Supplementary Material for: Refined Convergence Rates of the Good-Turing Estimator

Amichai Painsky
Tel Aviv University

May 13, 2021

Appendix A The MSE of GT for Fixed k

The squared error of the GT estimator follows

$$\begin{aligned}\left(\hat{M}_k^{GT} - M_k\right)^2 = &\left(\sum_{u \in \mathcal{X}} \frac{k+1}{n} \mathbb{1}(N_u(X^n) = k+1) - p(u)\mathbb{1}(N_u(X^n) = k)\right) \cdot \\ &\left(\sum_{v \in \mathcal{X}} \frac{k+1}{n} \mathbb{1}(N_v(X^n) = k+1) - p(v)\mathbb{1}(N_v(X^n) = k)\right) = \\ &\left(\frac{k+1}{n}\right)^2 \sum_{u,v \in \mathcal{X}} \mathbb{1}(N_u(X^n) = k+1)\mathbb{1}(N_v(X^n) = k+1) - \\ &\frac{2(k+1)}{n} \sum_{u,v \in \mathcal{X}} p(u)\mathbb{1}(N_u(X^n) = k)\mathbb{1}(N_v(X^n) = k+1) + \\ &\sum_{u,v \in \mathcal{X}} p(u)p(v)\mathbb{1}(N_u(X^n) = k)\mathbb{1}(N_v(X^n) = k).\end{aligned} \tag{1}$$

Therefore, the expected error satisfies

$$\mathbb{E}_{X^n\sim p}\left(\hat{M}_k^{GT}-M_k\right)^2 = \tag{2}$$
$$\frac{1}{n^2}\sum_{u,v\in\mathcal{X}}(k+1)^2P_n(k+1,k+1)-2(k+1)np(u)P_n(k,k+1)+n^2p(u)p(v)P_n(k,k).$$

where $P_n(i,j)=\mathbb{E}_{X^n\sim p}\left(\mathbb{1}(N_u(X^n)=i)\mathbb{1}(N_v(X^n)=j)\right)$, and

$$P_n(i,j)=\begin{cases}\binom{n}{i\ j}p^i(u)p^j(v)(1-p(u)-p(v))^{n-i-j} & u\neq v, i+j\leq n\\ \binom{n}{i}p^i(u)(1-p(u))^{n-i} & u=v,\ i=j\\ 0 & o.w.\end{cases} \tag{3}$$

Define $P(u,v)=p^{k+1}(u)p^{k+1}(v)(1-p(u)-p(v))^{n-2k-2}$. Plugging the above to (2) yields

$$\mathbb{E}_{X^n\sim p}\left(\hat{M}_k^{GT}-M_k\right)^2=\frac{1}{n^2}\sum_{u\neq v}P(u,v)\Bigg((k+1)^2\binom{n}{k+1\ k+1}- \tag{4}$$
$$2n(k+1)\binom{n}{k\ k+1}(1-p(u)-p(v))+n^2\binom{n}{k\ k}(1-p(u)-p(v))^2\Bigg)+$$
$$\left(\frac{k+1}{n}\right)^2\binom{n}{k+1}\sum_u p^{k+1}(u)(1-p(u))^{n-k-1}+\binom{n}{k}\sum_u p^{k+2}(u)(1-p(u))^{n-k}$$

for $2k<n$. Let us now focus on the first summation. Notice that $\binom{n}{k\ k+1}=\binom{n}{k\ k}\frac{n-2k}{k+1}$ and $\binom{n}{k+1\ k+1}=\binom{n}{k\ k}\frac{(n-2k)(n-2k-1)}{(k+1)^2}$. Then, the first summation equals

$$\frac{1}{n^2}\binom{n}{k\ k}\sum_{u\neq v}P(u,v)\big(2k(2k+1)-n-4nk(p(u)+p(v))+n^2(p(u)+p(v))^2\big). \tag{5}$$

Let us study the different terms in (5). First,

$$\binom{n}{k\ k}\sum_{u\neq v}P(u,v)(p(u)+p(v))^2=\binom{n}{k\ k}\sum_{u\neq v}P(u,v)\left(p^2(u)+2p(v)p(v)+p^2(v)\right). \quad (6)$$

Lemma 1 in [3] states that

$$\sum_{u\neq v}p^i(u)p^j(v)(1-p(u)-p(v))^n\leq\frac{(i-1)!(j-1)!n!}{(n+i+j-2)!} \quad (7)$$

Plugging (7) to (6) yields

$$\binom{n}{k\ k}\sum_{u\neq v}P(u,v)(p(u)+p(v))^2=o\left(\frac{1}{n}\right). \quad (8)$$

Similarly, we have

$$\frac{1}{n^2}\binom{n}{k\ k}\sum_{u\neq v}P(u,v)\big(2k(2k+1)-4nk(p(u)+p(v))\big)=o\left(\frac{1}{n}\right) \quad (9)$$

Therefore, the first term in (4) equals

$$-\frac{1}{n}\binom{n}{k\ k}\sum_{u\neq v}P(u,v)+o\left(\frac{1}{n}\right) \quad (10)$$

and

$$\mathbb{E}_{X^n\sim p}\left(\hat{M}_k^{GT}-M_k\right)^2 = -\frac{1}{n}\binom{n}{k\ k}\sum_{u\neq v}P(u,v)+ \tag{11}$$
$$\left(\frac{k+1}{n}\right)^2\binom{n}{k+1}\sum_u p^{k+1}(u)(1-p(u))^{n-k-1}+$$
$$\binom{n}{k}\sum_u p^{k+2}(u)(1-p(u))^{n-k}+o\left(\frac{1}{n}\right).$$

We now rewrite (11) in a more compact manner. First, we have that

$$\mathbb{E}_{X^n\sim p}\left(\Phi_k\left(X^n\right)\right) = \mathbb{E}_{X^n\sim p}\left(\sum_u \mathbb{1}(N_u(X^n)=k)\right) = \tag{12}$$
$$\mathbb{E}_{X^n\sim p}\left(\sum_{u=v}\mathbb{1}(N_u(X^n)=k)\mathbb{1}(N_v(X^n)=k)\right) =$$
$$\sum_{u=v}P_n(k,k) = \binom{n}{k}\sum_u p^k(u)(1-p(u))^{n-k}.$$

We begin with the first term in (11),

$$\frac{1}{n}\binom{n}{k\ k}\sum_{u\neq v}P(u,v) = \frac{1}{n}\left(\binom{n}{k\ k}\Big/\binom{n}{k+1\ k+1}\right)\sum_{u\neq v}P_n(k+1,k+1) = \tag{13}$$
$$\frac{(k+1)^2}{n(n-2k)(n-2k-1)}\mathbb{E}_{X^n\sim p}\sum_{u\neq v}\mathbb{1}(N_u(X^n)=k+1)\mathbb{1}(N_v(X^n)=k+1) =$$
$$\frac{(k+1)^2}{n(n-2k)(n-2k-1)}\mathbb{E}_{X^n\sim p}\left(\left(\sum_u \mathbb{1}(N_u(X^n)=k+1)\right)^2-\sum_u \mathbb{1}(N_u(X^n)=k+1)\right) =$$
$$\frac{(k+1)^2}{n(n-2k)(n-2k-1)}\mathbb{E}_{X^n\sim p}\left(\Phi_{k+1}^2(X^n)-\Phi_{k+1}(X^n)\right).$$

Notice that

$$\mathbb{E}_{X^n\sim p}\left(\Phi_{k+1}(X^n)\right) = \binom{n}{k+1}\sum_u p^{k+1}(u)(1-p(u))^{n-k-1} \leq \binom{n}{k+1}\frac{k!(n-k-1)!}{(n-1)!} = \frac{n}{k+1}. \tag{14}$$

where the first equality is due to (12) and the inequality follows from Lemma 2 in [3],

$$\sum_{u\in\mathcal{X}} p(u)^i(1-p(u))^n \leq \frac{(i-1)!n!}{(n-1+i)!}. \tag{15}$$

Plugging (14) to (13), we obtain

$$-\frac{1}{n}\binom{n}{k\ k}\sum_{u\neq v} P(u,v) = \frac{-(k+1)^2}{n(n-2k)(n-2k-1)}\mathbb{E}_{X^n\sim p}\left(\Phi^2_{k+1}(X^n)\right) + o\left(\frac{1}{n}\right). \tag{16}$$

We now continue to the second term in (11). Here, we have that

$$\left(\frac{k+1}{n}\right)^2\binom{n}{k+1}\sum_u p^{k+1}(u)(1-p(u))^{n-k-1} = \left(\frac{k+1}{n}\right)^2\mathbb{E}_{X^n\sim p}\left(\Phi_{k+1}(X^n)\right) \tag{17}$$

where the equality follows from (12). Finally, the third term in (11) satisfies

$$\binom{n}{k}\sum_u p^{k+2}(u)(1-p(u))^{n-k} = \binom{n}{k}\sum_u p^{k+2}(u)(1-p(u))^{n-k-2}(1-p(u))^2 = \tag{18}$$
$$\binom{n}{k}\sum_u p^{k+2}(u)(1-p(u))^{n-k-2} - 2\binom{n}{k}\sum_u p^{k+3}(u)(1-p(u))^{n-k-2} +$$
$$\binom{n}{k}\sum_u p^{k+4}(u)(1-p(u))^{n-k-2} = \binom{n}{k}\sum_u p^{k+2}(u)(1-p(u))^{n-k-2} + o\left(\frac{1}{n}\right) =$$
$$\frac{(k+1)(k+2)}{(n-k)(n-k-1)}\mathbb{E}_{X^n\sim p}\left(\Phi_{k+2}(X^n)\right) + o\left(\frac{1}{n}\right).$$

where the third equality follows from (15) and the final equality is due to (12). Putting together (16), (17) and (18), we conclude that

$$\mathbb{E}_{X^n\sim p}\left(\hat{M}_k^{GT} - M_k\right)^2 = \frac{-(k+1)^2}{n(n-2k)(n-2k-1)}\mathbb{E}_{X^n\sim p}\left(\Phi^2_{k+1}(X^n)\right) + \tag{19}$$
$$\left(\frac{k+1}{n}\right)^2\mathbb{E}_{X^n\sim p}\left(\Phi_{k+1}(X^n)\right) + \frac{(k+1)(k+2)}{(n-k)(n-k-1)}\mathbb{E}_{X^n\sim p}\left(\Phi_{k+2}(X^n)\right) + o\left(\frac{1}{n}\right) \leq$$
$$\frac{-(k+1)^2}{n(n-2k)(n-2k-1)}f^2_{n,k+1}(p) + \left(\frac{k+1}{n}\right)^2 f_{n,k+1}(p) + \frac{(k+1)(k+2)}{(n-k)(n-k-1)}f_{n,k+2}(p) + o\left(\frac{1}{n}\right)$$

where $f_{n,k}(p) \triangleq \mathbb{E}_{X^n\sim p}\left(\Phi_k(X^n)\right) = \binom{n}{k}\sum_u p^k(u)(1-p(u))^{n-k}$ and the second inequality is due to $\mathbb{E}(X^2) \geq \mathbb{E}^2(X)$. Let us now bound (19) for every possible probability distribution. First, we apply Proposition 3 from the main text to $f_{n,k}(p)$ and obtain

$$f_{n,k}(p) = \binom{n}{k}\sum_u p^k(u)(1-p(u))^{n-k} \leq \binom{n}{k}\max_{q\in[0,1]} q^{k-1}(1-q)^{n-k} = \tag{20}$$
$$\binom{n}{k}\left(\frac{k-1}{n-1}\right)^{k-1}\left(1-\frac{k-1}{n-1}\right)^{n-k} = \left(\frac{n-1}{k-1}\right)\text{Bin}\left(k; n, \frac{k-1}{n-1}\right).$$

Therefore, the last term in (19) satisfies

$$\frac{(k+1)(k+2)}{(n-k)(n-k-1)}f_{n,k+2}(p) \leq \frac{(k+2)(n-1)}{(n-k)(n-k-1)}\text{Bin}\left(k+2;n,\frac{k+1}{n-1}\right). \tag{21}$$

Further, we study the first two terms of (19). Namely,

$$\frac{-(k+1)^2}{n(n-2k)(n-2k-1)}f^2_{n,k+1}(p) + \left(\frac{k+1}{n}\right)^2 f_{n,k+1}(p) \tag{22}$$

We notice that (22) is quadratic (and concave) in $f_{n,k+1}(p)$. Therefore, its maximum is obtained either on the local optimum, $(n-2k)(n-2k-1)/2n$, or on the boundary of $f_{n,k+1}(p)$. Therefore, we conclude that

- If $\frac{k(n-2k)(n-2k-1)}{2n(n-1)} \leq \text{Bin}\left(k+1;n,\frac{k}{n-1}\right)$ then,

$$\mathbb{E}_{X^n\sim p}\left(\hat{M}_k^{GT}-M_k\right)^2 \leq \tag{23}$$
$$\frac{(n-2k)(n-2k-1)(k+1)^2}{4n^3} + \frac{(k+2)(n-1)}{(n-k)(n-k-1)}\text{Bin}\left(k+2;n,\frac{k+1}{n-1}\right) + o\left(\frac{1}{n}\right).$$

- If $\frac{k(n-2k)(n-2k-1)}{2n(n-1)} > \text{Bin}\left(k+1;n,\frac{k}{n-1}\right)$ then,

$$\begin{aligned}\mathbb{E}_{X^n\sim p}\left(\hat{M}_k^{GT}-M_k\right)^2 \leq & \frac{-(k+1)^2(n-1)^2}{k^2n(n-2k)(n-2k-1)}\cdot\text{Bin}^2\left(k+1;n,\frac{k}{n-1}\right) + \\ & \left(\frac{k+1}{n}\right)^2\left(\frac{n-1}{k}\right)\cdot\text{Bin}\left(k+1;n,\frac{k}{n-1}\right) + \\ & \frac{(k+2)(n-1)}{(n-k)(n-k-1)}\text{Bin}\left(k+2;n,\frac{k+1}{n-1}\right) + o\left(\frac{1}{n}\right).\end{aligned} \tag{24}$$

It is well-known that a Binomial distribution $\text{Bin}(k;n,q)$ converges to a Poisson distribution Pois($k; \lambda = nq$) in cases where n grows and nq remains fixed. Specifically, Prokhorov showed that $|\text{Bin}(k; n, q) - \text{Pois}(k; nq)| \leq cq$ for a fixed constant c, and $k = 0, \ldots, n$ [2]. We apply Prokhorov result to the above and replace the Binomial terms with a Poisson distribution. Further we notice that as n grows, $f^*_{n,k+1} > f^{max}_{n,k+1}$. Therefore, we have that in this setup

$$R^*_n(\hat{M}^{GT}_k) \leq \frac{-(k+1)^2(n-1)^2}{k^2 n(n-2k)(n-2k-1)} \frac{\left(\frac{kn}{n-1}\right)^{2k+2} \exp\left(-2\frac{kn}{n-1}\right)}{\left((k+1)!\right)^2} + \left(\frac{k+1}{n}\right)^2 \left(\frac{n-1}{k}\right) \cdot \frac{\left(\frac{kn}{n-1}\right)^{k+1} \exp\left(-\frac{kn}{n-1}\right)}{(k+1)!} + \frac{(k+2)(n-1)}{(n-k)(n-k-1)} \frac{\left(\frac{(k+1)n}{n-1}\right)^{k+2} \exp\left(-\frac{(k+1)n}{n-1}\right)}{(k+2)!} + o\left(\frac{1}{n}\right). \tag{25}$$

Finally, we apply Sterling bounds, $\sqrt{2\pi} k^{k+1/2} \exp(-k) \leq k! \leq k^{k+1/2} \exp(-k+1)$ and conclude with the following theorem. For a fixed $k \geq 1$ and $n >> k$, the MSE of the GT estimator satisfies

$$R^*_n(\hat{M}^{GT}_k) \leq \frac{g(k)}{n} + o\left(\frac{1}{n}\right) \tag{26}$$

where

$$g(k) = \frac{e}{\sqrt{2\pi}} \Bigg(-\frac{\sqrt{2\pi}}{k+1} \left(\frac{k}{k+1}\right)^{2k} + \sqrt{k+1} \left(\frac{k}{k+1}\right)^{k} + \sqrt{k+2} \left(\frac{k+1}{k+2}\right)^{k+2} \Bigg). \tag{27}$$

Appendix B The MSE of GT for Large k

As shown in the main text, for $2k < n$ we have

$$\mathbb{E}\left(\hat{M}_k^{GT} - M_k\right)^2 = \tag{28}$$
$$\frac{1}{n^2}\binom{n}{k\ k}\sum_{u\neq v} P(u,v)\left(2k(2k+1) - n - 4k(p(u)+p(v)) + n^2(p(u)+p(v))^2\right) +$$
$$\frac{(k+1)^2}{n^2}\binom{n}{k+1}\sum_u p^{k+1}(u)(1-p(u))^{n-k-1} + \binom{n}{k}\sum_u p^{k+2}(u)(1-p(u))^{n-k}$$

where $P(u,v) = p^{k+1}(u)p^{k+1}(v)(1-p(u)-p(v))^{n-2k-2}$. Applying Propositions 1 and 2 from the main text we obtain

$$\mathbb{E}\left(\hat{M}_k^{GT} - M_k\right)^2 \le \frac{1}{n^2}\binom{n}{k\ k}\max_{q_1,q_2\in\Delta_2}\rho(q_1,q_2) + \frac{1}{n^2}\binom{n}{k}\max_{q\in[0,1]}\eta(q) \tag{29}$$

where

$$\rho(q_1,q_2) = q_1^k q_2^k (1-q_1-q_2)^{n-2k-2}\left(2k(2k+1) - n - 4kn(q_1+q_2) + n^2(q_1+q_2)^2\right) \tag{30}$$
$$\eta(q) = q^k(1-q)^{n-k-1}\left((n-k)(k+1) + n^2 q(1-q)\right). \tag{31}$$

We show that the first term in (30) depends on a single variable.

Proposition B.1.

$$\max_{q_1,q_2\in\Delta_2}\rho(q_1,q_2) = \max_{q\in[0,1/2]}\rho_1(q_1) \tag{32}$$

where $\rho_1(q_1) = \rho(q_1,q_1)$.

Proof. We first notice that $\max_{q_1,q_2\in\Delta_2}\rho(q_1,q_2)\geq 0$ since $\rho(0,0)=0$. Next, we show that for every pair $q_1,q_2\in\Delta_2$ such that $\rho(q_1,q_2)\geq 0$, we have $\rho(q_1,q_2)\leq\rho((q_1+q_2)/2,(q_1+q_2)/2)$. Therefore, we would like to show that

$$q_1^kq_2^k(1-q_1-q_2)^{n-2k-2}\big(2k(2k+1)-n-4kn(q_1+q_2)+n^2(q_1+q_2)^2\big)\leq \\ \left(\frac{q_1+q_2}{2}\right)^{2k}(1-q_1-q_2)^{n-2k-2}\big(2k(2k+1)-n-4kn(q_1+q_2)+n^2(q_1+q_2)^2\big) \tag{33}$$

for every q_1,q_2 such that $2k(2k+1)-n-4kn(q_1+q_2)+n^2(q_1+q_2)^2$ is non-negative. This inequality holds since $q_1^kq_2^k\leq\left(\frac{q_1+q_2}{2}\right)^{2k}$, as shown in the proof of Proposition C.1. $\square$

Plugging Proposition B.1 to (29) we obtain

$$\mathbb{E}\left(\hat{M}_k^{GT}-M_k\right)^2\leq\frac{1}{n^2}\binom{n}{k\ k}\max_{q_1\in[0,1/2]}\rho_1(q_1)+\frac{1}{n^2}\binom{n}{k}\max_{q\in[0,1]}\eta(q) \tag{34}$$

Let us characterize the maxima of $\rho_1(q_1)$ and $\eta(q)$. We begin with $\rho_1(q_1)$. Notice we have

$$\begin{aligned}\rho_1(q_1)=&q_1^{2k}(1-2q_1)^{n-2k-2}\big(2k(2k+1)-n-8knq_1+4n^2q_1^2\big)=\\ &q_1^{2k}(1-2q_1)^{n-2k-2}\big((2nq_1-2k)^2+2k-n\big)\leq q_1^{2k}(1-2q_1)^{n-2k-2}(2nq_1-2k)^2\end{aligned} \tag{35}$$

where the last inequality is due to $2k<n$. Let us characterize the maxima of this upper bound. We have

$$\frac{d}{dq_1}q_1^{2k}(1-2q_1)^{n-2k-2}(2nq_1-2k)^2= \\ q_1^{2k-1}(1-2q_1)^{n-2k-3}(2nq_1-2k)\left(-n^2q_1^2+(-2k+2kn+n)q_1-k^2\right)=0. \tag{36}$$

Therefore, the candidates for a maximum are obtained from a quadratic form, and

$$q_1^* = \frac{k(n-1)}{n^2} + \frac{1}{2n} \pm \frac{\sqrt{(n-2k)(n-2k+4kn)}}{2n^2}. \tag{37}$$

For $\eta(q)$ we have

$$\max_{q\in[0,1]} \eta(q) \le \max_{t_1\in[0,1]} \eta_1(t_1) + \max_{t_2\in[0,1]} \eta_2(t_2) \tag{38}$$

where

$$\begin{aligned} \eta_1(t_1) &= (n-k)(k+1)t_1^k(1-t_1)^{n-k-1} \\ \eta_2(t_2) &= n^2 t_2^{k+1}(1-t_2)^{n-k}. \end{aligned} \tag{39}$$

Simple calculus shows that $t_1^* = \frac{k}{n-1}$ and $t_2^* = \frac{k+1}{n+1}$. Notice that for sufficiently large n, the two maximizers are approximately equivalent. Putting together (34), (37) and the above, we obtain

$$\begin{aligned} \mathbb{E}\left(\hat{M}_k^{GT} - M_k\right)^2 \le & \frac{1}{n^2}\binom{n}{k\ k}(q_1^*)^{2k}(1-2q_1^*)^{n-2k-2}(2nq_1^* - 2k)^2 + \\ & \frac{1}{n^2}\binom{n}{k}\left((n-k)(k+1)(t_1^*)^k(1-t_1^*)^{n-k-1} + n^2(t_2^*)^{k+1}(1-t_2^*)^{n-k}\right) \end{aligned} \tag{40}$$

where q_1^*, t_1^* and t_2^* are defined in (37) and (39) respectively. We now derive a more compact form of our proposed bound. Notice we have

$$\frac{1}{n^2}\binom{n}{k\ k}(q_1^*)^{2k}(1-2q_1^*)^{n-2k-2}(2nq_1^*-2k)^2 = \tag{41}$$
$$\frac{1}{n^2}\binom{2k}{k}(2nq_1^*-2k)^2\left(\frac{1}{2}\right)^{2k}(1-2q_1^*)^{-2}\binom{n}{2k}(2q_1^*)^{2k}(1-2q_1^*)^{n-2k} =$$
$$\frac{1}{n^2}\binom{2k}{k}(2nq_1^*-2k)^2\left(\frac{1}{2}\right)^{2k}(1-2q_1^*)^{-2}\mathrm{Bin}(2k;n,2q_1^*).$$

Similarly,

$$\frac{1}{n^2}\binom{n}{k}\left((n-k)(k+1)(t_1^*)^k(1-t_1^*)^{n-k-1}+n^2(t_2^*)^{k+1}(1-t_2^*)^{n-k}\right) = \tag{42}$$
$$\frac{(n-k)(k+1)}{n^2}(1-t_1^*)^{-1}\mathrm{Bin}(k;n,t_1^*)+t_2^*\cdot\mathrm{Bin}(k;n,t_2^*)$$

where $\mathrm{Bin}(k;n,q)$ is a Binomial distribution with parameters n and q. Putting together (40) (41) and (42) we obtain

$$\mathbb{E}\left(\hat{M}_k^{GT}-M_k\right)^2 \leq \frac{1}{n^2}\binom{2k}{k}(2nq_1^*-2k)^2\left(\frac{1}{2}\right)^{2k}(1-2q_1^*)^{-2}\mathrm{Bin}(2k;n,2q_1^*)+ \tag{43}$$
$$\frac{(n-k)(k+1)}{n^2}(1-t_1^*)^{-1}\mathrm{Bin}(k;n,t_1^*)+t_2^*\cdot\mathrm{Bin}(k;n,t_2^*)$$

We now bound from above the Binomial terms, using Sterling bounds. Notice we have

$$\log\mathrm{Bin}(k;n,q) = \log\binom{n}{k}+k\log(q)+(n-k)\log(1-q) = \tag{44}$$
$$\log\binom{n}{rn}+n\left(r\log(q)+(1-r)\log(1-q)\right),$$

where $r = k/n$. The binomial coefficient satisfies

$$\log \binom{n}{rn} = \log n! - \log(rn)! - \log(n - rn)! \leq \tag{45}$$
$$-\log\sqrt{2\pi} + \frac{1}{12n} - \frac{1}{12rn+1} - \frac{1}{12(1-r)n+1} +$$
$$\left(n + \frac{1}{2}\right)\log(n) - \left(rn + \frac{1}{2}\right)\log(rn) - \left(n - rn + \frac{1}{2}\right)\log(n - rn) \leq$$
$$-\frac{1}{2}\log\left(2\pi nr(1-r)\right) + nH(r)$$

where the first inequality follows from Robbin's refined version of Sterling's bound [4],

$$\sqrt{2\pi}\ n^{n+\frac{1}{2}}e^{-n}e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi}\ n^{n+\frac{1}{2}}e^{-n}e^{\frac{1}{12n}},$$

the second inequality follows from

$$\frac{1}{12n} - \frac{1}{12rn+1} - \frac{1}{12(1-r)n+1} \leq \frac{1}{12n} - \frac{2}{6n+1} < 0$$

for $0 \leq r \leq 1$, and $H(r)$ is the binary entropy of r,

$$H(r) = -r\log(r) - (1-r)\log(1-r).$$

Therefore,

$$\log \mathrm{Bin}(rn; n, q) \leq -\frac{1}{2}\log\left(2\pi nr(1-r)\right) + nH(r) + n\left(r\log(q) + (1-r)\log(1-q)\right) =$$
$$-\frac{1}{2}\log\left(2\pi nr(1-r)\right) - nD_{KL}(r||q)$$

where $D_{KL}(r||q)$ is the Kullback-Leibler divergence,

$$D_{KL}(r||q) = r \log \frac{r}{q} + (1-r) \log \frac{(1-r)}{(1-q)}.$$

This means that

$$\text{Bin}(k; n, q) \le \frac{1}{\sqrt{2\pi k(1-k/n)}} \exp\left(-nD_{KL}\left(\frac{k}{n}\middle|\middle| q\right)\right). \tag{46}$$

Finally, we apply Sterling's bound $\binom{2k}{k} \le \frac{e}{\sqrt{2\pi}} \frac{2^{2k}}{\sqrt{k}}$ to (43), and plug (46), to obtain

$$\begin{aligned} R_n^*(\hat{M}_k^{GT}) \le & \frac{4(q_1^* - k/n)^2}{(2q_1^* - 1)^2 \sqrt{8\pi^3 e^{-2} k^2 (1 - 2k/n)}} \exp\left(-nD_{KL}(2k/n||2q_1^*)\right) + \\ & \frac{(1-k/n)(k/n + 1/n)}{(1-t_1^*)\sqrt{2\pi k(1-k/n)}} \exp\left(-nD_{KL}(k/n||t_1^*)\right) + \\ & \frac{t_2^*}{\sqrt{2\pi k(1-k/n)}} \exp\left(-nD_{KL}(k/n||t_2^*)\right). \end{aligned} \tag{47}$$

Let us further derive the above. First, we have that $\exp\left(-nD_{KL}(p||q)\right) \le \exp\left(-n(p-q)^2\right)$, following [1]. This allows us to quantify the order of the exponential terms and show that they are all $O(1)$. We now apply q_1^*,t_1^* and t_2^* to the different terms in (47), and conclude:

$$\frac{4(q_1^* - k/n)^2}{(2q_1^* - 1)^2 \sqrt{8\pi^3 e^{-2} k^2 (1 - 2k/n)}} \exp\left(-nD_{KL}(2k/n||2q_1^*)\right) = O\left(\frac{1}{n^2}\right). \tag{48}$$

Proceeding to the second and third terms in (47),

$$\frac{(1-k/n)(k/n+1/n)}{(1-t_1^*)\sqrt{2\pi k(1-k/n)}}+\frac{t_2^*}{\sqrt{2\pi k(1-k/n)}}= \tag{49}$$
$$\frac{1}{\sqrt{2\pi k(1-k/n)}}\left(\frac{(1-k/n)(k/n+1/n)}{(1-\frac{k}{n-1})}+\frac{k+1}{n+1}\right)\leq$$
$$\frac{1}{\sqrt{2\pi k(1-k/n)}}\left(\frac{(1-k/n)(k/n+1/n)}{(1-\frac{k}{n-1})}+\frac{k}{n}+\frac{1}{n}\right)=$$
$$\frac{k/n+1/n}{\sqrt{2\pi k(1-k/n)}}\left(\frac{(1-\frac{k}{n})}{(1-\frac{k}{n-1})}+1\right)\leq$$
$$\frac{k/n+1/n}{\sqrt{2\pi k(1-k/n)}}\left(2+\frac{k/n}{n-1-k}\right)=$$
$$\sqrt{\frac{2}{\pi}}\left(\frac{\sqrt{k}}{n\sqrt{1-k/n}}+\frac{1}{n\sqrt{k(1-k/n)}}\right)+O\left(\frac{k^{3/2}}{n^3}\right)=$$
$$\sqrt{\frac{2}{\pi}}\left(\frac{\sqrt{k}}{n\sqrt{1-k/n}}\right)+O\left(\frac{1}{\sqrt{k}n}\right).$$

Putting together (47,48,49) we conclude that for $2k<n$,

$$R_n^*(\hat{M}_k^{GT})\leq\sqrt{\frac{2}{\pi}}\left(\frac{\sqrt{k}}{n\sqrt{1-k/n}}\right)+o\left(\frac{1}{n}\right). \tag{50}$$

Notice that this bound is in fact $O\left(\frac{\sqrt{k}}{n}\right)$, as $\frac{1}{\sqrt{1-k/n}}\leq\sqrt{2}$.

Appendix C The MSE of the ML estimator

As shown in the main text, for $2k \leq n$, we have that

$$\mathbb{E}\left(\hat{M}_k^{ML} - M_k\right)^2 = \tag{51}$$
$$\binom{n}{k\ k} \sum_{u \neq v} \left(p(u) - \frac{k}{n}\right)\left(p(v) - \frac{k}{n}\right) p^k(u) p^k(v) (1 - p(u) - p(v))^{n-2k} +$$
$$\binom{n}{k} \sum_{u} \left(p(u) - \frac{k}{n}\right)^2 p^k(u) (1 - p(u))^{n-k}.$$

Applying Propositions 1 and 2 from the main text to the above yields

$$\mathbb{E}\left(\hat{M}_k^{ML} - M_k\right)^2 \leq \binom{n}{k\ k} \max_{q_1, q_2 \in \Delta_2} \psi(q_1, q_2) + \binom{n}{k} \max_{q \in [0,1]} \phi(q) \tag{52}$$

where

$$\psi(q_1, q_2) = \left(q_1 - \frac{k}{n}\right)\left(q_2 - \frac{k}{n}\right) q_1^{k-1} q_2^{k-1} (1 - q_1 - q_2)^{n-2k} \tag{53}$$

$$\phi(q) = \left(q - \frac{k}{n}\right)^2 q^{k-1} (1 - q)^{n-k} \tag{54}$$

Let us first focus on the first term in (52).

Proposition C.1.

$$\max_{q_1, q_2 \in \Delta_2} \psi(q_1, q_2) = \max_{q_1 \in [0,1/2]} \psi_2(q_2) \tag{55}$$

where $\psi_2(q_2) = \psi(q_2, q_2)$

Proof. We study $\psi(q_1, q_2)$ for different possible pairs of $q_1, q_2 \in \Delta_2$.

- For $q_1 \le \frac{k}{n}$ and $q_2 \ge \frac{k}{n}$ (and vice versa):
  We have that $\psi(q_1, q_2) < 0$, while $\psi(\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2}) \ge 0$. Therefore, $\psi(q_1, q_2) \le \psi(\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2})$.

- For $q_1, q_2 \ge \frac{k}{n}$:
  We would like to show that $\psi(q_1, q_2) \le \psi(\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2})$. Plugging (53), we require that

$$\left(q_1 - \frac{k}{n}\right)\left(q_2 - \frac{k}{n}\right) q_1^{k-1} q_2^{k-1} \le \left(\frac{q_1+q_2}{2} - \frac{k}{n}\right)^2 \left(\frac{q_1+q_2}{2}\right)^{2k-2}$$

  Since both sides of the inequality are positive, this is equivalent to

$$\begin{aligned} &\log\left(q_1 - \frac{k}{n}\right) + \log\left(q_2 - \frac{k}{n}\right) + (k-1)\left(\log(q_1) + \log(q_2)\right) \le \\ &2\log\left(\frac{q_1+q_2}{2} - \frac{k}{n}\right) + (2k-2)\log\left(\frac{q_1+q_2}{2}\right). \end{aligned} \tag{56}$$

  Due to the concavity of the log function, we have

$$\begin{aligned} &\frac{1}{2}\log\left(q_1 - \frac{k}{n}\right) + \frac{1}{2}\log\left(q_2 - \frac{k}{n}\right) \le \log\left(\frac{q_1+q_2}{2} - \frac{k}{n}\right) \\ &\frac{1}{2}\log(q_1) + \frac{1}{2}\log(q_2) \le \log\left(\frac{q_1+q_2}{2}\right) \end{aligned} \tag{57}$$

  which justifies (56).

- For $q_1, q_2 \le \frac{k}{n}$:
  Again, we would like to show that $\psi(q_1, q_2) \le \psi(\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2})$. First, we show in (56) that $q_1^{k-1} q_2^{k-1} \le \left(\frac{q_1+q_2}{2}\right)^{2k-2}$. Therefore, we still need to show that $\left(q_1 - \frac{k}{n}\right)\left(q_2 - \frac{k}{n}\right) \le$

$\left(\frac{q_1+q_2}{2} - \frac{k}{n}\right)^2$. However, we have that

$$\left(\frac{q_1+q_2}{2} - \frac{k}{n}\right)^2 - \left(q_1 - \frac{k}{n}\right)\left(q_2 - \frac{k}{n}\right) = \left(\frac{q_1-q_2}{2}\right)^2 \geq 0 \tag{58}$$

which concludes the proof.

□

Plugging Proposition C.1 to (52), we have that

$$\mathbb{E}\left(\hat{M}_k^{ML} - M_k\right)^2 \leq \binom{n}{k\ k} \max_{q_2\in[0,1/2]} \psi_2(q_2) + \binom{n}{k} \max_{q\in[0,1]} \phi(q). \tag{59}$$

Let us now characterize the maxima of $\psi_2(q_2)$ and $\phi(q)$. We begin with $\psi_2(q_2)$. Taking the derivative, we obtain

$$\begin{aligned}&\frac{d}{dq_2}\left(q_2 - \frac{k}{n}\right)^2 q_2^{2k-2}(1-2q_2)^{n-2k} = \\ &q_2^{2k-3}(1-2q_2)^{n-2k-1}\left(q_2 - \frac{k}{n}\right)\left(-2nq_2^2 + \left(4k - 4\frac{k}{n}\right)q_2 + \frac{2k}{n}(1-k)\right) = 0.\end{aligned} \tag{60}$$

Therefore, the candidates for a maximum, for $k > 1$, are obtained from a quadratic function,

$$q_2^* = \frac{k}{n} - \frac{k}{n^2} \pm \frac{\sqrt{k\left(1 + \frac{k}{n^2} - \frac{2k}{n}\right)}}{n}. \tag{61}$$

For $\phi(q)$ we have

$$\frac{d}{dq}\left(q-\frac{k}{n}\right)^2 q^{k-1}(1-q)^{n-k} = \tag{62}$$
$$q^{k-2}(1-q)^{n-k-1}\left(q-\frac{k}{n}\right)\left((-n-1)q^2+\left(1+2k-\frac{k}{n}\right)q+\frac{k}{n}(1-k)\right)=0.$$

Once again, the candidates for a maximum, for $k>1$, are obtained from a quadratic function, and

$$q^* = \frac{k}{n}\left(\frac{2n-1}{2n+2}\right)+\frac{1}{2n+2}\pm\frac{\sqrt{\left(1+\frac{k}{n}\right)^2+8k\left(1-\frac{k}{n}\right)}}{2n+2}. \tag{63}$$

We now derive a more compact form for our proposed bound. Notice we have

$$\mathbb{E}\left(\hat{M}_k^{ML}-M_k\right)^2 \le \tag{64}$$
$$\binom{n}{k\ k}\left(q_2^*-\frac{k}{n}\right)^2(q_2^*)^{2k-2}(1-2q_2^*)^{n-2k}+\binom{n}{k}\left(q^*-\frac{k}{n}\right)^2(q^*)^{k-1}(1-q^*)^{n-k}=$$
$$\binom{2k}{k}\left(q_2^*-\frac{k}{n}\right)^2\left(\frac{1}{2}\right)^{2k}\left(\frac{1}{q_2^*}\right)^2\text{Bin}(2k;n,2q_2^*)+\left(q^*-\frac{k}{n}\right)^2\left(\frac{1}{q^*}\right)\text{Bin}(k;n,q^*).$$

where $\text{Bin}(k;n,q)$ is a Binomial distribution with parameters n and q.

Finally, we apply (46) and conclude that

$$R_n^*(\hat{M}_k^{ML}) \le \frac{\left(1-\frac{k/n}{q_2^*}\right)^2}{\sqrt{8\pi^3e^{-2}k^2(1-2k/n)}}\exp\left(-nD_{KL}\left(\frac{2k}{n}\middle\|2q_2^*\right)\right)+ \tag{65}$$
$$\frac{\left(q^*-\frac{k}{n}\right)^2}{q^*\sqrt{2\pi k(1-k/n)}}\exp\left(-nD_{KL}\left(\frac{k}{n}\middle\|q^*\right)\right)$$

References

[1] Amichai Painsky and Gregory Wornell. On the universality of the logistic loss function. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 936–940. IEEE, 2018.

[2] Yurii Vasil'evich Prokhorov. Asymptotic behavior of the binomial distribution. *Uspekhi Matematicheskikh Nauk*, 8(3):135–142, 1953.

[3] Nikhilesh Rajaraman, Andrew Thangaraj, and Ananda Theertha Suresh. Minimax risk for missing mass estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 3025–3029. IEEE, 2017.

[4] Herbert Robbins. A remark on stirling's formula. *The American mathematical monthly*, 62(1):26–29, 1955.