Adaptive and minimax optimal estimation of the tail coefficient

Alexandra Carpentier and Arlene K. H. Kim

University of Cambridge

Supplementary Material

This is a supplementary material for the paper: "Adaptive and minimax optimal estimation of the tail coefficient."

S1. Important preliminary result

Lemma 1 contains a classical and simple, yet important result for the paper.

Lemma 1 (Bernstein inequality for Bernoulli random variables)**.** *Let $X_1, \ldots, X_n$ be i.i.d. observations from F, and we define $p_k = 1 - F(e^k)$ and $\hat{p}_k = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_i > e^k\}$. Let $\delta > 0$ and also let n be large enough so that $p_k \geq \frac{4\log(2/\delta)}{n}$. Then with probability $1-\delta$,*

$$|\hat{p}_k - p_k| \leq 2\sqrt{\frac{p_k \log(2/\delta)}{n}}. \tag{S1.1}$$

Proof of Lemma 1. The proof is using Bernstein inequality (e.g. see Lemma 19.32 of Van der Vaart (2000)) of the following form; for any bounded, measurable function g, we have for every $t > 0$,

$$\mathbb{P}\left(\left|\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n g(X_i) - \mathbb{E}g(X)\right)\right| > t\right) \leq 2\exp\left(-\frac{1}{4}\frac{t^2}{\mathbb{E}g^2 + t||g||_\infty/\sqrt{n}}\right).$$

We use $g(\cdot) = \mathbf{1}\{\cdot > e^k\}$ and $t = 2\sqrt{p_k\log(2/\delta)}$ in the above inequality. Using the fact that $t = 2\sqrt{p_k \log(2/\delta)} \leq \sqrt{n}p_k$ by the assumption of $p_k \geq (4\log(2/\delta))/n$, we have

$$\begin{aligned}
\mathbb{P}\left(\sqrt{n}|\hat{p}_k - p_k| > t\right) &\leq 2\exp\left(-\frac{1}{4}\frac{t^2}{p_k + t/\sqrt{n}}\right)\\
&\leq 2\max\left[\exp\left(-\frac{1}{4}\frac{t^2}{p_k}\right), \exp\left(-\frac{1}{4}\sqrt{n}t\right)\right]\\
&\leq 2\exp\left(-\frac{1}{4}\frac{t^2}{p_k}\right)\\
&= \delta,
\end{aligned}$$

where the last equality follows by definition of t. $\square$

S2. Proof of Lemma 1

A. Since $p_k \geq 16\log(2/\delta)/n$, we can use Lemma 1. Rewriting the inequality (S1.1), we have with probability larger than $1-\delta$

$$\log\left(1-2\sqrt{\frac{\log(2/\delta)}{np_k}}\right) \leq \log(\hat{p}_k) - \log(p_k) \leq \log\left(1+2\sqrt{\frac{\log(2/\delta)}{np_k}}\right).$$

Then using the simple inequalities $\log(1+u) \leq u$, and $\log(1-u) \geq (-3u)/2$ for $u < 1/2$,

$$\log(p_k) - 3\sqrt{\frac{\log(2/\delta)}{np_k}} \leq \log(\hat{p}_k) \leq \log(p_k) + 2\sqrt{\frac{\log(2/\delta)}{np_k}}.$$

By using a similar inequality for $\log(\hat{p}_{k+1})$, with probability larger than $1-2\delta$,

$$\begin{aligned}\big|\hat{\alpha}(k) - (\log(p_k) - \log(p_{k+1}))\big| &\leq 3\sqrt{\frac{\log(2/\delta)}{np_k}} + 3\sqrt{\frac{\log(2/\delta)}{np_{k+1}}} \\ &\leq 6\sqrt{\frac{\log(2/\delta)}{np_{k+1}}}. \qquad \text{(S2.1)}\end{aligned}$$

B. By definition of second-order Pareto distributions, we have $\big|p_k - Ce^{-k\alpha}\big| \leq C'e^{-k\alpha(1+\beta)}$, or equivalently,

$$\left|\frac{e^{k\alpha}p_k}{C} - 1\right| \leq \frac{C'}{C}e^{-k\alpha\beta}.$$

Since we assume $\frac{C'}{C}e^{-k\alpha\beta} \leq 1/2$, we have

$$\big|\log(p_k) - \log(C) + k\alpha\big| \leq \frac{3C'}{2C}e^{-k\alpha\beta}.$$

A similar result also holds for p_{k+1}, and thus

$$\big|\log(p_k) - \log(p_{k+1}) - \alpha\big| \leq \frac{3C'}{C}e^{-k\alpha\beta}. \qquad \text{(S2.2)}$$

Combining Equations (S2.1) and (S2.2), we obtain the large deviation inequality (3.3)*[1]. Now, using the property of the second-order Pareto distributions, we can bound p_{k+1} from below.

$$\begin{aligned}p_{k+1} &\geq Ce^{-(k+1)\alpha}\left(1 - \frac{C'}{C}e^{-(k+1)\alpha\beta}\right) \\ &\geq \frac{C}{2}e^{-(k+1)\alpha} \geq Ce^{-(k+1)\alpha-1},\end{aligned}$$

[1] *All the equation numbers without "S" are from the main text.

where the second inequality comes from the assumption that $e^{-k\alpha\beta} \leq C/(2C')$. By substituting this into the inequality (3.3), the final inequality (3.4) follows.

S3. Proof of Theorem 1

The proof consists of the two steps—bounding the bias, and bounding the deviations of the estimate—as in the proof of the Lemma 1.B.

First, we bound the bias (more precisely, a proxy for the bias) using the property of the distribution class $\mathcal{A}$. By definition, we know that for any ϵ such that $C/2 > \epsilon > 0$, there exists a constant $B > 0$ such that for $x > B$,

$$\left|1 - F(x) - Cx^{-\alpha}\right| \leq \epsilon x^{-\alpha}.$$

Since $k_n \to \infty$ as $n \to \infty$, for any n larger than some large enough N_1 (i.e. such that $\forall n \geq N_1$, $e^{k_n} > B$), we have

$$\left|p_{k_n} - Ce^{-k_n\alpha}\right| \leq \epsilon e^{-k_n\alpha}, \tag{S3.1}$$

which yields since $\epsilon < C/2$, $\left|\log(p_{k_n}) - \log(C) + k_n\alpha\right| \leq \frac{3\epsilon}{2C}$ using the same technique as for the proof of Lemma 1. This holds also for $k_n + 1$ and thus

$$\left|\log(p_{k_n}) - \log(p_{k_n+1}) - \alpha\right| \leq \frac{3\epsilon}{C}. \tag{S3.2}$$

Note also that Equation (S3.1) can be used to bound the p_{k_n+1} below as follows.

$$p_{k_n+1} \geq (C - \epsilon)e^{-(k_n+1)\alpha} \geq \frac{C}{e^{\alpha+1}}e^{-k_n\alpha}. \tag{S3.3}$$

Since $(\log(n)e^{k_n\alpha})/n \to 0$ as $n \to \infty$, we know that there exists N_2 large enough, such that for any $n \geq N_2$, $p_{k_n+1} \geq 32\log(n)/n$.

Then we can bound the proxy for the standard deviation using the result (3.2) in Lemma 1.A. For $n \geq \max(N_1, N_2)$, combining Equation (S3.2) and Equation (3.2) with $\delta = 2/n^2$, we have with probability larger than $1 - 4/n^2$,

$$\left|\hat{\alpha}(k_n) - \alpha\right| \leq 6\sqrt{\frac{\log(n^2)}{np_{k_n+1}}} + \frac{3\epsilon}{C}.$$

Then we bound the first term in the right side of the above inequality using (S3.3). That is,

$$6\sqrt{\frac{\log(n^2)}{np_{k_n+1}}} \leq 6\sqrt{e^{\alpha+1}\frac{\log(n^2)}{Cne^{-k_n\alpha}}} \leq \frac{6e^{(\alpha/2)+1}}{\sqrt{C}}\sqrt{\frac{\log(n)e^{k_n\alpha}}{n}}$$

By the assumption that $(\log(n)e^{k_n\alpha})/n \to 0$, and since the above inequality holds for any $\epsilon > 0$, we conclude that α_n converges in probability to α. Moreover, since $\sum_n(4/n^2) < \infty$, Borel–Cantelli Lemma says that $\hat{\alpha}(k_n)$ converges to α almost surely.

S4. Proof of Theorem 2

Let n satisfy the following,

$$n > \max\Big((\frac{2C'}{C})^{\frac{2\beta+1}{\beta}}, (\frac{32\log(2/\delta)e^{2\alpha}}{C})^{\frac{2\beta+1}{2\beta}} \Big). \tag{S4.1}$$

We let $k^* = k^*_n$ such that $k^*_n := \Big\lfloor \log(n^{\frac{1}{\alpha(2\beta+1)}}) + 1 \Big\rfloor$. Note that for n larger than $(2C'/C)^{\frac{2\beta+1}{\beta}}$, we have $e^{-k^*\alpha\beta} \leq C/(2C')$. This implies, together with the second-order Pareto assumption,

$$p_{k^*+1} \geq \frac{C}{2} n^{-\frac{1}{2\beta+1}} e^{-2\alpha} \geq \frac{16\log(2/\delta)}{n}$$

where the last inequality follows by assuming $n \geq (\frac{32\log(2/\delta)e^{2\alpha}}{C})^{\frac{2\beta+1}{2\beta}}$.

By (3.4) and by the choice of k_n, we have with probability larger than $1-2\delta$,

$$\big|\hat{\alpha}(k^*) - \alpha\big| \leq \left(6\sqrt{e^{2\alpha+1}\frac{\log(2/\delta)}{C}} + \frac{3C'}{C} \right) n^{-\frac{\beta}{2\beta+1}}.$$

S5. Proof of Theorem 3

The following lemma is a useful preliminary result for the proof of Theorem 3.

Lemma 2. *We define K such that $p_K \geq \frac{16\log(2/\delta)}{n}$ and also $p_{K+1} < \frac{16\log(2/\delta)}{n}$. Then for any $k \geq K+1$, with probability larger than $1-\delta$,*

$$\hat{p}_k \leq \frac{24\log(2/\delta)}{n}. \tag{S5.1}$$

Proof of Lemma 2. We let $q := 16\log(2/\delta)/n$ and define a Bernoulli random variable $Y_i(q)$ (independent from $X_1, \ldots, X_n$) where $P(Y_i(q) = 1) = q$ for $i = 1, \ldots, n$. Then we compare $m_q := \frac{1}{n}\sum_{i=1}^n Y_i(q)$ and $\hat{p}_{K+1} = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_i > e^{K+1}\}$. Since $q > p_{K+1}$, the distribution of $\hat{p}_{K+1}$ is stochastically dominated by the distribution of m_q (that is, $P(\hat{p}_{K+1} > t) \leq P(m_q > t)$). By Lemma 1, we have with probability larger than $1-\delta$,

$$|m_q - q| \leq 2\sqrt{\frac{q\log(2/\delta)}{n}} = \frac{8\log(2/\delta)}{n}.$$

Then by stochastic dominance, with probability $1-\delta$,

$$\hat{p}_{K+1} \leq q + 2\sqrt{\frac{q\log(2/\delta)}{n}} = \frac{24\log(2/\delta)}{n}.$$

Thus, for any $k \geq K+1$ using the monotonicity of $\hat{p}_k$ (that is, $\hat{p}_k \geq \hat{p}_{k+1}$), we obtain that (S5.1) holds with probability larger than $1-\delta$ as required. □

The proof is based on 5 steps. We first define an event ξ in (S5.4) of high probability where the deviation of empirical probabilities $\hat{p}_k$ from p_k is well upper bounded (with the same bound in the large deviation inequality in (S1.1) but without a probability statement) for a given subset of indices $k \leq K$, where K is of order of $\log n$. Then we define $\bar{k}$ which is slightly smaller than the oracle k^* and also $\bar{k} \leq K$ so that on ξ the deviation of $\hat{\alpha}(\bar{k})$ from α (i.e. $|\hat{\alpha}(\bar{k}) - \alpha|$) is upper bounded as in (S5.7). In the third step, we show that $\hat{p}_{\bar{k}+1} > 24\log(2/\delta)/n$ on ξ so that $\bar{k}$ is one possible index for $\hat{k}_n$. Also we prove that $\hat{k}_n \leq \bar{k}$ in Step 4 which leads us to bound $|\hat{\alpha}(\bar{k}) - \hat{\alpha}(\hat{k}_n)|$ from above on ξ using the definition of $\hat{k}_n$. This combined with the second step finally gives an upper bound of $|\hat{\alpha}(\hat{k}_n) - \alpha|$ on ξ. More precisely, we prove that on the set ξ, we have $|\hat{\alpha}(\hat{k}_n) - \alpha| \leq (B_2 + \frac{3C'}{C})(\frac{n}{\log(2/\delta)})^{-\beta/(2\beta+1)}$ where B_2 is a constant which will be defined in the last stage of the proof. Then we can bound $\mathbb{P}(|\hat{\alpha}(\hat{k}_n) - \alpha| \geq (B_2 + \frac{3C'}{C})(\frac{n}{\log(2/\delta)})^{-\beta/(2\beta+1)}) \leq \mathbb{P}(\xi^c)$ which has a small probability.

Let $F \in \mathcal{S}(\alpha, \beta, C, C')$ and $1/4 > \delta > 0$. Also we let n satisfy the following,

$$n > \log\big(\frac{2}{\delta}\big) \max\left[32\Big(\frac{2C'}{C^{1+\beta}}\Big)^{1/\beta}, \Big(\frac{32e^{2\alpha}}{C}\Big)^{\frac{2\beta+1}{2\beta}}, \Big(\frac{2C'}{C}\Big)^{\frac{2\beta+1}{\beta}}, \Big(\frac{96e^{2\alpha}}{C}\Big)^{\frac{2\beta+1}{\beta}}\right]. \tag{S5.2}$$

Step 1: Definition of an event of high probability

First, we define $K \in \mathbb{N}$ such that $p_K \geq \frac{16\log(2/\delta)}{n} > p_{K+1}$. By inverting the condition for the second-order Pareto distributions, $\frac{16\log(2/\delta)}{n} \leq p_K \leq (C + C')e^{-K\alpha}$ gives $K \leq \frac{1}{\alpha}\log\left(\frac{(C+C')n}{16\log(2/\delta)}\right)$. Set $u = \frac{1}{\alpha}\log\left(\frac{Cn}{32\log(2/\delta)}\right) - 1$. Then since $n > 32(\frac{2C'}{C^{1+\beta}})^{1/\beta}\log(2/\delta)$, we know by definition of $\mathcal{S}$ that $1 - F(e^{u+1}) > \frac{16\log(2/\delta)}{n}$. Using the fact that $1 - F(e^x)$ is a decreasing function of x and $\frac{16\log(2/\delta)}{n} > p_{K+1}$, we have $u < K$. Thus we obtain the range of K by

$$\frac{1}{\alpha}\log\left(\frac{Cn}{32\log(2/\delta)}\right) - 1 < K \leq \frac{1}{\alpha}\log\left(\frac{(C+C')n}{16\log(2/\delta)}\right). \tag{S5.3}$$

We define the following event

$$\xi = \Big\{\omega : \forall k \leq K, \big|\hat{p}_k(\omega) - p_k\big| \leq 2\sqrt{\frac{p_k\log(2/\delta)}{n}}, \hat{p}_{K+1}(\omega) \leq \frac{24\log(2/\delta)}{n}\Big\}. \tag{S5.4}$$

By definition, we have $p_K \geq \frac{16\log(2/\delta)}{n}$, which gives the Bernstein inequality (S1.1) with probability $1-\delta$ for $k \leq K$. In addition, Lemma 2 gives (S5.1) with probabiltiy $1-\delta$. Thus, an union bound implies that $\mathbb{P}(\xi) \geq 1 - (K+1)\delta$. By monotonicity of $\hat{p}_k$, we have on the event ξ, for any $k \geq K+1$, $\hat{p}_k \leq \frac{24\log(2/\delta)}{n}$. This implies that on the event ξ, the k, k' considered in Equation (3.5) are smaller than K and in particular, we have $\hat{k}_n \leq K$.

Step 2: Bounding the deviation of $\hat{\alpha}(k)$ from α on ξ (where $k \leq K$)

We define $\bar{k}_n = \bar{k} \in \mathbb{N}$ such that

$$\bar{k} := \left\lfloor \log\left(\left(\frac{n}{\log(2/\delta)}\right)^{\frac{1}{\alpha(2\beta+1)}} \right) + 1 \right\rfloor.$$

By definition of $\bar{k}$, we know that $\bar{k} < K$. Indeed, by assuming $n \geq (32\frac{e^{2\alpha}}{C})^{\frac{2\beta+1}{2\beta}} \log(2/\delta)$ and by (S5.3),

$$\bar{k} \leq \log\left(\left(\frac{n}{\log(2/\delta)}\right)^{\frac{1}{\alpha(2\beta+1)}} \right) + 1 \leq \frac{1}{\alpha} \log\left(\frac{Cn}{32\log(2/\delta)} \right) - 1 < K.$$

Thus,

$$e^{-K\alpha\beta} \leq e^{-\bar{k}\alpha\beta} \leq C/(2C'), \tag{S5.5}$$

where the second inequality follows since $n > \log(2/\delta)(\frac{2C'}{C})^{\frac{2\beta+1}{\beta}}$.

Note also that $\bar{k} \leq k^*$, where $k^* := \left\lfloor \log\left(n^{\frac{1}{\alpha(2\beta+1)}} \right) + 1 \right\rfloor$ as before.

If $k < K$ satisfies $e^{-k\alpha\beta} \leq C/(2C')$, then since $p_{k+1} \geq p_K \geq (16\log(2/\delta))/n$, then using the exactly same proof as for Lemma 1.B, we have on ξ that

$$|\hat{\alpha}(k) - \alpha| \leq 6\sqrt{\frac{e^{(k+1)\alpha+1}\log(2/\delta)}{Cn}} + \frac{3C'}{C} e^{-k\alpha\beta}. \tag{S5.6}$$

Since $e^{-\bar{k}\alpha\beta} \leq C/(2C')$ by (S5.5) and $\bar{k} < K$, Equation (S5.6) is verified for $\bar{k}$ on ξ. Then by definition of $\bar{k}$ in Equation (S5.6), we have on ξ that

$$|\hat{\alpha}(\bar{k}) - \alpha| \leq \left(6\sqrt{\frac{e^{2\alpha+1}}{C}} + \frac{3C'}{C} \right) \left(\frac{n}{\log(2/\delta)} \right)^{-\frac{\beta}{2\beta+1}}. \tag{S5.7}$$

Step 3: Proof that $\hat{p}_{\bar{k}+1} > \frac{24\log(2/\delta)}{n}$ on ξ

By definition, we have on ξ, using $\bar{k} \leq K-1$ and $p_{\bar{k}+1} \geq p_K \geq (16\log(2/\delta))/n$,

$$\hat{p}_{\bar{k}+1} \geq p_{\bar{k}+1} \left(1 - 2\sqrt{\frac{\log(2/\delta)}{np_{\bar{k}+1}}} \right) \geq \frac{p_{\bar{k}+1}}{2}.$$

Then using the second order Pareto property with $(C'/C)e^{-\bar{k}\alpha\beta} \leq 1/2$, we have $p_{\bar{k}+1} \geq (Ce^{-(\bar{k}+1)\alpha})/2$, which gives

$$\hat{p}_{\bar{k}+1} \geq \frac{Ce^{-(\bar{k}+1)\alpha}}{4} \geq \frac{Ce^{-2\alpha}}{4} \left(\frac{\log(2/\delta)}{n} \right)^{1/(2\beta+1)}, \tag{S5.8}$$

where the second inequality follows from $n > \log(2/\delta)(\frac{2C'}{C})^{\frac{2\beta+1}{\beta}}$ and from the definition of $\bar{k}$. Since $n > \left(\frac{96e^{2\alpha}}{C}\right)^{\frac{2\beta+1}{\beta}} \log(2/\delta)$, we have shown that $\hat{p}_{\bar{k}+1}$ is larger than $\frac{24\log(2/\delta)}{n}$ on ξ.

Step 4: Proof that $\hat{k}_n \le \bar{k}$ on ξ

Suppose that $\hat{k}_n > \bar{k}$. Then by definition of $\hat{k}_n$ in (3.5), on ξ, there exists $k > \bar{k}$ such that $\hat{p}_{k+1} > \frac{24\log(2/\delta)}{n}$ (this imposes $k < K$ on ξ) and

$$|\hat{\alpha}(k) - \hat{\alpha}(\bar{k})| > A(\delta)\sqrt{\frac{1}{n\hat{p}_{k+1}}} \ge \frac{A(\delta)}{\sqrt{2(C+C')}}\sqrt{\frac{e^{k\alpha}}{n}}, \tag{S5.9}$$

where the second inequality in the above follows by bounding $\hat{p}_{k+1}$ above by definition of ξ,

$$\hat{p}_{k+1} \le p_{k+1}\left(1 + 2\sqrt{\frac{\log(2/\delta)}{np_{k+1}}}\right) \le \frac{3}{2}p_{k+1} \le 2(C+C')e^{-k\alpha},$$

where the penultimate inequality is obtained by $p_k \ge p_K \ge 16\log(2/\delta)/n$ (since $k \le K$), and the last inequality follows by definition of the second order Pareto condition.

Since $k \ge \bar{k}+1$, we bound $e^{-k\alpha\beta} \le e^{-\bar{k}\alpha\beta} \le C/(2C')$ by (S5.5). Also we have $p_{k+1} \ge \frac{16\log(2/\delta)}{n}$, since $p_{k+1} \ge p_K$. Equation (S5.6) is thus verified on ξ for such $k > \bar{k}$. Now using $\sqrt{\frac{e^{k\alpha}\log(2/\delta)}{n}} > e^{-k\alpha\beta}$ (since $k > \bar{k}$), we have

$$|\hat{\alpha}(k) - \alpha| \le \Big(6\sqrt{\frac{e^{\alpha+1}}{C}} + \frac{3C'}{C}\Big)\sqrt{\frac{e^{k\alpha}\log(2/\delta)}{n}}. \tag{S5.10}$$

Equations (S5.9) and (S5.10) imply that on ξ,

$$\begin{aligned}
|\hat{\alpha}(\bar{k}) - \alpha| &> \Big(\frac{A(\delta)}{\sqrt{2(C+C')}} - \sqrt{\log(2/\delta)}\big(6\sqrt{\frac{e^{\alpha+1}}{C}} + \frac{3C'}{C}\big)\Big)\sqrt{\frac{e^{k\alpha}}{n}} \\
&\ge \big(6\sqrt{\frac{e^{2\alpha+1}}{C}} + \frac{3C'}{C}\big)\left(\frac{n}{\log(2/\delta)}\right)^{-\frac{\beta}{2\beta+1}},
\end{aligned}$$

since we assume that $\frac{A(\delta)}{\sqrt{2(C+C')}} \ge 2\sqrt{\log(2/\delta)}\big(6\sqrt{\frac{e^{2\alpha+1}}{C}} + \frac{3C'}{C}\big)$. This contradicts Equation (S5.7), and this means that on ξ, $\hat{k}_n \le \bar{k}$.

Step 5: Large deviation inequality for an adaptive estimator

We have $\hat{p}_{\bar{k}+1} \ge \frac{24\log(2/\delta)}{n}$ from Step 3, and $\hat{k}_n \le \bar{k}$ from Step 4 on ξ. Thus by definition of $\hat{k}_n$ in (3.5), we have on ξ that

$$\begin{aligned}
|\hat{\alpha}(\bar{k}) - \hat{\alpha}(\hat{k}_n)| &\le A(\delta)\sqrt{\frac{1}{n\hat{p}_{\bar{k}+1}}} \\
&\le 2A(\delta)\sqrt{\frac{e^{2\alpha}}{C}}\left(\log\left(\frac{2}{\delta}\right)\right)^{-\frac{1}{2(2\beta+1)}} n^{-\frac{\beta}{2\beta+1}} \\
&= 2A(\delta)\sqrt{\frac{e^{2\alpha}}{C\log(2/\delta)}}\left(\frac{n}{\log(2/\delta)}\right)^{-\frac{\beta}{2\beta+1}},
\end{aligned} \tag{S5.11}$$

where the second inequality follows on ξ by Equation (S5.8).

Hence, Equations (S5.11) and (S5.7) imply that on ξ

$$|\hat{\alpha}(\hat{k}_n) - \alpha| \leq \left((6\sqrt{\frac{e^{2\alpha+1}}{C}} + \frac{3C'}{C}) + 2A(\delta)\sqrt{\frac{e^{2\alpha}}{C\log(2/\delta)}} \right) \left(\frac{n}{\log(2/\delta)} \right)^{-\frac{\beta}{2\beta+1}}.$$

Denote $B_1 = 6\sqrt{\frac{e^{2\alpha+1}}{C}\log(2/\delta)}$ and $B_2 = \left(B_1 + 2A(\delta)\sqrt{\frac{e^{2\alpha}}{C}}\right)\frac{1}{\sqrt{\log(2/\delta)}}$. Then since $\mathbb{P}(\xi) \geq 1 - (K+1)\delta$, we have shown that

$$\begin{aligned} \sup_{F\in\mathcal{S}} \mathbb{P}_F \left(|\hat{\alpha}(\hat{k}_n) - \alpha| \geq \left(B_2 + \frac{3C'}{C}\right) \left(\frac{n}{\log(2/\delta)} \right)^{-\frac{\beta}{2\beta+1}} \right) \\ \leq (K+1)\delta \leq \left(\frac{1}{\alpha} \log\left(\frac{(C+C')n}{16} \right) + 1 \right) \delta \end{aligned}$$

where the last inequality follows by (S5.3). This concludes the proof.

S6. Proof of Corollary 1

Set

$$\begin{aligned} \epsilon &= \left(1 + \frac{1}{\alpha_1}\log\left((C_2 + C')n\right)\right)\delta, \\ A(\epsilon) &= 6\sqrt{2(C_2+C')}\left(\sqrt{\log\left(\frac{2}{\epsilon}\left(1 + \frac{\log((C_2+C')n)}{\alpha_1}\right)\right)}\left(2\sqrt{\frac{e^{2\alpha_2+1}}{C_1}} + \frac{C'}{C_1}\right)\right), \end{aligned} \tag{S6.1}$$

and plug $\delta = \delta(\epsilon) = \epsilon/(1 + \log((C_2+C')n))/\alpha_1)$ and $A(\epsilon) := A(\delta(\epsilon))$ in the adaptive method described in Theorem 3. Set

$$B_3 := 6\sqrt{\frac{e^{2\alpha_2+1}}{C_1}} + 24\frac{e^{2\alpha_2}}{C_1}\sqrt{2e(C_2+C')} + 12e^{\alpha_2}\frac{C'}{C_1}\sqrt{2\frac{(C_2+C')}{C_1}} + \frac{3C'}{C_1}. \tag{S6.2}$$

It holds for any $\alpha \in [\alpha_1, \alpha_2]$, $C \in [C_1, C_2]$ and $\beta > \beta_1$ that the constant in Theorem 3 can be bounded as

$$\begin{aligned} B_2 + \frac{3C'}{C} &= 6\sqrt{\frac{e^{2\alpha+1}}{C}} + 12\sqrt{2\frac{e^{2\alpha}}{C}(C_2+C')}\left(2\sqrt{\frac{e^{2\alpha_2+1}}{C_1}} + \frac{C'}{C_1}\right) + \frac{3C'}{C} \\ &\leq B_3, \end{aligned}$$

so B_3 is a uniform bound on the constant in Theorem 3 for all considered values of α, C, β. Also, the uniform condition for the sample size is derived from Equation (S5.2) by

$$n > \log\left(\frac{2}{\epsilon}\Big(1+\frac{\log((C_2+C')n)}{\alpha_1}\Big)\right)$$
$$\times \max\left[32\Big(\frac{2\bar{C}'}{\bar{C}_1^{1+\beta_1}}\Big)^{\frac{1}{\beta_1}}, \Big(\frac{2\bar{C}'}{\bar{C}_1}\Big)^{2+\frac{1}{\beta_1}}, \Big(\frac{32e^{2\alpha_2}}{\bar{C}_1}\Big)^{1+\frac{1}{2\beta_1}}, \Big(\frac{96e^{2\alpha_2}}{\bar{C}_1}\Big)^{2+\frac{1}{\beta_1}}\right], \quad \text{(S6.3)}$$

where $\bar{C}_1 = \min(1, C_1)$ and $\bar{C}' = \max(1, C')$.

S7. Proof of Theorem 4

We prove the minimax lower bound by Fano's method (see e.g. Section 2.7 in Tsybakov (2008)). We define a set of approximately $\log(n)$ functions F_i whose first and second order parameters are respectively α_i and β_i. Until a point K_i, each distribution F_i matches a Pareto distribution with the first order parameter α, which is the same for all of the F_i. After this point K_i, F_i is Pareto with parameter α_i. These functions satisfy several specific properties summarized in Lemma 3. For instance, they are such that the for any $i \neq j$, the distance between α_i and α_j is at least of order $(\frac{n}{\log\log(n)})^{-\frac{\beta_i}{2\beta_i+1}}$. Moreover, the Kullback Leibler (KL) divergence between F_i and F_j is small enough so that F_i and F_j cannot be distinguishable as n increases. These two main properties enable us to apply Fano's lemma, which results in the lower bound of Theorem 4. For the proof, we assume that n is sufficiently large.

Step 1: Construction of a finite set of distributions

Let $\alpha > 0$ and $\beta > 1$. Let $\upsilon := \min\big(1, \frac{\alpha^2}{8\exp(\frac{1}{\alpha(2\beta-1)})}\big)$. Let $M > 1$ be an integer such that

$$\lfloor \log(n/\log(M)) \rfloor + 1 = M,$$

which implies that $\log(n)/2 < M < 2\log(n)$ for large n. Set for any integer $1 \leq i \leq M$

$$\begin{aligned}
\beta_i &= \beta - \frac{i}{M}\\
\gamma_i &= \frac{\beta_i}{2\beta_i+1}\Big(1+\frac{\log(\upsilon)}{\log\log M}\Big)\\
K_i &= n^{\frac{1}{\alpha(2\beta_i+1)}}\big(\log M\big)^{-\frac{\gamma_i}{\alpha\beta_i}} = \Big(\frac{n}{\upsilon\log(M)}\Big)^{\frac{1}{\alpha(2\beta_i+1)}} \qquad \text{(S7.1)}\\
t_i &= K_i^{-\alpha\beta_i} = n^{-\frac{\beta_i}{2\beta_i+1}}\big(\log M\big)^{\gamma_i} = \Big(\frac{n}{\upsilon\log(M)}\Big)^{-\frac{\beta_i}{2\beta_i+1}}\\
\alpha_i &= \alpha - t_i = \alpha - n^{-\beta_i/(2\beta_i+1)}(\log(M))^{\gamma_i}.
\end{aligned}$$

By definition, for $i < j$, we have $\beta_i > \beta_j$, $\gamma_i > \gamma_j$, $K_i < K_j$, $t_i < t_j$ and $\alpha_i > \alpha_j$. By assuming n large enough, we suppose that $\gamma_i > 0$ for all $i = 1, \ldots, M$, and

$\frac{\min(\alpha,1/\alpha)}{2} n^{\frac{\beta_i}{2\beta_i+1}} > M^{\frac{\beta_i}{2\beta_i+1}+1}$. Also we have $\beta_i \geq \beta - 1$, $K_i > 1$, and $\alpha - t_i \geq \alpha/2 =: \alpha_1$ for large enough n.

Using these notation, we introduce the distribution

$$1 - F_0(x) = x^{-\alpha}, \tag{S7.2}$$

and for any integer $1 \leq i \leq M$, we introduce *perturbed versions* of the distribution F_0

$$1 - F_i(x) = x^{-\alpha}\mathbf{1}\{1 \leq x \leq K_i\} + K_i^{-t_i} x^{-\alpha+t_i}\mathbf{1}\{x > K_i\}. \tag{S7.3}$$

We write $\{f_0, f_1, \ldots, f_M\}$ for the densities associated with distributions $\{F_0, F_1, \ldots, F_M\}$ with respect to Lebesgue measure.

Step 2: Properties of the constructed distributions

The following lemma highlights important characteristics of distributions $\{F_i, i = 1, \ldots, M\}$ and their parameters corresponding to the second order Pareto distributions.

Lemma 3. *Let $1 \leq i \leq M$ and $1 \leq j \leq M$. It holds that for F_i defined as (S7.3) and using notation in (S7.1),*

$$F_i \in \mathcal{S}\left(\alpha - t_i, \beta_i, K_i^{-t_i}, \frac{1}{\alpha(\beta-1)}\right). \tag{S7.4}$$

Moreover

$$\exp\big(-\frac{1}{\alpha(2\beta-1)}\big) \leq K_i^{-t_j} \leq 1, \tag{S7.5}$$

and if $i \neq j$,

$$|\alpha_i - \alpha_j| \geq c(\beta)\max(t_i, t_j), \tag{S7.6}$$

where $c(\beta) := 1 - \exp\left(-\frac{1}{2(2\beta+1)^2}\right)$.

Step 3: Computation of the Kullback-Leibler (KL) divergence

In this step, we first compute the KL divergence between F_0 and F_i, which is defined as $KL(F_0, F_i) = \int f_0(x) \log \frac{f_0(x)}{f_i(x)} dx$. Then we prove that it has the same order of the KL divergence between F_i and F_0. Second, we prove that the KL divergence between F_i and F_j is at most of the same order of $\max\{KL(F_0, F_i), KL(F_j, F_0)\}$.

Lemma 4 provides the order of the KL divergence between F_i and F_0.

Lemma 4. *Let $1 \leq i \leq M$. It holds that for F_0 in (S7.2), F_i in (S7.3) and using notation in (S7.1),*

$$\max\big(KL(F_0, F_i), KL(F_i, F_0)\big) \leq \frac{2t_i^2 K_i^{-\alpha}}{\alpha^2}.$$

Using Lemma 4, we obtain bounds on the KL divergence between F_i and F_j in the following lemma.

Lemma 5. *Let $(i, j) \in \{1, \ldots, M\}^2$. It holds that for F_i in (S7.3) and using notation in (S7.1),*

$$KL(F_i, F_j) \leq \frac{2\exp(\frac{1}{\alpha(2\beta-1)})}{\alpha^2}\left(t_i^2 K_i^{-\alpha} + t_j^2 K_j^{-\alpha}\right). \tag{S7.7}$$

Step 4: Use of Fano's method.

Here we follow ideas in the Fano's method using the above results in Step 1-3. Let $\hat{\alpha} = \hat{\alpha}(X_1, \ldots, X_n) =: \hat{\alpha}(X)$ be an estimator of α. Then we define the following discrete random variable

$$Z = Z(X) := \arg \min_{j \in \{1,\ldots,M\}} |\hat{\alpha}(X) - \alpha_j|,$$

which implies that $|\hat{\alpha} - \alpha_j| > c(\beta)t_j/2$ if $Z \neq j$ by Equation (S7.6). Also we consider another random variable Y, uniformly distributed on $\{1, \ldots, M\}$ where $X|Y = j \sim F_j^n$. By bounding the maximum by the average,

$$\begin{aligned}
\max_{j \in \{1,\ldots,M\}} \mathbb{P}_{F_j}\left(|\hat{\alpha} - \alpha_j| \geq \frac{c(\beta)t_j}{2}\right) &\geq \frac{1}{M}\sum_{j=1}^{M} \mathbb{P}\left(Z \neq j | Y = j\right) \\
&= \mathbb{P}(Z \neq Y) \\
&\geq 1 - \frac{1}{\log M}\left(\frac{1}{M^2}\sum_{j,j'} KL(F_j^n, F_{j'}^n) + \log 2\right),
\end{aligned}$$

where the last inequality is obtained by Fano's inequality (see Section 2.1 in Cover and Thomas (2012), or see Appendix in Subsection S9 for a proof of how this inequality is derived).

Using the fact that $KL(F_1^n, F_2^n) = nKL(F_1, F_2)$, and by Equation (S7.7),

$$\begin{aligned}
\frac{1}{M^2}\sum_{j,j'} KL(F_j^n, F_{j'}^n) &\leq \frac{n}{M^2}\frac{2\exp(\frac{1}{\alpha(2\beta-1)})}{\alpha^2}\sum_{j,j'}\left(t_j^2 K_j^{-\alpha} + t_{j'}^2 K_{j'}^{-\alpha}\right) = \frac{n}{M}\frac{4\exp(\frac{1}{\alpha(2\beta-1)})}{\alpha^2}\sum_j t_j^2 K_j^{-\alpha} \\
&= \frac{n}{M}\frac{4\exp(\frac{1}{\alpha(2\beta-1)})}{\alpha^2}\sum_j \frac{\upsilon\log(M)}{n} \\
&= \frac{4\exp(\frac{1}{\alpha(2\beta-1)})}{\alpha^2}(\log(M)) \times \upsilon \leq \frac{1}{2}\log(M).
\end{aligned}$$

where the second equality follows by $t_j^2 K_j^{-\alpha} = K_j^{-\alpha(2\beta_j+1)} = \frac{\upsilon\log(M)}{n}$ and the last inequality is by $\upsilon \leq \frac{\alpha^2}{8\exp(\frac{1}{\alpha(2\beta-1)})}$. Hence, for a sufficiently large n, we have

$$\max_{j \in \{1,\ldots,M\}} \mathbb{P}_{F_j}\left(|\hat{\alpha} - \alpha_j| \geq \frac{c(\beta)t_j}{2}\right) \geq \frac{1}{4}.$$

More specifically, using $c(\beta) := 1 - \exp(-\frac{1}{2(2\beta+1)^2}) \geq \frac{1}{2(2\beta+1)^2}$ and since $t_j = \left(\frac{\upsilon\log(M)}{n}\right)^{\frac{\beta_j}{2\beta_j+1}} \geq$

$v^{\frac{\beta_j}{2\beta_j+1}}\Big(\frac{\log\big((\log(n))/2\big)}{n}\Big)^{\frac{\beta_j}{2\beta_j+1}}$, we have

$$\max_{j\in\{1,\dots,M\}} \mathbb{P}_{F_j}\left(|\hat{\alpha}-\alpha_j| \geq B(\alpha,\beta,\beta_j)\Big(\frac{\log\big((\log(n))/2\big)}{n}\Big)^{\frac{\beta_j}{2\beta_j+1}}\right) \geq \frac{1}{4},$$

where

$$B(\alpha,\beta,\beta_j) := \frac{1}{4(2\beta+1)^2}\min\left[1, \Big(\frac{\alpha^2}{8\exp(\frac{1}{\alpha(2\beta-1)})}\Big)^{\frac{\beta_j}{2\beta_j+1}}\right]. \tag{S7.8}$$

By definition of $\{F_1,\dots,F_M\}$, we have (by Lemma 3)

$$\{F_1,\dots,F_M\} \subset \Big\{F \in \mathcal{S}(\alpha^*,\beta^*,C,\tilde{C}') : \alpha^* \in [\alpha/2,\alpha], \beta^* \in [\beta-1,\beta], C \in [\tilde{C}_1,\tilde{C}_2]\Big\},$$

where $\tilde{C}_1(\alpha,\beta) := \exp\Big(-\frac{1}{\alpha(2\beta-1)}\Big)$, $\tilde{C}_2 := 1$, and $\tilde{C}'(\alpha,\beta) = \frac{1}{\alpha(\beta-1)}$.

Then by bounding the supremum by the maximum over the finite subset, we finally provide the following lower bound result.

$$\begin{aligned}
&\sup_{\substack{\alpha^*\in[\alpha/2,\alpha],\beta^*\in[\beta-1,\beta]\\ C\in[\tilde{C}_1,\tilde{C}_2]}} \sup_{F\in\mathcal{S}(\alpha^*,\beta^*,C,\tilde{C}')} \mathbb{P}_F\left(|\hat{\alpha}-\alpha^*| \geq B(\alpha,\beta,\beta^*)\Big(\frac{\log\big((\log(n))/2\big)}{n}\Big)^{\frac{\beta^*}{2\beta^*+1}}\right)\\
&\qquad \geq \max_{j\in\{1,\dots,M\}} \mathbb{P}_{F_j}\left(|\hat{\alpha}-\alpha_j| \geq B(\alpha,\beta,\beta_j)\Big(\frac{\log\big((\log(n))/2\big)}{n}\Big)^{\frac{\beta_j}{2\beta_j+1}}\right)\\
&\qquad \geq \frac{1}{4}.
\end{aligned}$$

By changing parametrization and setting $\alpha_1 = \alpha/2$ and $\beta_1 = \beta - 1$, we proved that

$$\sup_{\substack{\alpha^*\in[\alpha_1,2\alpha_1],\beta^*\in[\beta_1,\infty)\\ C\in[C_1,C_2]}} \sup_{F\in\mathcal{S}(\alpha^*,\beta^*,C,C')} \mathbb{P}_F\left(|\hat{\alpha}-\alpha^*| \geq B_4\Big(\frac{n}{\log\big(\log(n)/2\big)}\Big)^{-\frac{\beta^*}{2\beta^*+1}}\right) \geq 1/4,$$

where $C' = \tilde{C}'(2\alpha_1,\beta_1+1)$ and

$$C_1 = \tilde{C}_1(2\alpha_1,\beta_1+1),\ C_2 = 1,\ B_4 = B(2\alpha_1,\beta_1+1,\infty). \tag{S7.9}$$

This concludes the proof.

Proof of Lemma 3. **(1) Proof of Equation** (S7.4)**:** For $1 \leq i \leq M$, $F_i \in \mathcal{A}(\alpha-t_i, K_i^{-t_i})$ by definition. For $x > K_i$, F_i satisfies the second-order Pareto condition. For any $1 \leq x \leq K_i$,

$$\begin{aligned}
\Big|1-F_i(x)-K_i^{-t_i}x^{-\alpha+t_i}\Big| &= \Big|x^{-\alpha}-K_i^{-t_i}x^{-\alpha+t_i}\Big| = x^{-\alpha}\Big|1-K_i^{-t_i}x^{t_i}\Big|\\
&\leq 2x^{-\alpha}\Big|t_i\log(K_i/x)\Big|.
\end{aligned}$$

The last inequality is obtained since $\forall u \in [0,1]$, $|e^{-u}-1| \le 2u$ and

$$t_i \log(K_i) \le n^{-\frac{\beta_i}{2\beta_i+1}} \big(\log M\big)^{\gamma_i} \big(\frac{1}{\alpha(2\beta_i+1)}\big) \log(n) \le \frac{1}{\alpha} n^{-\frac{\beta_i}{2\beta_i+1}} \log(n)^{\gamma_i+1} \le 1$$

by assuming large n. Then for any $1 \le x \le K_i$

$$\begin{aligned}\Big|1 - F_i(x) - K_i^{-t_i} x^{-\alpha+t_i}\Big| &\le 2x^{-\alpha} K_i^{-\alpha\beta_i} \log(K_i/x) = 2x^{-\alpha} x^{-\alpha\beta_i} \Big(\frac{K_i}{x}\Big)^{-\alpha\beta_i} \log(K_i/x) \\ &\le 2x^{-\alpha} x^{-\alpha\beta_i} \Big(\frac{K_i}{x}\Big)^{-\alpha(\beta-1)} \log(K_i/x) \\ &\le \frac{1}{\alpha(\beta-1)} x^{-\alpha(\beta_i+1)},\end{aligned}$$

where the ultimate inequality follows from the fact that for any $u \ge 1, t > 0$, we have $u^{-t}\log(u) \le 1/(et)$. Thus, we have shown the first result (S7.4).

(2) Proof of Equation (S7.5)**:** Let $1 \le j \le M$. Since $K_i > 1$ and $t_j > 0$ for all $i = 1, \dots, M$ and all $j = 1, \dots, M$, we have $K_i^{-t_j} \le 1$. By definition (S7.1) and bounding $M \le n$,

$$\begin{aligned}K_i^{-t_j} &\ge \Big(n^{\frac{1}{\alpha(2\beta_i+1)}}\Big)^{-n^{-\frac{\beta_j}{2\beta_j+1}}\big(\log M\big)^{\gamma_j}} = \exp\Big(-\frac{\log(n)}{\alpha(2\beta_i+1)} n^{-\frac{\beta_j}{2\beta_j+1}} \big(\log M\big)^{\gamma_j}\Big) \\ &\ge \exp\Big(-\frac{\log(n)^{1+\gamma_j}}{\alpha(2\beta-1)} n^{-\frac{\beta_j}{2\beta_j+1}}\Big) \\ &\ge \exp\Big(-\frac{1}{\alpha(2\beta-1)}\Big),\end{aligned}$$

where the final inequality follows for a sufficiently large n.

(3) Proof of Equation (S7.6)**:** Consider now $i < j$. From (S7.4), each F_i corresponds to the tail index $\alpha_i = \alpha - t_i = \alpha - (n/(\upsilon\log(M))^{-\beta_i/(2\beta_i+1)}$. For $i < j$, we have $\alpha_i > \alpha_j$ and $t_i < t_j$ as we described in the Step 1. Also, using $\beta_j - \beta_i = (i-j)/M$,

$$\begin{aligned}|\alpha_i - \alpha_j| &= \Big|t_j(1 - \frac{t_i}{t_j})\Big| = t_j \Big|1 - \big(\frac{n}{\upsilon\log(M)}\big)^{-\frac{\beta_i}{2\beta_i+1}+\frac{\beta_j}{2\beta_j+1}}\Big| \\ &= t_j\left[1 - \Big(\frac{n}{\upsilon\log(M)}\Big)^{\frac{(i-j)/M}{(2\beta_i+1)(2\beta_j+1)}}\right] = t_j\left[1 - \exp\Big(\frac{(i-j)}{M(2\beta_i+1)(2\beta_j+1)}\log\big(\frac{n}{\upsilon\log M}\big)\Big)\right] \\ &\ge t_j\left(1 - \exp\Big(\frac{(i-j)}{(2\beta_i+1)(2\beta_j+1)}\frac{(M-1)}{M}\Big)\right) \\ &\ge t_j\left[1 - \exp\Big(\frac{(i-j)}{2(2\beta_i+1)(2\beta_j+1)}\Big)\right],\end{aligned}$$

where the penultimate inequality is obtained since $\upsilon \le 1$, and since $\log\big(\frac{n}{\log(M)}\big) + 1 \ge M \ge 2$. This implies Equation (S7.6).

$\square$

Proof of Lemma 4. **(1) KL divergence between F_0 and F_i**

Let $1 \le i \le M$. By definition of KL divergence,

$$\begin{aligned} KL(F_0, F_i) &= \int_1^\infty f_0(x) \log\left(\frac{f_0(x)}{f_i(x)}\right) dx \\ &= -t_i \int_{K_i}^\infty \alpha x^{-\alpha-1} \log\left(\left(\frac{\alpha - t_i}{\alpha}\right)^{\frac{1}{t_i}} \frac{x}{K_i}\right) dx. \end{aligned}$$

By the change of variable $u = \left(\frac{\alpha-t_i}{\alpha}\right)^{1/t_i} x/K_i$, and letting $a_i = \left(\frac{\alpha-t_i}{\alpha}\right)^{1/t_i}$,

$$\begin{aligned} KL(F_0, F_i) &= -t_i \int_{a_i}^\infty \alpha\Big(\big(\frac{\alpha}{\alpha - t_i}\big)^{1/t_i} K_i u\Big)^{-\alpha-1} \log(u) du \times \Big(\big(\frac{\alpha}{\alpha - t_i}\big)^{1/t_i} K_i\Big) \\ &= t_i\Big(a_i^{-1} K_i\Big)^{-\alpha} \int_{a_i}^\infty (-\alpha) u^{-\alpha-1} \log(u) du. \end{aligned}$$

Now by performing an integration by parts, we obtain

$$\begin{aligned} KL(F_0, F_i) &= t_i\Big(a_i^{-1} K_i\Big)^{-\alpha} \left(u^{-\alpha} \log(u)\big|_{a_i}^\infty - \int_{a_i}^\infty u^{-\alpha-1} du \right) \\ &= t_i K_i^{-\alpha}\Big(\log(1/a_i) - \frac{1}{\alpha}\Big) = K_i^{-\alpha}\left(\log\left(\frac{\alpha}{\alpha - t_i}\right) - \frac{t_i}{\alpha}\right). \end{aligned}$$

Using $\alpha - t_i \ge \alpha/2$, we further upper bound this divergence

$$\begin{aligned} KL(F_0, F_i) &= K_i^{-\alpha}\left(\log\left(1 + \frac{t_i}{\alpha - t_i}\right) - \frac{t_i}{\alpha}\right) \le K_i^{-\alpha}\left(\frac{t_i}{\alpha - t_i} - \frac{t_i}{\alpha}\right) = K_i^{-\alpha} \frac{t_i^2}{\alpha(\alpha - t_i)} \\ &= \frac{2t_i^2 K_i^{-\alpha}}{\alpha^2}. \end{aligned}$$

(2) KL divergence between F_i and F_0

Similar calculations as above give

$$\begin{aligned} KL(F_i, F_0) &= \int_1^\infty f_i(x) \log \frac{f_i(x)}{f_0(x)} dx \\ &= t_i a_i^{-\alpha+t_i} K_i^{-\alpha} \int_{a_i}^\infty (\alpha - t_i) u^{-\alpha+t_i-1} \log(u) du \\ &= K_i^{-\alpha}\left(\log\left(\frac{\alpha - t_i}{\alpha}\right) + \frac{t_i}{\alpha - t_i}\right) \le \frac{2t_i^2 K_i^{-\alpha}}{\alpha^2}. \end{aligned}$$

$\square$

Proof of Lemma 5. **(1) KL divergence between F_i and F_j with $i < j$**

Consider the case $i < j$. First, note that

$$KL(F_i, F_j) := \int f_i(x) \log \frac{f_i(x)}{f_j(x)} dx$$
$$= KL(F_i, F_0) + \int_{K_j}^{\infty} f_i(x) \log \frac{f_0(x)}{f_j(x)} dx. \tag{S7.10}$$

Thus it suffices to bound the second term $\int_{K_j}^{\infty} f_i \log \frac{f_0}{f_j}$ in (S7.10).

We use the similar calculations used in the proof of Lemma 4. With the notation $a_j = (\frac{\alpha - t_j}{\alpha})^{1/t_j}$,

$$\begin{aligned}\int_{K_j}^{\infty} f_i(x) \log \frac{f_0(x)}{f_j(x)} dx &= t_j K_i^{-t_i} K_j^{-\alpha+t_i} a_j^{\alpha - t_j} \int_{a_j}^{\infty} -(\alpha - t_i) u^{-\alpha + t_i - 1} \log(u) du \\ &= \big(\frac{K_j}{K_i}\big)^{t_i} K_j^{-\alpha} a_j^{t_i - t_j} \left(\log \frac{1}{a_j} - \frac{1}{\alpha - t_i} \right) \\ &\leq 2 \exp\big(\frac{1}{\alpha(2\beta - 1)}\big) t_j^2 K_j^{-\alpha},\end{aligned}$$

where the final inequality follows by bounding $(K_j/K_i)^{t_i} \leq \exp\big(\frac{1}{\alpha(2\beta-1)}\big)$ using Lemma 3, and by bounding $a_j^{t_i - t_j}(\log(1/a_j) - 1/(\alpha - t_i)) \leq t_j^2/\alpha^2$ for a sufficiently large n.

Combining this upper bound with bounds on $KL(F_0, F_j)$ and $KL(F_i, F_0)$ in Lemma 4 and also with Equation (S7.10),

$$KL(F_i, F_j) \leq KL(F_i, F_0) + \exp\left(\frac{1}{\alpha(2\beta - 1)}\right) KL(F_0, F_j)$$
$$\leq \frac{2\exp(\frac{1}{\alpha(2\beta-1)})}{\alpha^2} \left(t_i^2 K_i^{-\alpha} + t_j^2 K_j^{-\alpha}\right). \tag{S7.11}$$

(2) KL between F_i and F_j with $i > j$

Now we turn to the case $i > j$. In the same way as for Equation (S7.10), we have

$$KL(F_i, F_j) = KL(F_i, F_0) + \int_{K_j}^{\infty} f_i(x) \log \frac{f_0(x)}{f_j(x)} dx. \tag{S7.12}$$

For the second term, first note that $\log \frac{f_0(x)}{f_j(x)}$ is a decreasing function for any $x \geq K_j$. Also since $\forall x \geq K_j$, $F_i(x) \leq F_0(x)$, and since $F_i(K_j) = F_0(K_j)$, the measure associated to F_i restricted to $[K_j, \infty)$ stochastically dominates F_0. This implies that

$$\int_{K_j}^{\infty} f_i(x) \log \frac{f_0(x)}{f_j(x)} dx \leq \int_{K_j}^{\infty} f_0(x) \log \frac{f_0(x)}{f_j(x)} dx.$$

Combining this with (S7.12) followed by Lemma 4, we have

$$KL(F_i, F_j) \leq KL(F_i, F_0) + KL(F_0, F_j) \leq \frac{2}{\alpha^2} \left(t_i^2 K_i^{-\alpha} + t_j^2 K_j^{-\alpha}\right). \tag{S7.13}$$

Finally, by Equations (S7.11) and (S7.13), we obtain the result (S7.7). $\square$

S8. Remarks on the proof of Theorem 4

Remark 1. We only proved the results for certain sets of $C_1, C_2, \beta_1, \beta_2, \alpha_1, \alpha_2, C'$. In fact, it is possible to modify this result to hold for different ranges of parameters (although, the ranges cannot be taken too tight, and C' cannot be taken too small). Note that the narrower the intervals $[C_1, C_2]$, $[\alpha_1, \alpha_2]$, the larger β_1 and the smaller C', the better the result is. Here are possible modification:

1. Range of α: from the proof, one could take $[\alpha_1, \alpha_1 + t_M]$ which is actually included in $[\alpha_1, \alpha_1 + n^{-\epsilon}]$ for some $\epsilon > 0$. So without additional effort, the interval can be taken at any position and the range of the interval can be made very small.
2. Range of β: for any $\beta_1 > 0$, the result holds for $[\beta_1, \beta_1 + 1]$ (although it is stated for $[\beta_1, \infty)$ to match the upper bound). The constants in the proof could be modified to consider a range $[\beta_1, \beta_1 + \epsilon]$ for any arbitrary small $\epsilon > 0$, by constructing M different β_i's uniformly spread on this interval.
3. Range of C: from the proof of the second result in Lemma 3, the tightest range of C is $[K_M^{-t_M}, K_1^{-t_1}]$ which is actually included in $[1 - n^{-\epsilon}, 1]$ for some $\epsilon > 0$. The range could be changed to any $[a - n^{-\epsilon}, a]$ for $a > 0$ by modifying distributions F_i so that the new distrubutions have a domain starting from $a^{-1/\alpha}$ instead of 1 in (S7.3). Then, the interval can be taken at any position a and the range of the interval can be made very small.

However, C' is an upper bound which characterizes the amount of deviation with respect to the Pareto assumption. It cannot be taken too small since if F_i's are too close to F_j's, they can not be distinguished.

S9. Appendix

Lemma 6 (Fano's inequality). *Suppose Y is a uniform random variable on $\{1, \ldots, M\}$, and let Z is a random variable of a function of X, where $X|Y = j \sim \mathbb{P}_j$ with $d\mathbb{P}_j/d\nu = p_j$ where ν is the dominating measure. Then*

$$\mathbb{P}\left(Z \neq Y\right) \geq 1 - \frac{1}{\log M}\left(\frac{1}{M^2}\sum_{j,j'} KL(\mathbb{P}_j, \mathbb{P}_{j'}) + \log 2\right).$$

Proof. Recall the definition of the entropy $H(Y) = -\sum_y p(y)\log p(y)$ for a discrete random variable Y with a probability mass function $p(y)$. Also we denote $H(Y|Z = z)$ by the conditional entropy of Y given $Z = z$, and we define $H(Y|Z) = -\sum_x \sum_y p(y, z) \log p(y|z)$. Following the terminology used in the information theory, we define *information* between Y and Z as the KL divergence between joint distribution and product of the marginal distribution, i.e. $I(Y, Z) = KL(P_{Y,Z}, P_Y \times P_Z)$ where we can show that

$$I(Y, Z) = KL(P_{Y,Z}, P_Y \times P_Z) = H(Y) - H(Y|Z) \tag{S9.1}$$

by splitting the probability distribution. Finally recall that for $Z = Z(X)$, $I(Y,Z) \leq I(Y,X)$.

Consider the event $E = \mathbf{1}\{Z \neq Y\}$. By splitting the probabilities with different order,

$$\begin{aligned} H(E,Y|Z) &= H(Y|Z) + H(E|Y,Z) := (1) \\ &= H(E|Z) + H(Y|E,Z) := (2), \end{aligned}$$

where $(1) = H(Y|Z)$ since E becomes a constant given Y and Z. Then we upper bound (2) as follows,

$$\begin{aligned} (2) &= H(E|Z) + H(Y|E,Z) \\ &\leq H(E) + H(Y|E,Z) \\ &= H(E) + \mathbb{P}(E=0)H(Y|E=0,Y) + \mathbb{P}(E=1)H(Y|E=1,Z) \\ &\leq \log 2 + \mathbb{P}(Z \neq Y)\log M. \end{aligned}$$

Combining both (1) and (2), we have

$$H(Y|Z) \leq \log 2 + \mathbb{P}(Z \neq Y)\log M,$$

in turn,

$$\mathbb{P}(Z \neq Y) \geq \frac{1}{\log M}\left(H(Y|Z) - \log 2\right). \tag{S9.2}$$

Now, using the fact (S9.1),

$$\begin{aligned} H(Y|Z) &= \log M - I(Y,Z) \\ &\geq \log M - I(Y,X) \\ &= \log M - \int \sum_y p(y)p(x|y)\log\frac{p(y)p(x|y)}{p(x)p(y)} \\ &= \log M - \int \sum_j \frac{1}{M}\mathbf{1}\{y=j\}p(x|y)\log\frac{p(x|y)}{p(x)} \\ &= \log M - \frac{1}{M}\sum_{j=1}^{M}\int p_j(x)\log\frac{p_j(x)}{\frac{1}{M}\sum_{j'}p_{j'}(x)}dx \\ &\geq \log M - \frac{1}{M^2}\sum_{j,j'} KL(\mathbb{P}_j, \mathbb{P}_{j'}), \end{aligned} \tag{S9.3}$$

where the penultimate equality is followed since $p(x) = \sum_j \mathbb{P}(Y=j)\mathbb{P}(X=x|Y=j) = \frac{1}{M}\sum_j p_j(x)$, and the last inequality is obtained by the concavity of the logarithm function. Combining (S9.2) and (S9.3), we obtain

$$\mathbb{P}(Z \neq Y) \geq 1 - \frac{1}{\log M}\left(\frac{1}{M^2}\sum_{j,j'} KL(\mathbb{P}_j, \mathbb{P}_{j'}) + \log 2\right).$$

□

References

Cover, T. M. and Thomas, J. A. (2012). *Elements of Information Theory*. Wiley-interscience.

Tsybakov, A. B.(2008). *Introduction to Nonparametric Estimation*. Springer.

van der Vaart, A. W.(2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge.