On Anomaly Ranking and Excess-Mass Curves, Supplementary Material

Nicolas Goix
UMR LTCI No. 5141
Telecom ParisTech/CNRS
Institut Mines-Telecom
Paris, 75013, France

Anne Sabourin
UMR LTCI No. 5141
Telecom ParisTech/CNRS
Institut Mines-Telecom
Paris, 75013, France

Stéphan Clémençon
UMR LTCI No. 5141
Telecom ParisTech/CNRS
Institut Mines-Telecom
Paris, 75013, France

1 Illustrations

Note that the scoring function we built in Algorithm 1 is an estimator of the density f (usually called the silhouette), since $f(x) = \int_0^\infty \mathbb{1}_{f\geq t} dt = \int_0^\infty \mathbb{1}_{\Omega_t^*} dt$ and $s(x) := \sum_{k=1}^K (t_k - t_{k-1}) \mathbb{1}_{x\in\hat{\Omega}_{t_k}}$ which is a discretization of $\int_0^\infty \mathbb{1}_{\hat{\Omega}_t} dt$. This fact is illustrated in Fig. 1

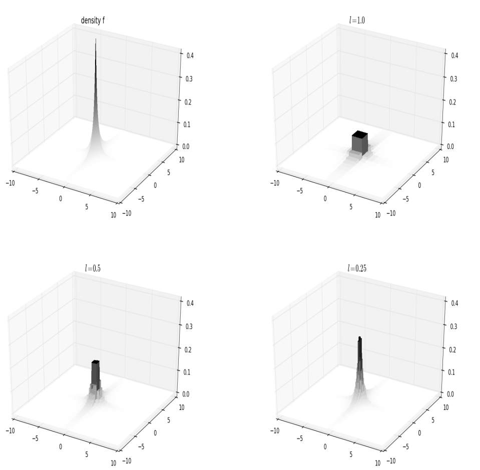

Figure 1: density and scoring functions

2 Detailed Proofs

Proof of Proposition 1

Let $t > 0$. Recall that $EM^*(t) = \alpha(t) - t\lambda(t)$ where $\alpha(t)$ denote the mass at level t, namely $\alpha(t) = \mathbb{P}(f(X) \geq t)$, and $\lambda(t)$ denote the volume at level t, i.e. $\lambda(t) = Leb(\{x, f(x) \geq t\})$. For $h > 0$, let $A(h)$ denote the quantity $A(h) = 1/h(\alpha(t+h) - \alpha(t))$ and $B(h) = 1/h(\lambda(t+h) - \lambda(t))$. It is straightforward to see that $A(h)$ and $B(h)$ converge when $h \to 0$, and expressing $EM^{*'} = \alpha'(t) - t\lambda'(t) - \lambda(t)$, it suffices to show that $\alpha'(t) - t\lambda'(t) = 0$, namely $\lim_{h\to 0} A(h) - t\ B(h) = 0$. Now we have $A(h) - t\ B(h) = \frac{1}{h}\int_{t\leq f\leq t+h} f - t \leq \frac{1}{h}\int_{t\leq f\leq t+h} h = Leb(t \leq f \leq t+h) \to 0$ because f has no flat part.

Proof of Lemma 1:

On the one hand, for every Ω measurable,

$$\begin{aligned}
\mathbb{P}(X \in \Omega) - t\ Leb(\Omega) &= \int_\Omega (f(x) - t) dx \\
&\leq \int_{\Omega\cap\{f\geq t\}} (f(x) - t) dx \\
&\leq \int_{\{f\geq t\}} (f(x) - t) dx \\
&= \mathbb{P}(f(X) \geq t) - t\ Leb(\{f \geq t\}).
\end{aligned}$$

It follows that $\{f \geq t\} \in \arg\max_{A meas.} \mathbb{P}(X \in A) - t\ Leb(A)$.

On the other hand, suppose $\Omega \in \arg\max_{A\ meas.} \mathbb{P}(X \in A) - t\ Leb(A)$ and $Leb(\{f > t\} \setminus \Omega) > 0$. Then there is $\epsilon > 0$ such that $Leb(\{f > t + \epsilon\} \setminus \Omega) > 0$ (by subadditivity of Leb, if it is not the case, then $Leb(\{f > t\} \setminus \Omega) = Leb(\cup_{\epsilon\in\mathbb{Q}_+} \{f > t + \epsilon\} \setminus \Omega) = 0$). We have thus

$$\int_{\{f>t\}\setminus\Omega} (f(x) - t) dx > \epsilon . Leb(\{f > t + \epsilon\} \setminus \Omega) > 0 ,$$

so that

$$\int_\Omega (f(x)-t)dx \;\leq\; \int_{\{f>t\}}(f(x)-t)dx - \int_{\{f>t\}\setminus\Omega}(f(x)-t)dx < \int_{\{f>t\}}(f(x)-t)dx\;,$$

i.e

$$\mathbb{P}(X\in\Omega) - t\; Leb(\Omega) < \mathbb{P}(f(X)\geq t) - t\; Leb(\{x, f(x)\geq t\})$$

which is a contradiction: $\{f>t\}\subset\Omega$ Leb-a.s. .

To show that $\Omega_t^*\subset\{x, f(x)\geq t\}$, suppose that $Leb(\Omega_t^*\cap\{f<t\})>0$. Then by sub-additivity of Leb just as above, there is $\epsilon>0$ s.t $Leb(\Omega_t^*\cap\{f<t-\epsilon\})>0$ and $\int_{\Omega_t^*\cap\{f<t-\epsilon\}} f-t\leq -\epsilon.Leb(\Omega_t^*\cap\{f<t-\epsilon\})<0$. It follows that $\mathbb{P}(X\in\Omega_t^*)-t\;Leb(\Omega_t^*)<\mathbb{P}(X\in\Omega_t^*\setminus\{f<t-\epsilon\})-t\;Leb(\Omega_t^*\setminus\{f<t-\epsilon\})$ which is a contradiction with the optimality of Ω_t^*.

Proof of Proposition 2

Proving the first assertion is immediate, since $\int_{f\geq t}(f(x)-t)dx\geq\int_{s\geq t}(f(x)-t)dx$. Let us now turn to the second assertion. We have:

$$EM^*(t)-EM_s(t) = \int_{f>t}(f(x)-t)dx - \sup_{u>0}\int_{s>u}(f(x)-t)dx = \inf_{u>0}\int_{f>t}(f(x)-t)dx - \int_{s>u}(f(x)-t)dx\;,$$

yet:

$$\int_{\{f>t\}\setminus\{s>u\}}(f(x)-t)dx + \int_{\{s>u\}\setminus\{f>t\}}(t-f(x))dx \leq (\|f\|_\infty - t).Leb\Big(\{f>t\}\setminus\{s>u\}\Big) + \;t\;Leb\Big(\{s>u\}\setminus\{f>t\}\Big),$$

so we obtain:

$$EM^*(t)-EM_s(t)\leq\;\max(t,\|f\|_\infty-t)\times Leb\Big(\{s>u\}\Delta\{f>t\}\Big) \leq\; \|f\|_\infty.Leb\Big(\{s>u\}\Delta\{f>t\}\Big).$$

To prove the third point, note that:

$$\inf_{u>0} Leb\Big(\{s>u\}\Delta\{f>t\}\Big) = \inf_{T\nearrow} Leb\Big(\{Ts>t\}\Delta\{f>t\}\Big)$$

Yet,

$$\begin{aligned}&Leb\Big(\{Ts>t\}\Delta\{f>t\}\Big)\\ &\leq Leb(\{f>t-\|Ts-f\|_\infty\}\setminus\{f>t+\|Ts-f\|_\infty\})\\ &=\lambda(t-\|Ts-f\|_\infty)\;-\;\lambda(t+\|Ts-f\|_\infty)\\ &=-\int_{t-\|Ts-f\|_\infty}^{t+\|Ts-f\|_\infty}\lambda'(u)du\;.\end{aligned}$$

On the other hand, we have $\lambda(t)=\int_{\mathbb{R}^d}\mathbb{1}_{f(x)\geq t}dx=\int_{\mathbb{R}^d}g(x)\|\nabla f(x)\|dx$ where we let $g(x)=\frac{1}{\|\nabla f(x)\|}\mathbb{1}_{\{x,\|\nabla f(x)\|>0, f(x)\geq t\}}$. The co-area formula (see [1], p.249, th3.2.12) gives in this case: $\lambda(t)=\int_{\mathbb{R}}du\int_{f^{-1}(u)}\frac{1}{\|\nabla f(x)\|}\mathbb{1}_{\{x,f(x)\geq t\}}d\mu(x)=\int_t^\infty du\int_{f^{-1}(u)}\frac{1}{\|\nabla f(x)\|}d\mu(x)$ so that $\lambda'(t)=-\int_{f^{-1}(u)}\frac{1}{\|\nabla f(x)\|}d\mu(x)$.

Let η_ϵ such that $\forall u>\epsilon$, $|\lambda'(u)|=\int_{f^{-1}(u)}\frac{1}{\|\nabla f(x)\|}d\mu(x)<\eta_\epsilon$. We obtain:

$$\sup_{t\in[\epsilon+\inf_{T\nearrow}\|f-Ts\|_\infty,\|f\|_\infty]} EM^*(t)-EM_s(t) \leq 2.\eta_\epsilon.\|f\|_\infty\inf_{T\nearrow}\|f-Ts\|_\infty.$$

In particular, if $\inf_{T\nearrow}\|f-Ts\|_\infty\leq\epsilon_1$,

$$\sup_{[\epsilon+\epsilon_1,\|f\|_\infty]}|EM^*-EM_s|\leq 2.\eta_\epsilon.\|f\|_\infty.\inf_{T\nearrow}\|f-Ts\|_\infty\;.$$

Proof of Proposition 3

Let i in $\{1,...,K\}$. First, note that:

$$H_{n,t_{i+1}}(\hat\Omega_{t_{i+1}}\cup\hat\Omega_{t_i}) = H_{n,t_{i+1}}(\hat\Omega_{t_{i+1}}) + H_{n,t_{i+1}}(\hat\Omega_{t_i}\setminus\hat\Omega_{t_{i+1}}),$$

$$H_{n,t_i}(\hat\Omega_{t_{i+1}}\cap\hat\Omega_{t_i}) = H_{n,t_i}(\hat\Omega_{t_i}) - H_{n,t_i}(\hat\Omega_{t_i}\setminus\hat\Omega_{t_{i+1}}).$$

It follows that

$$\begin{aligned}&H_{n,t_{i+1}}(\hat\Omega_{t_{i+1}}\cup\hat\Omega_{t_i}) + H_{n,t_i}(\hat\Omega_{t_{i+1}}\cap\hat\Omega_{t_i})\\ &= H_{n,t_{i+1}}(\hat\Omega_{t_{i+1}}) + H_{n,t_i}(\hat\Omega_{t_i}) + H_{n,t_{i+1}}(\hat\Omega_{t_i}\setminus\hat\Omega_{t_{i+1}}) - H_{n,t_i}(\hat\Omega_{t_i}\setminus\hat\Omega_{t_{i+1}})\;,\end{aligned}$$

with $H_{n,t_{i+1}}(\hat\Omega_{t_i}\setminus\hat\Omega_{t_{i+1}})-H_{n,t_i}(\hat\Omega_{t_i}\setminus\hat\Omega_{t_{i+1}})\geq 0$ since $H_{n,t}$ is decreasing in t. But on the other hand, by definition of $\hat{\Omega}_{t_{i+1}}$ and $\hat{\Omega}_{t_i}$ we have:

$$H_{n,t_{i+1}}(\hat{\Omega}_{t_{i+1}} \cup \hat{\Omega}_{t_i}) \leq H_{n,t_{i+1}}(\hat{\Omega}_{t_{i+1}})\,,$$
$$H_{n,t_i}(\hat{\Omega}_{t_{i+1}} \cap \hat{\Omega}_{t_i}) \leq H_{n,t_i}(\hat{\Omega}_{t_i})\,.$$

Finally we get:

$$H_{n,t_{i+1}}(\hat{\Omega}_{t_{i+1}} \cup \hat{\Omega}_{t_i}) = H_{n,t_{i+1}}(\hat{\Omega}_{t_{i+1}})\,,$$
$$H_{n,t_i}(\hat{\Omega}_{t_{i+1}} \cap \hat{\Omega}_{t_i}) = H_{n,t_i}(\hat{\Omega}_{t_i})\,.$$

Proceeding by induction we have, for every m such that $k+m \leq K$:

$$H_{n,t_{i+m}}(\hat{\Omega}_{t_i} \cup \hat{\Omega}_{t_{i+1}} \cup ... \cup \hat{\Omega}_{t_{i+m}}) = H_{n,t_{i+m}}(\hat{\Omega}_{t_{i+m}})\ ,$$
$$H_{n,t_i}(\hat{\Omega}_{t_i} \cap \hat{\Omega}_{t_{i+1}} \cap ... \cap \hat{\Omega}_{t_{i+m}}) = H_{n,t_i}(\hat{\Omega}_{t_i})\ .$$

Taking (i=1, m=k-1) for the first equation and (i=k, m=K-k) for the second completes the proof.

Proof of Theorem 1

We shall use the following lemma:

Lemma 2.1. *With probability at least* $1-\delta$, *for* $k \in \{1,...,K\}$, $0 \leq EM^*(t_k) - EM_{s_K}(t_k) \leq 2\Phi_n(\delta)$.

Proof of Lemma 2.1:

Remember that by definition of $\hat{\Omega}_{t_k}$: $H_{n,t_k}(\hat{\Omega}_{t_k}) = \max_{\Omega\in\mathcal{G}} H_{n,t_k}(\Omega)$ and note that:

$$EM^*(t_k) = \max_{\Omega\ meas.} H_{t_k}(\Omega) = \max_{\Omega\in\mathcal{G}} H_{t_k}(\Omega) \geq H_{t_k}(\hat{\Omega}_{t_k}).$$

On the other hand, using (5), with probability at least $1-\delta$, for every $G\in\mathcal{G}$, $|\mathbb{P}(G)-\mathbb{P}_n(G)| \leq \Phi_n(\delta)$. Hence, with probability at least $1-\delta$, for all $\Omega\in\mathcal{G}$:

$$H_{n,t_k}(\Omega) - \Phi_n(\delta) \leq H_{t_k}(\Omega) \leq H_{n,t_k}(\Omega) + \Phi_n(\delta)$$

so that, with probability at least $(1-\delta)$, for $k \in \{1..,K\}$,

$$\begin{aligned} H_{n,t_k}(\hat{\Omega}_{t_k}) - \Phi_n(\delta) &\leq H_{t_k}(\hat{\Omega}_{t_k}) \\ &\leq EM^*(t_k) \\ &\leq H_{n,t_k}(\hat{\Omega}_{t_k}) + \Phi_n(\delta)\ , \end{aligned}$$

whereby, with probability at least $(1-\delta)$, for $k \in \{1,..,K\}$,

$$0 \leq EM^*(t_k) - H_{t_k}(\hat{\Omega}_{t_k})\ \leq\ 2\Phi_n(\delta)\ .$$

The following Lemma is a consequence of the derivative property of EM^* (Proposition 1)

Lemma 2.2. *Let* k *in* $\{1,...,K-1\}$. *Then for every* t *in* $]t_{k+1},t_k]$, $0 \leq EM^*(t) - EM^*(t_k) \leq \lambda(t_{k+1})(t_k - t_{k+1})$.

Combined with Lemma 2.1 and the fact that EM_{s_K} is non-increasing, and writing $EM^*(t) - EM_{s_K}(t) = (EM^*(t) - EM^*(t_k)) + (EM^*(t_k) - EM_{s_K}(t_k)) + (EM_{s_K}(t_k) - EM_{s_K}(t))$ this result leads to:

$$\forall k \in\ \{0,...,K-1\},\ \forall t \in\ ]t_{k+1},t_k],$$
$$0 \leq EM^*(t) - EM_{s_K}(t) \leq 2\Phi_n(\delta) + \lambda(t_{k+1})(t_k - t_{k+1})$$

which gives Lemma 2 stated in section Technical Details. Notice that we have not yet used the fact that f has a compact support.

The compactness support assumption allows an extension of Lemma 2.2 to $k=K$, namely the inequality holds true for t in $]t_{K+1},t_K] =]0,t_K]$ as soon as we let $\lambda(t_{K+1}) := Leb(suppf)$. Indeed the compactness of $suppf$ implies that $\lambda(t) \to Leb(suppf)$ as $t \to 0$. Observing that Lemma 2.1 already contains the case $k=K$, this leads to, for k in $\{0,...,K\}$ and $t \in\ ]t_{k+1},t_k]$, $|EM^*(t) - EM_{s_K}(t)| \leq 2\Phi_n(\delta) + \lambda(t_{k+1})(t_k - t_{k+1})$. Therefore, λ being a decreasing function bounded by $\lambda(Leb(suppf))$, we obtain the following: with probability at least $1-\delta$, we have for all t in $]0,t_1]$:

$$\begin{aligned} &|\mathrm{EM}^*(t) - \mathrm{EM}_{s_K}(t)| \\ &\leq \left(A + \sqrt{2log(1/\delta)}\right)\frac{1}{\sqrt{n}} \\ &\qquad\qquad + \lambda(Leb(suppf)) \sup_{1\leq k\leq K}(t_k - t_{k+1}). \end{aligned}$$

Proof of Theorem 2

The first part of this theorem is a consequence of (10) combined with:

$$\begin{aligned} \sup_{t\in]0,t_N]} |EM^*(t) - EM_{s_N}(t)|\ &\leq\ 1 - EM_{s_N}(t_N) \\ &\leq\ 1 - EM^*(t_N) + 2\Phi_n(\delta)\ , \end{aligned}$$

where we use the fact that $0 \leq EM^*(t_N) - EM_{s_N}(t_N) \leq 2\Phi_n(\delta)$ following from Lemma 2.1.
To see the convergence of $s_N(x)$, note that:

$$\begin{aligned} s_N(x)\ &=\ \frac{t_1}{\sqrt{n}}\sum_{k=1}^{\infty}\frac{1}{(1+\frac{1}{\sqrt{n}})^k}\mathbb{1}_{x\in\hat{\Omega}_{t_k}}\mathbb{1}_{\{k\leq N\}} \\ &\leq\ \frac{t_1}{\sqrt{n}}\sum_{k=1}^{\infty}\frac{1}{(1+\frac{1}{\sqrt{n}})^k}\ <\ \infty, \end{aligned}$$

and analogically to remark 1 observe that $EM_{s_N} \leq EM_{s_\infty}$ so that $\sup_{t\in]0,t_1]}|EM^*(t) - EM_{s_\infty}(t)| \leq \sup_{t\in]0,t_1]}|EM^*(t) - EM_{s_N}(t)|$ which prooves the last part of the theorem.

Proof of Lemma 3

By definition, for every class of set $\mathcal{H}$, $EM^*_{\mathcal{H}}(t) = \max_{\Omega \in \mathcal{H}} H_t(\Omega)$. The bias $EM^*(t) - EM^*_{\mathcal{G}}(t)$ of the model $\mathcal{G}$ is majored by $EM^*(t) - EM^*_{\mathcal{F}}(t)$ since $\mathcal{F} \subset \mathcal{G}$. Remember that $f_F(x) := \sum_{i \geq 1} \mathbb{1}_{x \in F_i} \frac{1}{|F_i|} \int_{F_i} f(y)dy$ and note that for all $t > 0$, $\{f_F > t\} \in \mathcal{F}$. It follows that:

$$EM^*(t) - EM^*_{\mathcal{F}}(t) = \int_{f>t} (f-t) - \sup_{C \in \mathcal{F}} \int_C (f-t)$$

$$\leq \int_{f>t} (f-t) - \int_{f_F>t} (f-t) \text{ since } \{f_F > t\} \in \mathcal{F}$$

$$= \int_{f>t} (f-t) - \int_{f_F>t} (f_F - t)$$

$$\text{since } \forall G \in \mathcal{F}, \int_G f = \int_G f_F$$

$$= \int_{f>t} (f-t) - \int_{f>t} (f_F - t) + \int_{f>t} (f_F - t) - \int_{f_F>t} (f_F - t)$$

$$= \int_{f>t} (f - f_F) + \int_{\{f>t\} \setminus \{f_F>t\}} (f_F - t) - \int_{\{f_F>t\} \setminus \{f>t\}} (f_F - t) \ .$$

Observe that the second and the third term in the bound are non-positive. Therefore:

$$EM^*(t) - EM^*_{\mathcal{F}}(t) \leq \int_{f>t} (f - f_F) \leq \int_{\mathbb{R}^d} |f - f_F| \ .$$

References

[1] H. Federer. *Geometric Measure Theory*. Springer, 1969.