property

$$\int_{\mathbb{R}^p} g_p(\|x\|_2)dx = 1$$

2.3 Identifiability

There is one degree of non-identifiability. Let $a > 0$, let $\Sigma' = \frac{\Sigma}{a}$, then we have that

$$\begin{aligned} f(x\,;\,\mu, \Sigma) &= \left|\frac{\Sigma}{a}\right|^{-1/2} a^{-p/2} g_p\left(\sqrt{\frac{1}{a}}\sqrt{x^\top \left(\frac{\Sigma}{a}\right)^{-1} x}\right) \\ &= |\Sigma'|^{-1/2} g_p'(\sqrt{x^\top \Sigma'^{-1} x}) \end{aligned}$$

where $g_p'(r) = a^{-p/2} g_p(r/\sqrt{a})$. It is easy to check that $\int_{\mathbb{R}^p} g_p'(\|x\|_2)dx = 1$. Thus, without loss of generality, we may assume that $\|\Sigma\|_2 = 1$.

To prove identifiability, we note the following lemma:

Lemma 1. *Suppose $A, B \succ 0$. Let $a, b > 0$, the sets $\{x : x^\top Ax = a\}$ and $\{x : x^\top Bx = b\}$ are equal iff $(bA)/a = B$.*

Proof. Let $S = \{x : x^\top Ax = a\}$. We have that for any $x \in S$, $x^\top((bA/a) - B)x = 0$. Since S contains p independent vectors, namely the elementary basis appropriately scaled, we have that $(bA/a) - B = 0$. $\square$

Now suppose (Σ, g_p) and (Σ', g_p') induce the same density f. We have then that $g_p(\sqrt{x^\top \Sigma^{-1} x}) = cg_p'(\sqrt{x^\top \Sigma'^{-1} x}) = f(x)$ for some $c > 0$.

[TODO:finish, intuition: we look at the level sets of g_p and g_p', i.e., $g_p^{-1}(\{a\})$ for some $a > 0$. If the level sets are singletons, this is easy. If the level sets are bounded, this is easy too. If the level sets are unbounded, what to do?

2.4 Characterizations

Let X follow a centered elliptical distribution. Then, we have that

$$X = \Omega^{1/2}\Phi Y$$

where Φ is random vector from $\mathbb{S}^{p-1}$ and Y is a non-negative random variable that follows the density

$$f_Y(y) = c_p y^{p-1} g_p(y)$$

$c_p = 2\frac{\pi^{p/2}}{\Gamma(p/2)}$.

3 Log-Concavity

A related lemma in [TODO:cite Bhattacharyya] states that f is unimodal iff g_p is non-increasing.

Lemma 2. *f is log-concave iff g_p is log-concave and non-increasing.*

Proof. Without loss of generality, suppose that $\mu = 0$.

Suppose g_p is log-concave and non-increasing. Then, we have that

$$\begin{aligned}
\log f(\lambda x + (1-\lambda)y) &= (-1/2)\log|\Sigma| + \log g_p(\|\lambda x + (1-\lambda)y\|_\Sigma) \\
&\geq (-1/2)\log|\Sigma| + \log g_p(\lambda\|x\|_\Sigma + (1-\lambda)\|y\|_\Sigma) \\
&\geq (-1/2)\log|\Sigma| + \lambda\log g_p(\|x\|_\Sigma) + (1-\lambda)\log g_p(\|y\|_\Sigma) \\
&= \lambda\log f(x) + (1-\lambda)\log f(y)
\end{aligned}$$

The first inequality follows because $\|\lambda x + (1-\lambda)y\|_\Sigma \leq \lambda\|x\|_\Sigma + (1-\lambda)\|y\|_\Sigma$; since $\|\cdot\|_\Sigma$ is a norm, it is convex. The first inequality follows also because $\log g_p$ is a non-increasing function. The second inequality follows from the log-concavity of g_p.

Now we turn to the converse. If g_p is increasing at any point, then it is clear that f is no longer unimodal and hence not log-concave. If g_p is not log-concave, then for some $t, s \in \mathbb{R}^+$, $\log g_p(\lambda t + (1-\lambda)s) < \lambda\log g_p(t) + (1-\lambda)\log g_p(s)$. Let $z \in \mathbb{R}^p$ satisfy $\|z\|_\Sigma = 1$, then

$$\begin{aligned}
\log f(\lambda tz + (1-\lambda)sz) &= \log g_p(\lambda t + (1-\lambda)s) \\
&< \lambda\log g_p(t) + (1-\lambda)\log g_p(s) \\
&= \lambda\log f(tz) + (1-\lambda)\log f(sz)
\end{aligned}$$

and thus $\log f$ is concave either.

□

3.1 Projection Operator

Define $\mathcal{F} = \{\phi(\mathbb{R}^+, \mathbb{R}^+) \; : \; \phi \text{ is concave, decreasing}\}$. We describe projection onto the class of p-variate densities of the form $f(x) = |\Sigma|^{-1/2}\exp(\phi(\|x\|_\Sigma))$ where ϕ is such that $\exp(\phi(r))r^{p-1}c_p$ is a density over $[0, \infty)$.

First, we fix $\Sigma \succ 0$. We can without loss of generality assume that $\|\Sigma\|_2 = 1$ as we have discussed in the section on identifiability.

Definition Let $\Sigma \succ 0$ be fixed. For a probability measure P over $\mathbb{R}^p$ and a function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$, we define

$$L_\Sigma(\phi, P) = \int \phi(\|x\|_\Sigma) dP - \int_0^\infty \exp(\phi(r)) r^{p-1} c_p dr$$

The projection of P is $\phi^* \in \mathcal{F}$ such that

$$L_\Sigma(\phi^*, P) = \sup_{\phi \in \mathcal{F}} L_\Sigma(\phi, P)$$

Note 1:

First, we check that if ϕ^* exists, then $\exp(\phi^*(r)) r^{p-1} c_p$ is indeed a density. To see this, note that

$$\partial_c L_\Sigma(\phi^* + c, P) = 1 - e^c \int_0^\infty \exp(\phi^*(r)) r^{p-1} c_p dr$$

By definition of ϕ^*, $c = 0$ implies that $\partial_c L_\Sigma(\phi^* + c, P) = 0$, implying further that $\exp(\phi^*(r)) r^{p-1} c_p$ is indeed a valid density.

Note 2:

It is clear that $L_\Sigma(\phi, P)$ is concave in ϕ.

Proposition 3. *Let $\Sigma \succ 0$ be fixed. Define $\mathcal{P}_d = \{P : \int \|x\| dP < \infty,\ P(\{0\}) < 1\}$. Then, we have,*

1. *If $\int \|x\| dP = \infty$, then $L_\Sigma(\phi, P) = -\infty$ for all $\phi \in \mathcal{F}$.*
2. *If $P(\{0\}) = 1$, then $\sup_{\phi \in \mathcal{F}} L_\Sigma(\phi, P) = \infty$*
3. *If $P(\{0\}) < 1$ and $\int \|x\| dP < \infty$, then $\sup_{\phi \in \mathcal{F}} L_\Sigma(\phi, P) < \infty$ and there exists a maximizer $\phi^* \in \mathcal{F}$ which achieves this value.*

Proof. Suppose that $\int \|x\| dP = \infty$. Then $\int \|x\|_\Sigma dP \geq \int \frac{\|x\|_2}{\|\Sigma\|_2} dP = \infty$.

First, suppose that $\lim_{r \to \infty} \phi(r) = c > -\infty$. Then $L(\phi, P) \leq \phi(0) - \int_0^\infty r^{p-1} e^c c_p dr = -\infty$. Thus, we may consider only ϕ such that $\lim_{r \to \infty} \phi(r) = -\infty$. For any such ϕ, there exists $a, b > 0$ such that $\phi(r) \leq a - b|r|$.

Hence, $L(\phi, P) \leq \int \phi dP \leq \int a - b\|x\|_\Sigma dP \leq -\infty$. This proves claim 1.

For claim 2, suppose $P(\{0\}) = 1$. Let $\phi_n(r) = n - e^n r$. Then, we have that

$$
\begin{aligned}
L(\phi_n, P) &= n - \int e^n \exp(-e^n r) r^{p-1} c_p dr \\
&= n - e^n \int e^{-s} \left(\frac{s}{e^n}\right)^{p-1} \frac{ds}{e^n} \\
&= n - (e^n)^{1-p}\Gamma(p)
\end{aligned}
$$

Thus, we have that $\lim_{n\to\infty} L(\phi_n, P) = \infty$. This proves the second claim.

Onto the third claim. We will first prove that if $P(\{0\}) < 1$ and $\int \|x\| dP < \infty$, then $-\infty < \sup_{\phi\in\mathcal{F}} L_\Sigma(\phi, P) < \infty$. Then we will prove that the maximizer exists.

By plugging in $\phi(r) = -r$, we have that $L_\Sigma(\phi, P) = -\int \|x\|_\Sigma dP - \int e^{-r} r^{p-1} c_p dr = -\int \|x\|_\Sigma dP - c_p\Gamma(p/2)$. Since $\int \|x\|_\Sigma dP \le \int \|x\| \|\Sigma^{-1}\|_2 dP < \infty$, we have shown that $L_\Sigma > -\infty$ for some ϕ.

Define $b^* = \inf\{b : P(B_\Sigma(0; b)) \ge \frac{P(\{0\})}{2} + \frac{1}{2}\}$. $b^* > 0$ since $P(\{0\}) < 1$. Let $b = b^*/2$, $c = P(B_\Sigma(0; b))$, then we have that $0 < c < 1$.

Suppose $\phi(0) = M$ and $\phi(b) = M'$, because ϕ is non-increasing, $M = \sup_r \phi(r)$ and $M' = \inf_{r\in[0,b]} \phi(r) = \sup_{r\in[b,\infty)} \phi(r)$.

$$
\int \phi dP \le \int_{B_\Sigma(0;b)} \phi dP + \int_{B_\Sigma(0;b)^c} \phi dP \le Mc + M'(1-c) = (M - M')c + M'
$$

Then, we have that

$$
\begin{aligned}
L_\Sigma(\phi, P) &= \int \phi dP - \int e^{\phi(r)} r^{p-1} c_p dr \\
&\le \int \phi dP - \int_0^b e^{\phi(r)} r^{p-1} c_p dr \\
&\le (M - M')c + M' - \int_0^b \exp(M - \frac{r}{b}(M - M')) r^{p-1} c_p dr \\
&\le \Delta(c-1) + M - e^M \int_0^b \exp(-\frac{r}{b}\Delta) r^{p-1} c_p dr
\end{aligned}
$$

where we have used the notation $\Delta = M - M'$.

First, let us suppose that $\Delta(1-c) \le 2M$. Then, we have that

$$
\begin{aligned}
L_\Sigma(\phi, P) &\leq M - e^M \int_0^b \exp(-\frac{r}{b}\frac{2M}{1-c}) r^{p-1} c_p dr \\
&\leq M - e^M \int_0^{2M/(1-c)} e^{-s} s^{p-1} c_p ds \left(\frac{b}{2M/(1-c)}\right)^p
\end{aligned}
$$

Which is bounded since the RHS goes to $-\infty$ as M goes to ∞.

Now, let us suppose that $\Delta(1-c) > 2M$, then we have that

$$
L_\Sigma(\phi, P) \leq -M
$$

Thus, we see that $L_\Sigma(\phi, P)$ is bounded and, furthermore, there exists a constant M^* such that $\sup\{L(\phi, P) \,:\, \phi \in \mathcal{F}\} = \sup\{L(\phi, P) \,:\, \phi \in \mathcal{F},\, \|\phi\|_\infty \leq M^*\}$.

Let $r^* = \sup\{r \,:\, P(B_\Sigma(0; r)) < 1\}$, Then

$\square$

Suppose $P(H) = 1$ for some hyperplane, then let $\Sigma_n \to A$ where H is the nullspace of A.

Suppose interior(csupp(P)) is non-empty, then its Lebesgue measure is some $c > 0$. The Lebesgue measure of $B_\Sigma(0; b)$ is at $c_p \frac{b^p}{p} |\Sigma|^{1/2}$. We assume that $\|\Sigma\|_2 = 1$.

4 Algorithm

Let $X_1, ..., X_n \sim P$ be the samples and let μ be zero and Σ be fixed. The log-likelihood is

$$
l(g_p\,;\, X_1, ..., X_n) = \sum_{i=1}^n \log g_p(\|X_i\|_\Sigma)
$$

where g_p is decreasing, log-concave, and satisfies $\int_0^\infty g_p(r) r^{p-1} c_p dr = 1$. If we reparametrize the problem by writing $\phi(r) = \log g_p(r)$, $Y_i = \|X_i\|_\Sigma$ and also put the integral constraint in the Lagrangian form, we get an equivalent optimization

$$
\max_{\phi \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \phi(Y_i) - \int_0^\infty \exp(\phi(r)) r^{p-1} c_p dr
$$

where $\mathcal{F} = \{\phi \,:\, \phi \text{ decreasing and concave }\}$. We use the notation from the previous section and denote $L_\Sigma(\phi, P_n) = \frac{1}{n}\sum_{i=1}^n \phi(Y_i) - \int_0^\infty \exp(\phi(r)) r^{p-1} c_p dr$.

Lemma 4. *Let $\phi \in \mathcal{F}$ and let $\bar{\phi}$ be the piecewise linear function with the property that $\bar{\phi}(Y_i) = \phi(Y_i)$ for all $i = 1, ..., n$ and $\bar{\phi}(0) = \phi(0)$. Then, we have that $\bar{\phi} \in \mathcal{F}$ and*

$$L_\Sigma(\bar{\phi}, P_n) \geq L_\Sigma(\phi, P_n)$$

The implication is that we need only optimize over piecewise linear functions whose knots are placed at $\{Y_1, ..., Y_n\} \cup \{0\}$.

Proof. It is clear that $\bar{\phi} \in \mathcal{F}$ and that $\phi \geq \bar{\phi}$.

Therefore,

$$\sum_{i=1}^n \phi(Y_i) = \sum_{i=1}^n \bar{\phi}(Y_i)$$
$$\int_0^\infty \exp(\phi(r)) r^{p-1} c_p dr \geq \int_0^\infty \exp(\bar{\phi}(r)) r^{p-1} c_p dr$$

□

4.1 Piecewise Linear Parametrization

Let $\bar{\mathcal{F}} = \{\phi : \phi \text{ is p.w. linear, decreasing, concave}\}$.

Given samples $Y_1, ..., Y_n \in \mathbb{R}^+$, any $\phi \in \bar{\mathcal{F}}$ can be parametrized by a vector $(\phi_1, ..., \phi_n)$,

$$\phi(r) = \sum_{i=1}^{n-1} \left[\left(\frac{Y_{i+1} - r}{Y_{i+1} - Y_i} \right) \phi_i + \left(\frac{r - Y_i}{Y_{i+1} - Y_i} \right) \phi_{i+1} \right] \mathbf{1}_{r \in [Y_i, Y_{i+1}]} + \phi_1 \mathbf{1}_{r \in [0, Y_1]}$$

Thus, we can write the full optimization as

$$\max_{\phi_1, \ldots, \phi_n} \frac{1}{n} \sum_{i=1}^n \phi_i - \sum_{i=1}^{n-1} \int_{Y_i}^{Y_{i+1}} \exp\left(\frac{Y_{i+1} - r}{Y_{i+1} - Y_i} \phi_i + \frac{r - Y_i}{Y_{i+1} - Y_i} \phi_{i+1} \right) r^{p-1} c_p dr - \int_0^{Y_1} \exp(\phi_1) r^{p-1} c_p dr$$

$$\text{subject to } \frac{\phi_{i+1} - \phi_i}{Y_{i+1} - Y_i} \geq \frac{\phi_{i+2} - \phi_{i+1}}{Y_{i+2} - Y_{i+1}} \quad \text{for all } i = 1, .., n-2$$
$$\frac{\phi_2 - \phi_1}{Y_2 - Y_1} \leq 0$$

4.1.1 Derivatives

Define the F function as the objective

$$F(\phi_1, ..., \phi_n) = \frac{1}{n} \sum_{i=1}^n \phi_i - \sum_{i=1}^{n-1} \int_{Y_i}^{Y_{i+1}} \exp\left(\frac{Y_{i+1} - r}{Y_{i+1} - Y_i} \phi_i + \frac{r - Y_i}{Y_{i+1} - Y_i} \phi_{i+1} \right) r^{p-1} c_p dr - \int_0^{Y_1} \exp(\phi_1) r^{p-1} c_p dr$$

We will rewrite F to facilitate the differentiation.

$$F(\phi) = \frac{1}{n}\mathbf{1}^\top \phi - \sum_{i=1}^{n-1} \int_{Y_i}^{Y_{i+1}} \exp(a_i(r)^\top \phi) r^{p-1} c_p dr - \int_0^{Y_1} \exp(\phi_1) r^{p-1} c_p dr$$

where $a_i(r) \in \mathbb{R}^n$ is the following form: $(0, ..., 0, \frac{Y_{i+1}-r}{Y_{i+1}-Y_i}, \frac{r-Y_i}{Y_{i+1}-Y_i}, 0, ..., 0)$ where the two non-zero coordinates are $i, i+1$. Then, we have that

$$\nabla F = \frac{1}{n}\mathbf{1} - \sum_{i=1}^{n-1} \int_{Y_i}^{Y_{i+1}} a_i(r) \exp(a_i(r)^\top \phi) r^{p-1} c_p dr - \int_0^{Y_1} e_1 \exp(\phi_1) r^{p-1} c_p dr$$

$$H\,F = -\sum_{i=1}^{n-1} \int_{Y_i}^{Y_{i+1}} a_i(r) a_i(r)^\top \exp(a_i(r)^\top \phi) r^{p-1} c_p dr - \int_0^{Y_1} e_1 e_1^\top \exp(\phi_1) r^{p-1} c_p dr$$

For a given ϕ, these can be evaluated by numerical integration.

4.1.2 Active Set

Let us express the constraints as $v_i^\top \phi \le 0$ for $i = 1, ..., n-1$, where

$$v_1 = (-\frac{1}{Y_2 - Y_1}, \frac{1}{Y_2 - Y_1}, 0, ..., 0)$$
$$v_2 = (\frac{1}{Y_2 - Y_1}, -\frac{1}{Y_3 - Y_2} - \frac{1}{Y_2 - Y_1}, \frac{1}{Y_3 - Y_2}, 0, ...0)$$
$$...$$
$$v_i = (0, ..., 0, \frac{1}{Y_i - Y_{i-1}}, -\frac{1}{Y_{i+1} - Y_i} - \frac{1}{Y_i - Y_{i-1}}, \frac{1}{Y_{i+1} - Y_i}, 0, ...0)$$
$$...$$

Define the active set $A \subset \{1, ..., n-1\}$ as $A = \{i\,;\, v_i^\top \phi = 0\}$. Define $I = \{1, ..., n-1\} - A$.

Proposition 5. *Define $V \in \mathbb{R}^{n \times n}$ such that the i-th row $V_i = v_i$ for $i = 1, ..., n-1$ and the n-th row $V_n = \mathbf{1}_n$.*

Then, define $B^\top = -V^{-1}$ and let b_i be the i-th row of B. We have that $b_i^\top v_i = -1$ and $b_i^\top v_j = 0$ for $j \neq i$.

The proof follows from the observation that V is invertible and that $B^\top V = -I$.

4.1.3 Optimization over an Active Set

In this section, we solve

$$\begin{aligned}&\min_{\phi} F(\phi)\\&\text{s.t. } v_i^\top \phi = 0 \qquad \text{for all } i \in A\end{aligned}$$

Given the active set A, let us define $I = \{1, ..., n\} - A$. We index the elements of I by $i_1, ..., i_T$ where T denotes the cardinality of I. By definition, $n \in I$ always.

Proposition 6. *The subspace $\{\phi : v_A^\top \phi = 0\}$ is equal to the subspace of ϕ where*

1. $\phi_I \in \mathbb{R}^T$
2. *For $j \in A$ where $j < i_1$, $\phi_j = \phi_{i_1}$*
3. *For $j \in A$ where $j > i_1$, $\phi_j = \frac{Y_{i_{t+1}} - Y_j}{Y_{i_{t+1}} - Y_{i_t}} \phi_{i_t} + \frac{Y_j - Y_{i_t}}{Y_{i_{t+1}} - Y_{i_t}} \phi_{i_{t+1}}$, $i_t < j < i_{t+1}$*

Given the proposition, we can solve the optimization over an active set with an unconstrained optimization.

$$\begin{aligned}\min_{\phi_I} \frac{1}{n} &\left(i_1 \phi_{i_1} + \sum_{t=1}^{T-1} \sum_{j=i_t+1}^{i_{t+1}} \frac{Y_{i_{t+1}} - Y_j}{Y_{i_{t+1}} - Y_{i_t}} \phi_{i_t} + \frac{Y_j - Y_{i_t}}{Y_{i_{t+1}} - Y_{i_t}} \phi_{i_{t+1}} \right) \\ &+ \sum_{t=1}^{T-1} \int_{Y_{i_t}}^{Y_{i_{t+1}}} \exp\left(\frac{Y_{i_{t+1}} - r}{Y_{i_{t+1}} - Y_{i_t}} \phi_i + \frac{r - Y_{i_t}}{Y_{i_{t+1}} - Y_{i_t}} \phi_{i_{t+1}} \right) r^{p-1} c_p dr - \int_0^{Y_{i_1}} \exp(\phi_{i_1}) r^{p-1} c_p dr\end{aligned}$$

We let $F(\phi_I)$ denote the objective function. Again, we can simplify the notation with vector representation.

$$F(\phi_I) = \frac{1}{n} w^\top \phi_I + \sum_{t=1}^{T} \int_{Y_{i_t}}^{Y_{i_{t+1}}} \exp(a_t(r)^\top \phi_I) r^{p-1} c_p dr - \int_0^{Y_{i_1}} \exp(\phi_{i_1}) r^{p-1} c_p dr$$

where $w \in \mathbb{R}^T$ is of the form $w_1 = i_1 + \sum_{j=i_1+1}^{i_2} \frac{Y_{i_2} - Y_j}{Y_{i_2} - Y_{i_1}}$ and $w_t = \sum_{j=i_{t-1}+1}^{i_t} \frac{Y_j - Y_{i_{t-1}}}{Y_{i_t} - Y_{i_{t-1}}} + \sum_{j=i_t+1}^{i_{t+1}} \frac{Y_{i_{t+1}} - Y_j}{Y_{i_{t+1}} - Y_{i_t}}$.

And, $a_t(r) \in \mathbb{R}^T$ is of the form $(0, ..., 0, \frac{Y_{i_{t+1}} - r}{Y_{i_{t+1}} - Y_{i_t}}, \frac{r - Y_{i_t}}{Y_{i_{t+1}} - Y_{i_t}}, 0, ..., 0)$ where the two non-zero coordinates are $t, t+1$.

5 Envelope bounds

Let Φ denote the class of decreasing, concave functions $\phi : [0,\infty) \to [-\infty,\infty)$, let $\mathcal{G} := \{e^{\phi} : \phi \in \Phi\}$, and let $\mathcal{H}$ denote the class of functions $h : [0,\infty) \to [0,\infty)$ of the form $h(r) = c_p r^{p-1} g(r)$ for some $g \in \mathcal{G}$, where

$$c_p \int_0^\infty r^{p-1} g(r)\,dr = 1 \tag{1}$$

$$c_p \int_0^\infty r^{p+1} g(r)\,dr = p. \tag{2}$$

Thus $\mathcal{H}$ consists of densities of random variables $\|X\|$, where X has a spherically symmetric, log-concave density on $\mathbb{R}^p$, and $\mathbb{E}(\|X\|^2) = p$.

The following result provides crude upper bounds for $\mathcal{H}$.

Lemma 7. *For all $r \in [0,\infty)$, we have*

$$\sup_{h\in\mathcal{H}} h(r) \le \begin{cases} \min(\sqrt{2}, 1/r) & \text{if } p = 1 \\ \min\Bigl\{\frac{(p+1)^{p/2}}{(p-1)!} r^{p-1}\,, 24r\,, \frac{p}{r}\Bigr\} & \text{if } p \ge 2. \end{cases} \tag{3}$$

Remark: The only difference between the cases $p = 1$ and $p \ge 2$ is that the bound $\sup_{h\in\mathcal{H}} h(r) \le 24r$ does not hold when $p = 1$. The bounds $\frac{(p+1)^{p/2}}{(p-1)!} r^{p-1}$ and p/r are sharp when $r = 0$ and $r = (p+2)^{1/2}$ respectively. The first of these facts is trivial unless $p = 1$, but in that case one can observe that if we define $h : [0,\infty) \to [0,\infty)$ by $h(r) := \sqrt{2}e^{-\sqrt{2}r}$ then $h \in \mathcal{H}$ and $h(0) = \sqrt{2}$. The second fact follows because if we define $h : [0,\infty) \to [0,\infty)$ by $h(r) := \frac{p}{(p+2)^{p/2}} r^{p-1} \mathbb{1}_{\{r \in [0,(p+2)^{1/2}]\}}$, then $h \in \mathcal{H}$ and $h(\sqrt{p+2}) = p/(p+2)^{1/2}$.

Remark for us: The second bound in (3) seems to be unnecessary.

Proof. For the first bound in (3) (treating the cases $p = 1$ and $p \ge 2$ simultaneously), for $r \in [0,\infty)$, let

$$g_0^*(r) := \frac{(p+1)^{p/2}}{c_p (p-1)!} e^{-(p+1)^{1/2} r},$$

so $g_0^* \in \mathcal{G}$, and let $h_0^*(r) := c_p r^{p-1} g_0^*(r)$. Then h_0^* is the $\Gamma(p,(p+1)^{1/2})$ density, so $h_0^* \in \mathcal{H}$. Suppose for a contradiction that $g \in \mathcal{G}$ satisfies the conditions the function $h : [0,\infty) \to [0,\infty)$ given by $h(r) := c_p r^{p-1} g(r)$ belongs to $\mathcal{H}$, and $g(0) > g_0^*(0)$. Then since $\log g_0^*$ is an affine function and h is a log-concave density, there exists $r_0 \in (0,\infty)$ such that $g(r) > g_0^*(r)$ for $r < r_0$ and $g(r) < g_0^*(r)$ for $r > r_0$. But then $h <_{\mathrm{st}} h^*$, so $c_p \int_0^\infty r^{p+1} g(r)\,dr < p$, which establishes our desired contradiction. But since every $\phi \in \Phi$ is decreasing, it follows that $r \mapsto \sup_{g\in\mathcal{G}} g(r)$ is decreasing, so

$$\sup_{h\in\mathcal{H}} h(r) = c_p \sup_{g\in\mathcal{G}} r^{p-1} g(r) \le c_p r^{p-1} \sup_{g\in\mathcal{G}} g(0) = c_p r^{p-1} g_0^*(0) = \frac{(p+1)^{p/2}}{(p-1)!} r^{p-1}.$$

Next we establish the third bound in (3), again treating $p = 1$ and $p \geq 2$ simultaneously. For $a \in (0, \infty)$ and $r \in (0, \infty)$, consider the function

$$g_a(r) := \frac{p}{c_p a^p} \mathbb{1}_{\{r \in [0,a]\}}.$$

Then $g_a \in \mathcal{G}$ and $c_p \int_0^\infty r^{p-1} g_a(r)\, dr = 1$. Thus if $g \in \mathcal{G}$ satisfies $g(a) > g_a(a)$, then $g(r) > g_a(r)$ for all $r \in [0, a]$ and $g(r) \geq g_a(r)$ for all $r \in [0, \infty)$. But then $c_p \int_0^\infty r^{p-1} g(r)\, dr > 1$, so the function $h : [0, \infty) \to [0, \infty)$ given by $h(r) := c_p r^{p-1} g(r)$ does not belong to $\mathcal{H}$. We deduce that for every $r \in (0, \infty)$,

$$\sup_{h \in \mathcal{H}} h(r) \leq c_p r^{p-1} g_r(r) = \frac{p}{r}.$$

Finally, we prove the second bound in (3) in the case $p \geq 2$. To this end, fix $M \geq \log 16$, and $m \in (-\infty, M-2]$. Suppose that $h \in \mathcal{H}$ satisfies $\log h(r_0) \geq M$ for some $r_0 \in (1/4, p^{1/2}]$, and for $t \in [m, M]$, let $D_t := \{r \in [0, \infty) : \log h(r) \geq t\}$. First note that for any $t \in [m, M]$ and $r \in D_m$, we have

$$\log h\biggl(\frac{t-m}{M-m} r_0 + \frac{M-t}{M-m} r\biggr) \geq \frac{(t-m)M}{M-m} + \frac{(M-t)m}{M-m} = t.$$

Hence, writing μ for Lebesgue measure on $\mathbb{R}$,

$$\mu(D_t) \geq \mu\biggl(\frac{t-m}{M-m} r_0 + \frac{M-t}{M-m} D_m\biggr) = \frac{M-t}{M-m} \mu(D_m).$$

Using Fubini's theorem, we can now compute

$$\begin{aligned}
1 &\geq \int_{D_m} h(r) - e^m \, dr \geq \int_{D_m} \int_m^M e^s \mathbb{1}_{\{\log h(r) \geq s\}} \, ds \, dr \\
&= \int_m^M e^s \mu(D_s) \, ds \geq \frac{\mu(D_m)}{M-m} \int_m^M (M-s) e^s \, ds = \frac{\mu(D_m) e^M}{M-m} \int_0^{M-m} t e^{-t} \, dt \\
&\geq \frac{\mu(D_m) e^M}{2(M-m)}.
\end{aligned}$$

Since D_m is an interval containing r_0, we conclude that $\log h(r) \leq m$ whenever $|r - r_0| \geq 2(M-m)e^{-M}$. Thus

$$\log h(r) \leq M - \frac{|r - r_0| e^M}{2}$$

for $|r - r_0| \geq 4e^{-M}$. Noting that $r_0 - 4e^{-M} > 0$ and using the bound $h(r) \leq p/r$, it now follows that

$$
\begin{aligned}
p = \int_0^\infty r^2 h(r)\,dr &\leq \int_0^{r_0 - 4e^{-M}} r^2 \exp\biggl\{M - \frac{(r_0 - r)e^M}{2}\biggr\}\,dr + p\int_{r_0-4e^{-M}}^{r_0+4e^{-M}} r\,dr \\
&\qquad\qquad\qquad\qquad + \int_{r_0+4e^{-M}}^\infty r^2 \exp\biggl\{M - \frac{(r-r_0)e^M}{2}\biggr\}\,dr \\
&\leq 2\int_2^\infty \biggl(r_0 - \frac{2s}{e^M}\biggr)^2 e^{-s}\,ds + 8e^{-M}r_0 p + 2\int_2^\infty \biggl(r_0 + \frac{2s}{e^M}\biggr)^2 e^{-s}\,ds \\
&= 4e^{-2}r_0^2 + 32e^{-2M} + 8e^{-M}r_0 p \leq p\biggl(\frac{2}{3} + 8e^{-M}r_0\biggr).
\end{aligned}
$$

We deduce that $e^{-M}r_0 \geq 1/24$, so $h(r) \leq \min(16, 24r)$ for $r \in (1/4, p^{1/2}]$. But our first bound in (3) is at most $5r$ for $r \leq 1$ and $p \geq 2$, and the conclusion follows. $\square$

Corollary 8. *Let $Z \sim h \in \mathcal{H}$. Then there exists a universal constant $c_0 > 0$ such that $\mathrm{Var}(Z) \geq c_0 p^{-1}$.*

Remark: Define $h : [0,\infty) \to [0,\infty)$ by $h(r) := \frac{p}{(p+2)^{p/2}} r^{p-1} \mathbb{1}_{\{r \in [0,(p+2)^{1/2}]\}}$. Then it can be shown that $h \in \mathcal{H}$, and if $Z \sim h$, then $\mathrm{Var}(Z) = p/(p+1)^2$. Thus the bound given in Corollary 8 is sharp in terms of its dependence on p.

Proof. From the first bound in Lemma 7, we have

$$\sup_{h \in \mathcal{H}} \sup_{r \in [0,\infty)} h(r) \leq \sqrt{2}$$

for $r \leq p^{1/2}/e$. Write $\mu := \mathbb{E}(Z)$ and $\sigma^2 := \mathrm{Var}(Z)$. By Lovász and Vempala [2007, Theorem 5.14(d)], we have

$$\frac{1}{128\sigma} \leq h(\mu) \leq \sup_{h \in \mathcal{H}} \sup_{r \in [0,\infty)} h(r) \leq e p^{1/2}.$$

The result follows. $\square$

An upper bound on the variance of $Z \sim h \in \mathcal{H}$ is readiy available.

Lemma 9. *Bobkov [2003, Lemma 1] Suppose h is a density of the form $r^{p-1}g(r)c_p$ for some log-concave function $g(r)$, suppose $Z \sim h$, then,*

$$\mathrm{Var}(Z) \leq \frac{1}{p}(\mathbb{E}Z)^2$$

Under our constraint on $h \in \mathcal{H}$, we have that $(\mathbb{E}Z)^2 \leq \mathbb{E}[Z^2] = p$. This gives us the following corollary:

Corollary 10. *Let $h \in \mathcal{H}$ and suppose $Z \sim h$. Then,*

$$\mathrm{Var}(Z) \leq 1$$

The upper bound is also tight. If we let $g(r) = e^{-ar}c$ where $a = \sqrt{\frac{(p+2)(p+1)}{p}}$ and c be chosen such that $cc_p = \frac{a^p}{\Gamma(p)}$, then we have that the mean is $\sqrt{\frac{p^3}{(p+2)(p+1)}}$ and the variance is $\frac{p^2}{(p+2)(p+1)}$. Thus, the variance of our chosen $g(r)$ gets arbitrarily close to 1 for increasing p.

We need one more ingredient before we can state our envelope bound.

Lemma 11. *Let $\mathcal{F}^{\mu,\sigma^2} = \{f$ log-concave density : $\mu_f = \mu,\ \sigma_f^2 = \sigma^2\}$. Then, there exists universal constants A, B such that*

$$\sup_{f \in \mathcal{F}^{\mu,\sigma^2}} f(x) \leq \frac{A}{\sigma} \exp\left(-B\frac{|x-\mu|}{\sigma}\right)$$

Proof. This follows directly from Kim et al. [2016b, Theorem 2] by specializing to $d = 1$ and performing a change of variables. □

The following theorem gives an envelope bound for the density class $\mathcal{H}$.

Theorem 12. *For any absolute constant $c_1, c_2 > 0$, there exists constants C_1, C_2 such that*

$$\sup_{f \in \mathcal{H}} f(x) \leq \begin{cases} \frac{A'}{c_0}\sqrt{p} & \left|x - \sqrt{p}\right| \leq \frac{c_0}{B\sqrt{p}} \\ \frac{A'}{eB}\frac{1}{|x-\sqrt{p}|} & \frac{c_0}{B\sqrt{p}} \leq \left|x - \sqrt{p}\right| \leq \frac{1}{B} \\ A'e^{-B|x-\sqrt{p}|} & \frac{1}{B} \leq \left|x - \sqrt{p}\right| \end{cases}$$

Proof. Define $\mathcal{H}_\sigma = \{f \in \mathcal{H} : \sigma_f = \sigma\}$ as the sub-class of $\mathcal{H}$ in which the densities have standard deviation σ. It is clear that $\mathcal{H}_\sigma = \emptyset$ for $\sigma \notin [\frac{c_0}{\sqrt{p}}, 1]$ by our upper and lower bounds on the variance of densities in $\mathcal{H}$ (Corollary 8, 10).

First, we observe

$$\sup_{f \in \mathcal{H}} f(x) = \sup_{\sigma \in [\frac{c_0}{\sqrt{p}}, 1]} \sup_{f \in \mathcal{H}_\sigma} f(x)$$

And, by Lemma 11 and by the fact that $\mu = \sqrt{\mathbb{E}Z^2 - \sigma^2} = \sqrt{p - \sigma^2}$,

$$
\begin{aligned}
\sup_{f \in \mathcal{H}_\sigma} f(x) &\leq \frac{A}{\sigma} \exp\left(-B \frac{|x - \sqrt{p - \sigma^2}|}{\sigma} \right) \\
&\leq \frac{A}{\sigma} \exp\left(-B \frac{|x - \sqrt{p}| - (\sqrt{p} - \sqrt{p - \sigma^2})}{\sigma} \right) \\
&\leq \frac{A}{\sigma} \exp\left(-B \frac{|x - \sqrt{p}|}{\sigma} + B \frac{\frac{\sigma^2}{2\sqrt{p - \sigma^2}}}{\sigma} \right) \\
&= \frac{A}{\sigma} \exp\left(-B \frac{|x - \sqrt{p}|}{\sigma} + B \frac{\sigma}{2\sqrt{p - \sigma^2}} \right) \\
&\leq \frac{A}{\sigma} \exp\left(-B \frac{|x - \sqrt{p}|}{\sigma} + B \frac{1}{2\sqrt{p - 1}} \right) \\
&\leq \frac{A'}{\sigma} \exp\left(-B \frac{|x - \sqrt{p}|}{\sigma} \right)
\end{aligned}
$$

The second inequality follows from triangle inequality. The third inequality follows because $\sqrt{p} - \sqrt{p - \sigma^2} \leq \frac{d\sqrt{p - \sigma^2}}{dp} \sigma^2$ by concavity of the square root function. The fourth inequality follows because $\sigma \leq 1$. And, in the last inequality, we define A' to be a constant since $B \frac{1}{2\sqrt{p-1}}$ is bounded by a constant for all p.

Let $H_\sigma(x) = \frac{A'}{\sigma} \exp\left(-B \frac{|x - \sqrt{p}|}{\sigma} \right)$.

If we write $\nu = \frac{1}{\sigma}$, then $\log H_\sigma(x)$ is a concave function of ν. We solve for the optimal and get that $\sigma = \frac{1}{\sqrt{p}}$ if $B|x - \sqrt{p}| \leq \sqrt{c_0}\sqrt{p}$, $\sigma = B|x - \sqrt{p}|$ if $\frac{c_0}{\sqrt{p}} \leq B|x - \sqrt{p}| \leq 1$, and $\sigma = 1$ if $1 \leq B|x - \sqrt{p}|$.

Thus, if $B|x - \sqrt{p}| \leq \frac{c_0}{\sqrt{p}}$, we have that $\sup_{\sigma \in [c_0/\sqrt{p}, 1]} H_\sigma(x) \leq A' \frac{\sqrt{p}}{c_0}$. If $\frac{c_0}{\sqrt{p}} \leq B|x - \sqrt{p}| \leq 1$, we have that $\sup_{\sigma \in [c_0/\sqrt{p}, 1]} H_\sigma(x) \leq \frac{A'}{eB|x - \sqrt{p}|}$. If $1 \leq B|x - \sqrt{p}|$, then $\sup_{\sigma \in [c_0/\sqrt{p}, 1]} H_\sigma(x) \leq A' \exp(-B|x - \sqrt{p}|)$.

□

6 Bracketing Entropy

We start with a proposition from Kim et al. [2016a]. Let $\mathcal{F}([a, b], -\infty, B)$ be the set of log-concave functions f such that f is supported on $[a, b]$ and that $\log f(x) \leq B$.

Proposition 13. *(Kim et al. [2016a, Proposition 14])*
There exists a universal constant $C > 0$ such that

$$
H_{[]}(\epsilon, \mathcal{F}([a, b], -\infty, B), d_H, [a, b]) \leq C(1 + B^{1/2}) \frac{e^{B/4}(b - a)^{1/4}}{\epsilon^{1/2}}
$$

We can slightly improve the result through a scaling argument.

Corollary 14. *There exists a universal constant $C > 0$ such that*

$$H_{[]}(\epsilon, \mathcal{F}([a,b], -\infty, B), d_H, [a,b]) \le C \frac{e^{B/4}(b-a)^{1/4}}{\epsilon^{1/2}}$$

Proof. Let $\sigma > 0$. For $f \in \mathcal{F}([a,b], -\infty, B)$, define $f_\sigma(x) = \frac{1}{\sigma} f(\frac{x}{\sigma})$ and define $\mathcal{F}_\sigma([a,b], -\infty, B) = \{f_\sigma \,:\, f \in \mathcal{F}([a,b], -\infty, B)\}$.

Since the Hellinger distance d_H is affine invariant, we have that

$$H_{[]}(\epsilon, \mathcal{F}([a,b], -\infty, B), d_H, [a,b]) = H_{[]}(\epsilon, \mathcal{F}_\sigma([a,b], -\infty, B), d_H, [a,b])$$

However, we also know that $\mathcal{F}_\sigma([a,b], -\infty, B) \subset \mathcal{F}([\sigma a, \sigma b], -\infty, B + \log \sigma)$. Thus,

$$\begin{aligned} H_{[]}(\epsilon, \mathcal{F}([a,b], -\infty, B), d_H, [a,b]) &\le H_{[]}(\epsilon, \mathcal{F}([\sigma a, \sigma b], -\infty, B + \log \sigma), d_H, [a,b]) \\ &\le C(1 + (B + \log \sigma)^{1/2}) \frac{e^{B/4}(b-a)^{1/4}}{\epsilon^{1/2}} \end{aligned}$$

Since this holds true for all $\sigma > 0$, the corollary follows. □

References

Sergey G Bobkov. Spectral gap and concentration for some spherically symmetric probability measures. In *Geometric aspects of functional analysis*, pages 37–43. Springer, 2003.

Arlene KH Kim, Adityanand Guntuboyina, and Richard J Samworth. Adaptation in log-concave density estimation. *arXiv preprint arXiv:1609.00861*, 2016a.

Arlene KH Kim, Richard J Samworth, et al. Global rates of convergence in log-concave density estimation. *The Annals of Statistics*, 44(6):2756–2779, 2016b.

László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms*, 30(3):307–358, 2007.