Supplementary Material for: A Data-driven Missing Mass Estimation Framework

Amichai Painsky
Department of Industrial Engineering, Tel Aviv University

January 10, 2022

Appendix A The Constrained Worst-case MSE

Theorem A.1. *Let $p \in \mathcal{P}_k$ be a probability distribution where $\mathcal{P}_k = \{p|\ a(u) \leq p(u) \leq b(u)\ \forall u \in \mathcal{X}\}$. Define $f_{n,r}^{max}(\mathcal{P}_k) = \max_{p\in\mathcal{P}_k} \sum_{u\in\mathcal{X}} p^r(u)e^{-np(u)}$. Let $\mathcal{C}_1 = \{u|\ a(u) > r/n\}$ be the set of all symbols for which $p^r(u)e^{-np(u)}$ is monotonically decreasing. Let $\mathcal{C}_2 = \mathcal{X} \setminus \mathcal{C}_1$ be the remaining symbols in the alphabet. Let $k_i = |\mathcal{C}_i|$, $A = 1 - \sum_{u\in\mathcal{C}_1} b(u)$ and $B = 1 - \sum_{u\in\mathcal{C}_1} a(u)$. Then,*

$$f_{n,r}^{max}(\mathcal{P}_k) \leq \sum_{u\in\mathcal{C}_1} a^r(u)e^{-na(u)} + g_{n,r}^{max}(\mathcal{C}_2) \tag{1}$$

where $g_{n,r}^{max}(\mathcal{C}_2)$ *is given by*

$$\begin{aligned} g_{n,r}^{max}(\mathcal{C}_2) = \max_{q(u)} \quad & \sum_{u \in \mathcal{C}_2} q^r(u) e^{-nq(u)} \\ \textit{subject to} \quad & A \le \sum_{u \in \mathcal{C}_2} q(u) \le B \\ & 0 \le q(u) \le 1 \ \ \textit{for all} \quad u \in \mathcal{C}_2 \end{aligned} \tag{2}$$

Further, the optimal solution to (2), $q^*(u)$*, satisfies*

- *if* $A \le k_2 r/n \le B$ *then* $q^*(u) = r/n$*.*
- *if* $k_2 r/n < A$ *then* $\sum_{u \in \mathcal{C}_2} q^*(u) = A$*.*
- *if* $k_2 r/n > B$ *then* $\sum_{u \in \mathcal{C}_2} q^*(u) = B$*.*

Proof. We are interested in the following optimization,

$$\begin{aligned} f_{n,r}^{max}(\mathcal{P}_k) \quad = \quad \max_{p(u)} \quad & \sum_{u \in \mathcal{X}} p^r(u) e^{-np(u)} \\ \text{subject to} \quad & \sum_{u \in \mathcal{X}} p(u) = 1 \\ & a(u) \le p(u) \le b(u) \ \ \text{for all} \quad u \in \mathcal{X}. \end{aligned} \tag{3}$$

Let us distinguish between two disjoint sets of variables. Define

$$\mathcal{C}_1 = \{u | \ a(u) > r/n\}. \tag{4}$$

In words, $\mathcal{C}_1$ is the set of all symbols for which $p^r(u)e^{-np(u)}$ is necessarily monotonically decreasing (that is, the $p(u)$'s are constrained to the monotonically decreasing region of the function). Further, $\mathcal{C}_2 = \mathcal{X} \setminus \mathcal{C}_1$ are the remaining symbols. Let us first assume that $\mathcal{C}_1$ is not an empty set. Reformulating (3), we obtain

$$\begin{aligned} \max_{p(u),\rho} \quad & \sum_{u\in\mathcal{C}_1} p^r(u)e^{-np(u)} + \sum_{u\in\mathcal{C}_2} p^r(u)e^{-np(u)} \\ \text{subject to} \quad & \sum_{u\in\mathcal{C}_1} p(u) = \rho,\ \sum_{u\in\mathcal{C}_2} p(u) = 1-\rho \\ & \sum_{u\in\mathcal{C}_1} a(u) \le \rho \le \sum_{u\in\mathcal{C}_1} b(u) \\ & a(u) \le p(u) \le b(u) \ \text{ for all } \quad u \in \mathcal{X} \end{aligned} \tag{5}$$

where the equivalence holds due to the separability of the objective function and the constraints. Let us now fix $\rho = \rho_0$ and study

$$\begin{aligned} \max_{p(u)} \quad & \sum_{u\in\mathcal{C}_1} p^r(u)e^{-np(u)} + \sum_{u\in\mathcal{C}_2} p^r(u)e^{-np(u)} \\ \text{subject to} \quad & \sum_{u\in\mathcal{C}_1} p(u) = \rho_0,\ \sum_{u\in\mathcal{C}_2} p(u) = 1-\rho_0 \\ & a(u) \le p(u) \le b(u) \ \text{ for all } \quad u \in \mathcal{X}. \end{aligned} \tag{6}$$

Define $f^{max}_{n,r,1}$ as the maximum of

$$\begin{aligned} \max_{p(u)} \quad & \sum_{u\in\mathcal{C}_1} p^r(u)e^{-np(u)} \\ \text{subject to} \quad & \sum_{u\in\mathcal{C}_1} p(u) = \rho_0, \\ & a(u) \le p(u) \le b(u) \ \text{ for all } \quad u \in \mathcal{C}_1 \end{aligned} \tag{7}$$

while $f^{max}_{n,r,2}$ is the maximum of

$$\begin{aligned}
\max_{p(u)} \quad & \sum_{u\in\mathcal{C}_2} p^r(u)e^{-np(u)} \\
\text{subject to} \quad & \sum_{u\in\mathcal{C}_2} p(u) = 1-\rho_0, \\
& a(u) \le p(u) \le b(u) \;\; \text{for all} \quad u \in \mathcal{C}_2.
\end{aligned} \tag{8}$$

Then, $f^{max}_{n,r,1} + f^{max}_{n,r,2}$ bounds from above (6). Now, observe that $f^{max}_{n,r,1} \le \sum_{u\in\mathcal{C}_1} a^r(u)e^{-na(u)}$ since $p^r(u)e^{-np(u)}$ is monotonically decreasing for all $u \in \mathcal{C}_1$. Further, we may bound $f^{max}_{n,r,2}$ from above by relaxing the inequality constraints of (8), as follows

$$\begin{aligned}
\max_{p(u)} \quad & \sum_{u\in\mathcal{C}_2} p^r(u)e^{-np(u)} \\
\text{subject to} \quad & \sum_{u\in\mathcal{C}_2} p(u) = 1-\rho_0 \\
& 0 \le p(u) \le 1 \;\; \text{for all} \quad u \in \mathcal{C}_2.
\end{aligned} \tag{9}$$

Going back to our original problem (before fixing ρ), we have that (5) is bounded from above by $\sum_{u\in\mathcal{C}_1} a^r(u)e^{-na(u)} + g^{max}_{n,r,2}$ where $g^{max}_{n,r,2}$ is defined as

$$\begin{aligned}
\max_{p(u),\rho} \quad & \sum_{u\in\mathcal{C}_2} p^r(u)e^{-np(u)} \\
\text{subject to} \quad & \sum_{u\in\mathcal{C}_2} p(u) = 1-\rho, \;\; 0 \le p(u) \le 1 \;\; \forall\, u \in \mathcal{C}_2 \\
& 1 - \sum_{u\in\mathcal{C}_1} b(u) \le 1-\rho \le 1 - \sum_{u\in\mathcal{C}_1} a(u)
\end{aligned} \tag{10}$$

That is, we conclude that (3) is bounded from above by the sum of $\sum_{u \in \mathcal{C}_1} a^r(u) e^{-na(u)}$ and

$$\begin{aligned} \max_{p(u)} \quad & \sum_{u \in \mathcal{C}_2} p^r(u) e^{-np(u)} \\ \text{subject to} \quad & A \leq \sum_{u \in \mathcal{C}_2} p(u) \leq B \\ & 0 \leq p(u) \leq 1 \ \text{ for all } \quad u \in \mathcal{C}_2 \end{aligned} \tag{11}$$

where $A = 1 - \sum_{u \in \mathcal{C}_1} b(u)$ and $B = 1 - \sum_{u \in \mathcal{C}_1} a(u)$. Notice that for the case where $\mathcal{C}_1$ is an empty set, we degenerate back to Theorem 3 of [1].

Finally, let us study (11). First, we relax the marginal constraints over $p(u)$ for the simplicity of the presentation, as later justified in the obtained solution. The Lagrangian of the problem is

$$\mathcal{L} = \sum_{u \in \mathcal{C}_2} p^r(u) e^{-np(u)} - \mu_1 \left(\sum_{u \in \mathcal{C}_2} p(u) - B \right) - \mu_2 \left(A - \sum_{u \in \mathcal{C}_2} p(u) \right) \tag{12}$$

and Karush–Kuhn–Tucker (KKT) optimality conditions imply

(a) $p^{r-1}(u) e^{-np(u)} \left(r - np(u) \right) - \mu_1 + \mu_2 = 0$

(b) $\mu_1 \left(\sum p(u) - B \right) = 0$

(c) $\mu_2 \left(A - \sum p(u) \right) = 0$

(d) $\mu_1, \mu_2 \geq 0$

We now distinguish between four different Cases. First, assume $\mu_1 = \mu_2 = 0$. In this case (a) implies $p(u) = r/n$, subject to the constraints $A \leq k_2 r/n \leq B$. Next, assume that $\mu_1 > 0$. This means that $\sum p(u) = B$ (b) and $\mu_2 = 0$ (c). Therefore, we have that

$p^{r-1}(u)e^{-np(u)}\left(r - np(u)\right) = \mu_1$ (a), which can hold only if $p(u) \le r/n$ for all $u \in \mathcal{C}_2$. In other words, $\mu_1 > 0$ may satisfy the KKT conditions only if $\sum p(u) \le k_2 r/n$. Alternatively, if $\sum p(u) \ge A > k_2 r/n$ then $\mu_1 > 0$ cannot satisfy the optimiality conditions. Further, assume that $\mu_2 > 0$. Here, $\sum p(u) = A$ (c) and $\mu_1 = 0$ (b). Therefore, we have that $p^{r-1}(u)e^{-np(u)}\left(r - np(u)\right) = -\mu_2$ (a), which can hold only if $p(u) \ge r/n$ for all $u \in \mathcal{C}_2$. As above, this conditions may hold only if $\sum p(u) \ge k_2 r/n$. Alternatively, if $\sum p(u) \le B > k_2 r/n$ then $\mu_2 > 0$ cannot satisfy the optimiality conditions. Finally, notice that for $\mu_1, \mu_2 > 0$, the KKT cannot hold due to (b) and (c). To conclude, we observe the following

- if $A \le k_2 r/n \le B$ then $p(u) = r/n$ is the optimal solution (corresponds to $\mu_1 = \mu_2 = 0$ while contradicts the other three cases).
- if $k_2 r/n < A$ then $\mu_2 > 0$ (the only case which does not imply a contradiction) and $\sum p(u) = A$ in this setup.
- if $k_2 r/n > B$ then $\mu_1 > 0$ and $\sum p(u) = B$ in this setup.

□

Appendix B The Properties of $g_{n,r}^{max}(\mathcal{C}_2)$

Let $\mathcal{P} \subseteq \Delta_{k_2}$ be the set of probability distributions that satisfy $\sum_{u\in\mathcal{C}_2} p(u) \le C$.Let

$$\begin{aligned} g_{n,r}^{max}(\mathcal{C}_2) = \max_{p(u)\in\mathcal{P}} \quad & \sum_{u\in\mathcal{C}_2} p^r(u)e^{-np(u)} \\ \text{subject to} \quad & \sum_{u\in\mathcal{C}_2} p(u) \le C \\ & 0 \le p(u) \ \text{ for all } \ u \in \mathcal{C}_2 \end{aligned} \tag{13}$$

Property 1. *Let* $g_{n,r} = \sum_{u \in \mathcal{C}_2} p^r(u)e^{-np(u)}$*. The summand,* $h_r(p(u)) = p^r(u)e^{-np(u)}$ *satisfies the following:*

1. $h_r(p(u))$ *is concave in* $p(u)$*, for* $\frac{r-\sqrt{r}}{n} \leq p(u) \leq \frac{r+\sqrt{r}}{n}$.

2. $h_r(p(u))$ *is convex in* $p(u)$*, for* $0 \leq p(u) \leq \frac{r-\sqrt{r}}{n}$ *and* $\frac{r+\sqrt{r}}{n} \leq p(u) \leq 1$.

Proof. The second derivative of $h_r(p(u))$,

$$\frac{d^2 h_r(p(u))}{dp^2(u)} = p^{r-2}(u)e^{-np(u)} \left(n^2 p^2(u) - 2nrp(u) + r(r-1)\right)$$

is negative for $\frac{r-\sqrt{r}}{n} < p(u) < \frac{r+\sqrt{r}}{n}$ and positive for $0 < p(u) < \frac{r-\sqrt{r}}{n}$ and $\frac{r+\sqrt{r}}{n} < p(u) < 1$. See Figure 1 for an illustration of $h_r(p(u))$. □

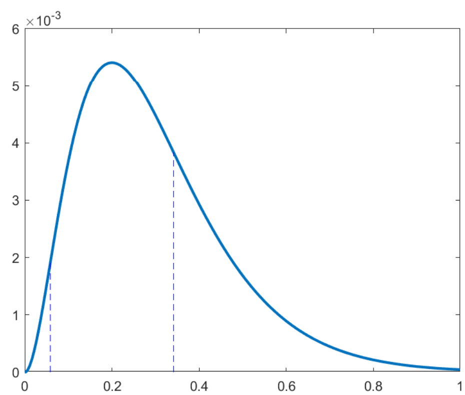

Figure 1: $h_r(p(u))$ for $r = 2$ and $n = 10$. The dashed lines are the bounds between the convex and concave regions

Property 2. *Let* $p^* \in \mathcal{P}$ *be the maximizer of* $g_{n,r} = \sum_{u \in \mathcal{C}_2} h_r(p(u))$ *over the set* $\mathcal{P}$ *. Then,* $p^*(u) = p^*(v)$ *for all* $p^*(u), p^*(v) \in \left[\frac{r-\sqrt{r}}{n}, \frac{r+\sqrt{r}}{n}\right]$.

Proof. By negation, assume there exists $p^*(u) \neq p^*(v)$ such that $p^*(u), p^*(v) \in \left[\frac{r-\sqrt{r}}{n}, \frac{r+\sqrt{r}}{n}\right]$. Define

$$\tilde{p}(l) = \begin{cases} p^*(l) & l \neq u, v \\ \frac{p^*(u)+p^*(v)}{2} & l = u, v \end{cases} \tag{14}$$

Then,

$$\begin{aligned} \sum_l h_r(\tilde{p}(l)) = \sum_{l \neq u,v} h_r(\tilde{p}(l)) + \sum_{l=u,v} h_r(\tilde{p}(l)) = \\ \sum_{l \neq u,v} h_r(p^*(l)) + 2h_r\left(\frac{p^*(u)+p^*(v)}{2}\right) > \\ \sum_{l \neq u,v} h_r(p^*(l)) + h_r(p^*(u)) + h_r(p^*(v)) = \sum_u h_r(p^*(u)). \end{aligned} \tag{15}$$

where the inequality follows from the concavity of $h_r(p(l))$ for every $p(l) \in \left[\frac{r-\sqrt{r}}{n}, \frac{r+\sqrt{r}}{n}\right]$. Therefore, we found $\tilde{p} \in \mathcal{P}$ for which $\sum_l h_r(\tilde{p}(l)) > \sum_l h_r(p^*(l))$, which contradicts the optimality of p^*. □

Property 3. *Let $p^* \in \mathcal{P}$ be the maximizer of $f_{n,r} = \sum_u h_r(p(u))$ over the set $\mathcal{P}$. Assume that $C \geq \frac{r+\sqrt{r}}{n}$. Then, there exists at most a single $p^*(u)$ such that $p^*(u) \in \left(\frac{r+\sqrt{r}}{n}, C\right]$.*

Proof. By negation, assume there exist $p^*(u)$ and $p^*(v)$ such that $p^*(u), p^*(v) \in \left(\frac{r+\sqrt{r}}{n}, C\right]$. Assume, without loss of generality, that $p^*(v) \leq p^*(u)$. Define $\delta = p^*(v) - \frac{r+\sqrt{r}}{n} > 0$. The function $h_r(p(u))$ is convex for $p(u) \in \left[\frac{r+\sqrt{r}}{n}, C\right]$ and strictly convex for $p(u) \in \left(\frac{r+\sqrt{r}}{n}, C\right]$. Therefore, we have

$$h_r\left(\frac{r+\sqrt{r}}{n}\right) \geq h_r\left(p^*(v)\right) - \delta h_r'\left(p^*(v)\right) \tag{16}$$

$$h_r\left(p^*(u)+\delta\right) > h_r\left(p^*(u)\right) + \delta h_r'\left(p^*(u)\right) \tag{17}$$

where $h_r'(p(u)) = \frac{dh_r(p(u))}{dp(u)}$. Putting together the above, we have

$$\begin{aligned} h_r\left(\frac{r+\sqrt{r}}{n}\right) &+ h_r\left(p^*(u)+\delta\right) > \\ &h_r\left(p^*(v)\right) + h_r\left(p^*(u)\right) + \delta\left(h_r'\left(p^*(u)\right) - h_r'\left(p^*(v)\right)\right). \end{aligned} \tag{18}$$

We observe that $h_r'(p(u))$ is an increasing function in $p(u)$, for $p(u) \in \left(\frac{r+\sqrt{r}}{n}, C\right]$, as its derivative, $\frac{d^2h_r(p(u))}{dp^2(u)}$ is positive in this range. Therefore, $h_r'\left(p^*(u)\right) \geq h_r'\left(p^*(v)\right)$ and

$$h_r\left(\frac{r+\sqrt{r}}{n}\right) + h_r\left(p^*(u)+\delta\right) > h_r\left(p^*(v)\right) + h_r\left(p^*(u)\right). \tag{19}$$

Therefore, we define $\tilde{p} \in \mathcal{P}$ such that

$$\tilde{p}(l) = \begin{cases} p^*(l) & l \neq u, v \\ p^*(l) - \delta & l = v \\ p^*(l) + \delta & l = u \end{cases} \tag{20}$$

and $\sum_l h_r(\tilde{p}(l)) > \sum_l h_r(p^*(l))$, which contradicts the optimality of p^*. □

Property 4. *Let $p^* \in \mathcal{P}$ be the maximizer of $f_{n,r} = \sum_u h_r(p(u))$ over $\mathcal{P}$. Then, there exists at most a single $p^*(u)$ such that $p^*(u) \in \left(0, \frac{r-\sqrt{r}}{n}\right)$.*

Proof. By negation, assume there exist $p^*(u)$ and $p^*(v)$ such that $p^*(u), p^*(v) \in \left(0, \frac{r-\sqrt{r}}{n}\right)$. Assume, without loss of generality, that $p^*(v) \leq p^*(u)$. Define $\delta = p^*(v) > 0$.

Let us first assume that $p^*(u) + \delta < \frac{r-\sqrt{r}}{n}$. The function $h_r(p(u))$ is convex for $p(u) \in$

$\left[0, \frac{r-\sqrt{r}}{n}\right]$ and strictly convex for $p(u) \in \left[0, \frac{r-\sqrt{r}}{n}\right)$. Therefore, we have

$$h_r\left(p^*(u)+\delta\right) > h_r\left(p^*(u)\right) + \delta h_r'\left(p^*(u)\right) \tag{21}$$

$$h_r\left(p^*(v)-\delta\right) \geq h_r\left(p^*(v)\right) - \delta h_r'\left(p^*(v)\right). \tag{22}$$

Putting together the above, we have

$$\begin{aligned} h_r\left(p^*(u)+\delta\right) + & h_r\left(p^*(v)-\delta\right) > \\ & h_r\left(p^*(u)\right) + h_r\left(p^*(v)\right) + \delta\left(h_r'\left(p^*(u)\right) - h_r'\left(p^*(v)\right)\right). \end{aligned} \tag{23}$$

We observe that $h_r'(p(u))$ is an increasing function in $p(u)$, for $p(u) \in \left(0, \frac{r-\sqrt{r}}{n}\right)$, as its derivative, $\frac{d^2 h_r(p(u))}{dp^2(u)}$ is positive in this range. Therefore, $h_r'\left(p^*(u)\right) \geq h_r'\left(p^*(v)\right)$ and

$$h_r\left(p^*(u)+\delta\right) + h_r\left(p^*(v)-\delta\right) > h_r\left(p^*(u)\right) + h_r\left(p^*(v)\right). \tag{24}$$

Therefore, we found $\tilde{p} \in \mathcal{P}$ such that

$$\tilde{p}(l) = \begin{cases} p^*(l) & l \neq u, v \\ 0 & l = v \\ p^*(u)+\delta & l = u \end{cases} \tag{25}$$

and $\sum_l h_r(\tilde{p}(l)) > \sum_l h_r(p^*(l))$, which contradicts the optimality of p^*.

Now, assume that $p^*(u) + \delta \geq \frac{r-\sqrt{r}}{n}$. Then, define $\tilde{\delta} = \frac{r-\sqrt{r}}{n} - p^*(u) > 0$. We have

$$h_r\left(p^*(u) + \tilde{\delta}\right) \geq h_r\left(p^*(u)\right) + \tilde{\delta} h_r'\left(p^*(u)\right) \tag{26}$$

$$h_r\left(p^*(v) - \tilde{\delta}\right) > h_r\left(p^*(v)\right) - \tilde{\delta} h_r'\left(p^*(v)\right). \tag{27}$$

Putting together the above, we have

$$\begin{aligned} h_r\left(p^*(u) + \tilde{\delta}\right) + & h_r\left(p^*(v) - \tilde{\delta}\right) > \\ & h_r\left(p^*(u)\right) + h_r\left(p^*(v)\right) + \tilde{\delta}\left(h_r'\left(p^*(u)\right) - h_r'\left(p^*(v)\right)\right). \end{aligned} \tag{28}$$

As above, we observe that $h_r'(p(u))$ is an increasing function in $p(u)$, for $p(u) \in \left(0, \frac{r-\sqrt{r}}{n}\right)$. Therefore, $h_r'\left(p^*(u)\right) \geq h_r'\left(p^*(v)\right)$ and

$$h_r\left(p^*(v) - \tilde{\delta}\right) + h_r\left(p^*(u) + \tilde{\delta}\right) > h_r\left(p^*(v)\right) + h_r\left(p^*(u)\right). \tag{29}$$

Therefore, we found $\tilde{p} \in \mathcal{P}$ such that

$$\tilde{p}(l) = \begin{cases} p^*(l) & l \neq u, v \\ p^*(v) - \tilde{\delta} & l = v \\ \frac{r-\sqrt{r}}{n} & l = u \end{cases} \tag{30}$$

and $\sum_l h_r(\tilde{p}(l)) > \sum_l h_r(p^*(l))$, which again contradicts the optimality of p^*. $\square$

Property 5. *Let $p^* \in \mathcal{P}$ be the maximizer of $f_{n,r} = \sum_u h_r(p(u))$ over $\mathcal{P}$. Assume that*

$k_2\frac{r+\sqrt{r}}{n} < C$. *Then, there exists a single* $p^*(u)$ *such that* $p^*(u) \in \left(\frac{r+\sqrt{r}}{n}, C\right]$, *while* $p^*(v) = \frac{1-p^*(u)}{k_2-1}$ *for all* $v \neq u$.

Proof. Property 1 characterizes $h_r(p(u))$ for different values of $p(u)$. Let us first assume that $p^*(u) \leq \frac{r+\sqrt{r}}{n}$ for every u. This means that $C = \sum_{u\in\mathcal{C}_2} p^*(u) \leq k_2\frac{r+\sqrt{r}}{n}$. This contradicts the assumption of Property 5, that $k_2\frac{r+\sqrt{r}}{n} < C$. Therefore, there exists at least a single $p^*(u)$ such that $p^*(u) \in \left(\frac{r+\sqrt{r}}{n}, C\right]$. On the other hand, Property 3 shows that there exists at most a single $p^*(u)$ such that $p^*(u) \in \left(\frac{r+\sqrt{r}}{n}, C\right]$. Therefore, we conclude that for $k\frac{r+\sqrt{r}}{n} < C$, there exists exactly a single $p^*(u)$ such that $p^*(u) \in \left(\frac{r+\sqrt{r}}{n}, C\right]$. Next, Property 4 shows that there exists at most a single $p^*(l) \in \left(0, \frac{r-\sqrt{r}}{n}\right)$. Define $\tilde{p}(l) = p^*(l) + \delta$ and $\tilde{p}(u) = p^*(u) - \delta$ where $\delta = \frac{r-\sqrt{r}}{n} - p^*(l)$. Notice we have that $\tilde{p}(u) \in \left(\frac{r+\sqrt{r}}{n}, C\right]$, as $k_2\frac{r+\sqrt{r}}{n} < C$. We have,

$$h_r(\tilde{p}(l)) = h_r\left(\frac{r-\sqrt{r}}{n}\right) > h_r(p^*(l)) \tag{31}$$

$$h_r(\tilde{p}(u)) > h_r(p^*(u)), \tag{32}$$

as $h_r(p(u))$ is monotonically increasing in $p(u) \in \left[0, \frac{r-\sqrt{r}}{n}\right]$ and monotonically decreasing in $p(u) \in \left[\frac{r+\sqrt{r}}{n}, C\right]$. Therefore, we show that by increasing $p^*(l)$ by δ, and reducing $p^*(u)$ by the same δ, we increasing the objective. We notice that the same argument applies for every $p^*(l) = 0$, as well. Therefore, we conclude that the maximizer of $f_{n,r}$ consists of a single $p^*(u) \in \left(\frac{r+\sqrt{r}}{n}, C\right]$, while $p^*(l) \in \left[\frac{r-\sqrt{r}}{n}, \frac{r+\sqrt{r}}{n}\right]$ for every $l \neq u$. Finally, we apply Property 2 to attain the desired result. $\square$

References

[1] Amichai Painsky. Generalized good-turing improves missing mass estimation. *Journal of the American Statistical Association*, (just-accepted):1–27, 2021.