Appendix A: A Proof for Theorem 1

For the purpose of this proof, we require the following property:

**Property 1** Let $p$ be a probability distribution over a countable alphabet $\mathcal{X}$ and $n > 1$. Then,

$$
\mathbb{E}_{X^n \sim p} (M(X^n) \Phi_1(X^n)) \leq \frac{1}{n} \mathbb{E}_{X^n \sim p} (\Phi_1^2(X^n))
$$

(1)

with equality iff $p$ is a degenerate distribution.

**Proof.** We have

$$
\mathbb{E}_{X^n \sim p} \left( \frac{1}{n} \Phi_1(X^n) \left( \frac{1}{n} \Phi_1(X^n) - M(X^n) \right) \right) = \frac{1}{n^2} \sum_{u,v} \mathbb{E}_{X^n \sim p} (\mathbbm{1}(N_u = 1) \mathbbm{1}(N_v = 1)) - \frac{1}{n} \sum_{u,v} p(v) \mathbb{E}_{X^n \sim p} (\mathbbm{1}(N_u = 1) \mathbbm{1}(N_v = 0)) = 
$$

$$
\sum_{u,v} \left( \frac{1}{n^2} P_n(1, 1) - \frac{1}{n} p(u) P_n(0, 1) \right) = 
$$

$$
\sum_{u \neq v} \left( \frac{n-1}{n} p(u)p(v)(1 - p(u) - p(v))^{n-2} - p(u)p(v)(1 - p(u) - p(v))^{n-1} \right) + 
$$

$$
\frac{1}{n} \sum_u p(u)(1 - p(u))^{n-1} \geq 
$$

$$
-\frac{1}{n} \sum_{u \neq v} p(u)p(v)(1 - p(u) - p(v))^{n-1} + \frac{1}{n} \sum_u p(u)(1 - p(u))^{n-1} \geq 
$$

$$
-\frac{1}{n} \sum_{u \neq v} p(u)p(v)(1 - p(u))^{n-1} + \frac{1}{n} \sum_u p(u)(1 - p(u))^{n-1} = 
$$

$$
\frac{1}{n} \sum_u p(u)(1 - p(u))^{n-1} \left( 1 - \sum_{v \neq u} p(v) \right) = \frac{1}{n} \sum_u p^2(u)(1 - p(u))^{n-1} \geq 0,
$$

where $P_n(i, j) = \mathbb{E} (\mathbbm{1}(N_u = i) \mathbbm{1}(N_v = j))$ follows (15) in the main text, and the inequalities follow from $(1 - p(u) - p(v))^{n-2} \geq (1 - p(u) - p(v))^{n-1}$ and $(1 - p(u) - p(v))^{n-1} \leq (1 - p(u))^{n-1}$, for all $0 \leq (1 - p(u) - p(v)) \leq 1$. Notice that equality holds if and only if $p$ is a degenerate distribution. □
Our missing mass minimax problem is defined as

\[ R^*_n(M_{\beta_1}, \Delta) = \inf_{\beta_1 \in \mathbb{R}} \sup_{p \in \Delta} \mathbb{E}_{X^n \sim p} (\beta_1 \Phi_1(X^n) - M(X^n))^2. \]  

(3)

We first observe that \( M(X^n) \in [0, 1] \), while \( \Phi_1(X^n) \in [0, n] \). This means that for an unbounded \( \beta_1 \) we attain an unbounded risk, which is not the minimax risk (for example, \( \beta_1 = 0 \) attains a lower risk). Therefore, we may restrict our attention to \( \beta_1 \in \mathcal{B} \) where \( \mathcal{B} \) is a bounded set (for example \( \mathcal{B} = [0, C] \), for some finite constant \( C \)). Now, let us first reformulate (3) as

\[
\min_{\beta_1 \in \mathcal{B}} \sup_{w \in \mathcal{W}} \int_{p \in \Delta} w(p) \mathbb{E}_{X^n \sim p} (\beta_1 \Phi_1(X^n) - M(X^n))^2 dp,
\]

(4)

where \( \mathcal{W} = \{ w(p) | \int_{p \in \Delta} w(p) dp = 1, \ w(p) \geq 0 \text{ for all } p \in \Delta \} \). In words, \( w(p) \) is a non-negative weight which corresponds to the distribution \( p \) in \( \Delta \). Notice that the equivalence above holds since for a fixed \( \beta_1 \), a weight function \( w(p) \) that puts all probability mass on the worst \( p \in \Delta \) is equivalent to supremum over \( p \in \Delta \). A similar derivation appears, for example, in [1].

Next, we require Sion’s minimax theorem [2], stated as follows.

**Theorem 1 (Sion’s Minimax Theorem [2])** Let \( \mathcal{U} \) be a compact convex subset of a linear topological space and \( \mathcal{V} \) be a convex subset of a linear topological space. If \( f(u, v) \) is a real-valued function on \( \mathcal{U} \times \mathcal{V} \) with

1. \( f(u, \cdot) \) is upper semi-continuous and quasi-concave on \( \mathcal{V} \) for all \( u \in \mathcal{U} \)
2. \( f(\cdot, v) \) is lower semi-continuous and quasi-convex on \( \mathcal{U} \) for all \( v \in \mathcal{V} \)

then, \( \min_{u \in \mathcal{U}} \sup_{v \in \mathcal{V}} f(u, v) = \sup_{v \in \mathcal{V}} \min_{u \in \mathcal{U}} f(u, v) \).

It is immediate to show that our objective (the integral in (4)), is continuous in \( w(p) \) and \( \beta_1 \), linear in \( w(p) \) and convex in \( \beta_1 \). Further, we observe that \( \mathcal{B} \) is a closed interval (compact, convex and a subset of a linear topological space) a while \( \mathcal{W} \) is a set of all probability distributions (convex and a subset of a linear topological space). Therefore, we may apply Sion’s minimax theorem and conclude that (4) is equivalent to

\[
\sup_{w \in \mathcal{W}} \min_{\beta_1 \in \mathcal{B}} \int_{p \in \Delta} w(p) \mathbb{E}_{X^n \sim p} (\beta_1 \Phi_1(X^n) - M(X^n))^2 dp
\]

(5)

Let us now focus on the minimization over \( \beta_1 \). Notice that this minimization is in fact a weighted mean square error problem. Specifically, \( \beta^*_1(w) \) is a weighted minimal mean squared error estimator, for estimating a random variable (the missing mass) from a random variable (the number of symbols with a single appearance). Simple calculus shows that for \( \Sigma_\Phi > 0 \),

\[
\beta^*_1(w) = \frac{\Sigma_M}{\Sigma_\Phi}
\]

(6)
and (4) equals

$$\sup_{w \in \mathcal{W}} \Sigma_M - \Sigma_{M\Phi}/\Sigma_{\Phi}$$

(7)

where

- $$\Sigma_M = \int_{p \in \Delta} w(p) E_{X^n \sim p} M^2(X^n) dp$$
- $$\Sigma_{M\Phi} = \int_{p \in \Delta} w(p) E_{X^n \sim p} (M(X^n) \Phi_1(X^n)) dp$$
- $$\Sigma_{\Phi} = \int_{p \in \Delta} w(p) E_{X^n \sim p} (\Phi_2^2(X^n)) dp$$.

We observe that for $$n = 1$$, $$\Sigma_{\Phi} = 0$$ iff $$w(p)$$ puts all the mass on a degenerate distribution. Such a weight function cannot attain the solution to (3), since the degenerate distribution is not the supremum of the objective ($$M(X^n) = \Phi_1(X^n) = 0$$ for all $$X^n \sim p$$). Therefore, we rule out the case where $$\Sigma_{\Phi} = 0$$.

Going back to Property 1 we have that for $$n > 1$$,

$$E_{X^n \sim p} (M(X^n) \Phi_1(X^n)) \leq \frac{1}{n} E_{X^n \sim p} (\Phi_2^2(X^n))$$

(8)

with equality iff $$p$$ is a degenerate distribution (resulting with both sides of the inequality equal zero). Averaging both sides over a weight measure $$w(p)$$ we obtain,

$$\Sigma_{M\Phi} = \int_{p \in \Delta} w(p) E_{X^n \sim p} (M(X^n) \Phi_1(X^n)) dp < \frac{1}{n} \int_{p \in \Delta} w(p) E_{X^n \sim p} (\Phi_2^2(X^n)) dp = \frac{1}{n} \Sigma_{\Phi},$$

(9)

for every $$w(p)$$ that does not puts all the mass on a degenerate distribution. On the other hand, such a choice of $$w(p)$$ cannot be the minimax solution of (3) for $$n > 1$$, as discussed above. Plugging the above to (6) yields the desired result.

Appendix B: A Proof for Theorem 2

We first introduce an important property that is used throughout our analysis.

**Proposition 1** Let $$p$$ be a probability distribution over a countable alphabet $$\mathcal{X}$$, $$r \geq 1$$ be a positive integer and $$n \in \mathbb{N}_+$$. Then the following holds,

$$\sum_{u \in \mathcal{X}} p^r(u) e^{-np(u)} \leq \frac{(r - 1)!}{n^{r-1}}.$$

(10)

**Proof.** Let $$X \sim p$$ and define a random variable $$T(x) = \frac{(np(x))^{r-1} e^{-np(x)}}{(r-1)!}.$$ Notice that $$T(x)$$ is a Poisson distribution, $$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$ with a parameter $$\lambda = np(x)$$ and $$k = r - 1$$. Therefore, $$T(x) \in [0, 1].$$
The expected value of $T(x)$ satisfies
\[
\mathbb{E}T(X) = \sum_{x \in X} p(x) \frac{(np(x))^{r-1} e^{-np(x)}}{(r-1)!} = \frac{n^{r-1}}{(r-1)!} \sum_{x \in X} p^r(x) e^{-np(x)} \leq 1
\] (11)
where the inequality follows from $T(x) \in [0, 1]$.

Let us now derive the bound for $\hat{M}_{\beta_1, \beta_2}(X^n)$. Plugging the definition of $P_n(i, j)$ (see (15) in the main text) to the estimation risk ((14) in the main text), we obtain
\[
\mathbb{E}_{X^n \sim p} \left( \hat{M}_{\beta_1, \beta_2}(X^n) - M(X^n) \right)^2 =
\left( \beta_1^2 n(n-1) - 2\beta_1 (1 - p(u) - p(v)) + (1 - p(u) - p(v))^2 \right) \sum_{u \neq v} p(u)p(v)(1 - p(u) - p(v))^{n-2} +
2\beta_1 \beta_2 \left( \frac{n}{2} \right) \sum_{u \neq v} p^2(u)p(v)(1 - p(u) - p(v))^{n-3} +
\beta_2 \left( \frac{n}{2} \right) \sum_{u \neq v} p^2(u)p^2(v)(1 - p(u) - p(v))^{n-4} -
2\beta_2 \left( \frac{n}{2} \right) \sum_{u \neq v} p^2(u)p(v)(1 - p(u) - p(v))^{n-2} +
n\beta_1^2 \sum_{u} p(u)(1 - p(u))^{n-1} + \sum_{u \neq v} p^2(u)(1 - p(u))^{n} + \beta_2 \left( \frac{n}{2} \right) \sum_{u \neq v} p^2(u)(1 - p(u))^{n-2}.
\] (12)

Let separately study each of the terms in (12). Define
\[
g(\beta_1, n) = \beta_1^2 n(n-1) - 2\beta_1 (1 - p(u) - p(v)) + (1 - p(u) - p(v))^2.
\]

Then, the first term in (12) is given by $g(\beta_1, n) \sum_{u \neq v} p(u)p(v)(1 - p(u) - p(v))^{n-2}$.

For $g(\beta_1, n) \geq 0$ we have
\[
g(\beta_1, n) \sum_{u \neq v} p(u)p(v)(1 - p(u) - p(v))^{n-2} \leq g(\beta_1, n) \sum_{u \neq v} p(u)p(v)e^{-(n-2)(p(u)+p(v))},
\] (13)
where the inequality follows from
\[
(1 - t)^n \leq e^{-nt}
\] (14)
for any $t \in \mathbb{R}$ and $n \in \mathbb{R}_+$ [3].
Alternatively, for $g(\beta_1, n) < 0$ we have

\[
g(\beta_1, n) \sum_{u \neq v} p(u)p(v)(1 - p(u) - p(v))^{n-2} \leq (15)
\]

\[
g(\beta_1, n) \sum_{u \neq v} p(u)p(v)e^{-(n-2)(p(u)+p(v))} - (n-2)g(\beta_1, n) \sum_{u \neq v} p(u)p(v)(p(u) + p(v))^2e^{-(n-2)(p(u)+p(v))},
\]

where the inequality follows from

\[
(1 - t)^n \geq e^{-nt}(1 - nt^2)
\]

for any $0 \leq t \leq 1$ and $n \in \mathbb{N}_+$ [3, 4].

Let us focus on the second term of (15). First, we notice that

\[
(n-2) \sum_{u \neq v} p(u)p(v)(p(u) + p(v))^2e^{-(n-2)(p(u)+p(v))} \leq (17)
\]

\[
(n-2) \sum_{u,v} p(u)p(v)(p(u) + p(v))^2e^{-(n-2)(p(u)+p(v))} =
\]

\[
2(n-2) \sum_{u,v} (p^3(u)p(v) + p^2(u)^2p(v))e^{-(n-2)(p(u)+p(v))} \leq \frac{6}{n-2},
\]

where the last inequality follows from Proposition 1. Second, we introduce the following property for $g(\beta_1, n)$.

**Proposition 2** $g(\beta_1, n) > -\frac{1}{n-1}$.

**Proof.** We observe that $g(\beta_1, n)$ is convex in $\beta_1$ and its minimum is attained at $\beta_1 = \frac{1-p(u)-p(v)}{n-1}$. Therefore, $g(\beta_1, n) \geq -\frac{(1-p(u)-p(v))^2}{n-1} \geq -\frac{1}{n-1}$. ■

Applying (17) and Proposition 2 to (15) yields

\[
g(\beta_1, n) \sum_{u \neq v} p(u)p(v)(1 - p(u) - p(v))^{n-2} \leq g(\beta_1, n) \sum_{u \neq v} p(u)p(v)e^{-(n-2)(p(u)+p(v))} + o\left(\frac{1}{n}\right). \quad (18)
\]

Next, we study the second and third terms in (12). We observe that both coefficients, $2\beta_1\beta_2\binom{n}{1,2}$, and $\beta_2^2\binom{n}{2,2}$ are non-negative. Therefore,

\[
2\beta_1\beta_2\binom{n}{1,2} \sum_{u \neq v} p^2(u)p(v)(1 - p(u) - p(v))^{n-3} \leq 2\beta_1\beta_2\binom{n}{1,2} \sum_{u \neq v} p^2(u)p(v)e^{-(n-3)(p(u)+p(v))} \quad (19)
\]

\[
\beta_2^2\binom{n}{2,2} \sum_{u \neq v} p^2(u)p^2(v)(1 - p(u) - p(v))^{n-4} \leq \beta_2^2\binom{n}{2,2} \sum_{u \neq v} p^2(u)p^2(v)e^{-(n-4)(p(u)+p(v))} \quad (20)
\]
following (14). Proceeding to the forth term in (12), we observe

$$-2\beta_2 \left( \frac{n}{2} \right) \sum_{u \neq v} p^2(u) p(v) (1 - p(u) - p(v))^{n-2} \leq$$

$$-2\beta_2 \left( \frac{n}{2} \right) \left( \sum_{u \neq v} p^2(u) p(v) e^{-(n-2)(p(u)+p(v))} - (n-2) \sum_{u \neq v} p^2(u) p(v) (p(u) + p(v))^2 e^{-(n-2)(p(u)+p(v))} \right)$$

following (16). Looking at the second term in (21), we have

$$2\beta_2 \left( \frac{n}{2} \right) (n-2) \sum_{u \neq v} p^2(u) p(v) (p(u) + p(v))^2 e^{-(n-2)(p(u)+p(v))} \leq$$

$$2\beta_2 \left( \frac{n}{2} \right) (n-2) \sum_{u \neq v} p^2(u) p(v) (p(u) + p(v))^2 e^{-(n-2)(p(u)+p(v))} \leq$$

$$2\beta_2 \left( \frac{n}{2} \right) (n-2) \left( \sum_{u,v} p^4(u) p(v) e^{-(n-2)(p(u)+p(v))} + 3 \sum_{u,v} p^3(u) p^2(v) e^{-(n-2)(p(u)+p(v))} \right) \leq o \left( \frac{1}{n} \right),$$

where the last inequality follows from Proposition 1 and $\beta_2 \leq O \left( \frac{1}{n^2} \right)$ for some $c > 1$. Finally, looking at the three last elements in (12), we observe they all have positive coefficients. Therefore,

$$n\beta_1^2 \sum p(u)(1-p(u))^{n-1} + \sum p^2(u)(1-p(u))^n + \beta_2^2 \left( \frac{n}{2} \right) \sum p^2(u)(1-p(u))^{n-2} \leq$$

$$n\beta_1^2 \sum p(u) e^{-(n-1)p(u)} + \sum p^2(u) e^{-np(u)} + \beta_2^2 \left( \frac{n}{2} \right) \sum p^2(u) e^{-(n-2)p(u)},$$

following (14). Putting together (18)-(23) we obtain

$$\mathbb{E}_{X^n \sim p} \left( M_{\beta_1,\beta_2} (X^n) - M(X^n) \right)^2 \leq$$

$$g(\beta_1, n) \sum_{u \neq v} p(u) p(v) e^{-(n-2)(p(u)+p(v))} + 2\beta_1 \beta_2 \left( \frac{n}{12} \right) \sum_{u \neq v} p^2(u) p(v) e^{-(n-3)(p(u)+p(v))} +$$

$$\beta_2^2 \left( \frac{n}{2} \right) \sum_{u \neq v} p^2(u) p^2(v) e^{-(n-4)(p(u)+p(v))} - 2\beta_2 \left( \frac{n}{2} \right) \sum_{u \neq v} p^2(u) p(v) e^{-(n-2)(p(u)+p(v))} +$$

$$n\beta_1^2 \sum p(u) e^{-(n-1)p(u)} + \sum p^2(u) e^{-np(u)} + \beta_2^2 \left( \frac{n}{2} \right) \sum p^2(u) e^{-(n-2)p(u)} + o \left( \frac{1}{n} \right).$$

Next, we would like the double summation terms to consider all possible $u,v$ (as opposed to $u \neq v$ in the current form). Therefore, we require the following proposition
Proposition 3  Let

\[ \begin{align*}
  t(p, \beta_1, \beta_2, n) &= g(\beta_1, n) \sum_{u=v} p(u) p(v) e^{-(n-2)(p(u)+p(v))} + 2\beta_1 \beta_2 \left( \frac{n}{2} \right) \sum_{u=v} p^2(u) p(v) e^{-(n-3)(p(u)+p(v))} + \\
  &+ \beta_2^2 \left( \frac{n}{2} \right) \sum_{u=v} p^3(u) e^{-(n-4)(p(u)+p(v))} - 2\beta_2 \left( \frac{n}{2} \right) \sum_{u=v} p^2(u) p(v) e^{-(n-2)(p(u)+p(v))}.
\end{align*} \]  

Then, \( t(p, \beta_1, \beta_2, n) + o \left( \frac{1}{n} \right) \geq 0 \).

Proof. Let us first notice that

\[ g(\beta_1, n) \bigg|_{u=v} = \beta_1^2 n(n-1) - 2n\beta_1 + 1 + 4p(u)(n\beta_1 - 1) + 4p^2(u) = \]

\[ (\beta_1 n - 1 + 2p(u))^2 - n\beta_1^2. \]

Now, we have that

\[ 0 \leq \left( \beta_1 n - 1 + 2p(u) \right) e^{-2p(u)} + \beta_2 \left( \frac{n}{2} \right) p(u) \sum_{u} p^2(u) e^{-2(n-4)p(u)} = \]

\[ \left( \beta_1 n - 1 + 2p(u) \right)^2 \sum_{u} p^2(u) e^{-2(n-2)p(u)} + 2\beta_1 \beta_2 n \left( \frac{n}{2} \right) \sum_{u} p^3(u) e^{-2(n-3)p(u)} - \\
  2\beta_2 \left( \frac{n}{2} \right) \sum_{u} p^3(u) e^{-2(n-3)p(u)} + 4\beta_2 \left( \frac{n}{2} \right) \sum_{u} p^4(u) e^{-2(n-3)p(u)} + \beta_2^2 \left( \frac{n}{2} \right)^2 \sum_{u} p^4(u) e^{-2(n-4)p(u)}. \]

Following Proposition 1, we notice that \( \sum_{u} p^4(u) e^{-2(n-3)p(u)} \leq \frac{6}{(n-3)^3} \). Therefore,

\[ 4\beta_2 \left( \frac{n}{2} \right) \sum_{u} p^4(u) e^{-2(n-3)p(u)} \leq o \left( \frac{1}{n} \right). \]  

(27)

In addition, we have that

\[ -2\beta_2 \left( \frac{n}{2} \right) \sum_{u} p^3(u) e^{-2(n-3)p(u)} \leq -2\beta_2 \left( \frac{n}{2} \right) \sum_{u} p^3(u) e^{-2(n-2)p(u)}. \]  

(29)

Plugging (26), (28) and (29) to (27) we obtain,

\[ 0 \leq t(p, \beta_1, \beta_2, n) + n\beta_1^2 \sum_{u} p^2(u) e^{-2(n-2)p(u)} + \\
  2\beta_1 \beta_2 \left( \frac{n}{2} - \left( \frac{n}{1} \right) \right) \sum_{u} p^3(u) e^{-2(n-3)p(u)} + \\
  \beta_2^2 \left( \frac{n}{2} \right)^2 - \left( \frac{n}{2} \right) \sum_{u} p^4(u) e^{-2(n-4)p(u)} + o \left( \frac{1}{n} \right). \]  

(30)
Finally, we have that that \( n\binom{n}{2} - \binom{n}{1} = 2\binom{n}{2} \) and \( \binom{n}{1}^2 - \binom{n}{2} = (2n - 3)\binom{n}{2} \). Plugging the above and Proposition 1 to (30) leads to

\[
0 \leq t(p, \beta_1, \beta_2, n) + o\left(\frac{1}{n}\right),
\]

as desired. ■

Applying Proposition 3 to (24), we obtain

\[
\begin{align*}
E_{X_n \sim p} \left( \bar{M}_{\beta_1, \beta_2}(X^n) - M(X^n) \right)^2 & \leq \\
g(\beta_1, n) \sum_{u,v} p(u)p(v)e^{-(n-2)(p(u)+p(v))} + 2\beta_1\beta_2 \binom{n}{1,2} \sum_{u,v} p^2(u)p(v)e^{-(n-3)(p(u)+p(v))} + \\
& \beta_2^2 \binom{n}{2} \sum_{u,v} p^2(u)p(v)e^{-(n-4)(p(u)+p(v))} - 2\beta_2 \binom{n}{2} \sum_{u,v} p^2(u)p(v)e^{-(n-2)(p(u)+p(v))} + \\
& n\beta_1^2 \sum_{u,v} p(u)e^{-(n-1)p(u)} + \sum_{u,v} p^2(u)e^{-np(u)} + \beta_2^2 \binom{n}{2} \sum_{u,v} p^2(u)e^{-(n-2)p(u)} + o\left(\frac{1}{n}\right) = \\
(n(n-1)\beta_1^2 - 2n\beta_1 + 1) \sum_{u,v} p(u)p(v)e^{-(n-2)(p(u)+p(v))} + 4(\beta_1 n - 1) \sum_{u,v} p^2(u)p(v)e^{-(n-2)(p(u)+p(v))} + \\
2\beta_1\beta_2 \binom{n}{1,2} \sum_{u,v} p^2(u)p(v)e^{-(n-3)(p(u)+p(v))} + \beta_2^2 \binom{n}{2} \sum_{u,v} p^2(u)p(v)e^{-(n-4)(p(u)+p(v))} - \\
2\beta_2 \binom{n}{2} \sum_{u,v} p^2(u)p(v)e^{-(n-2)(p(u)+p(v))} + n\beta_1^2 \sum_{u,v} p(u)e^{-(n-1)p(u)} + \sum_{u,v} p^2(u)e^{-np(u)} + \\
\beta_2^2 \binom{n}{2} \sum_{u,v} p^2(u)e^{-(n-2)p(u)} + o\left(\frac{1}{n}\right),
\end{align*}
\]

where the equality follows from

\[
g(\beta_1, n) = \beta_1^2 n(n-1) - 2n\beta_1 + 1 + (p(u) + p(v))(2n\beta_1 - 2) + (p(u) + p(v))^2
\]

and \( \sum_{u,v} p(u)p(v)(p(u) + p(v))^2e^{-(n-2)(p(u)+p(v))} \leq o\left(\frac{1}{n}\right) \) (by Proposition 1). Define

\[
f_{n,r} = \sum_u p_r(u)e^{-np(u)}
\]

Then, (32) bounded from above by

\[
\begin{align*}
E_{X_n \sim p} \left( \bar{M}_{\beta_1, \beta_2}(X^n) - M(X^n) \right)^2 & \leq \\
f_{n-2,1}^2 \binom{n(n-1)\beta_1^2 - 2n\beta_1 + 1}{2} + f_{n-2,1,f_{n-2,2}} \binom{4(\beta_1 n - 1) - 2\beta_2 \binom{n}{2}}{2} + \\
f_{n-3,1}f_{n-3,2} \binom{2\beta_1\beta_2 \binom{n}{1,2}}{2} + f_{n-4,2}^2 \binom{\beta_2^2 \binom{n}{2}}{2} + f_{n-1,1} \binom{n\beta_1^2}{2} + f_{n,2} + f_{n-2,2} \binom{\beta_2^2 \binom{n}{2}}{2} + o\left(\frac{1}{n}\right).
\end{align*}
\]
Finally, we observe that

\[ 0 \leq f_{n,r} - f_{n+m,r} = \sum_u p^r(u)e^{-np(u)} - \sum_u p^r(u)e^{-(n+m)p(u)} = \sum_u p^r(u)e^{-np(u)} \left( 1 - e^{-mp(u)} \right) \leq m \sum_u p^{r+1}(u)e^{-np(u)} \leq m \frac{r!}{n^r}, \]

where the second inequality follows from \( e^x \geq 1 + x \) for all \( x \in \mathbb{R} \), and the last inequality follows from Proposition 1. Applying the above, we have

\[ f_{n-3,1}f_{n-3,2} \leq \left( f_{n-2,1} + \frac{1}{n-3} \right) \left( f_{n-2,2} + \frac{2}{(n-3)^2} \right) \leq f_{n-2,1}f_{n-2,2} + O \left( \frac{1}{n^2} \right), \]

where the last inequality follows from Proposition 1. Similarly,

\[ f_{n-4,2}^2 \leq \left( f_{n-2,2} + \frac{4}{(n-4)^2} \right)^2 \leq f_{n-2,2}^2 + O \left( \frac{1}{n^2} \right), \]

where the last inequality follows from Proposition 1. Applying (35)-(37) to (34) we conclude that

\[ \mathbb{E}_{X \sim p} \left( \hat{M}_{\beta_1, \beta_2}(X^n) - M(X^n) \right)^2 \leq \left( n(n-1)\beta_1^2 - 2n\beta_1 + 1 \right) f_{n-2,1}^2 + \left( 4(\beta_1n-1) - 2\beta_2 \binom{n}{2} + 2\beta_1\beta_2 \binom{n}{1,2} \right) f_{n-2,1}f_{n-2,2} + \left( \binom{n}{2} - 1 \right)f_{n-2,2} + o \left( \frac{1}{n} \right). \]

As a corollary, we show that for \( \beta_1 = \frac{1}{n} \) and \( \beta_2 = 0 \), (38) degenerates to the GT bound [5].

**Corollary 1** Let \( p \) be a probability distribution over a countable alphabet \( \mathcal{X} \). Let \( f_{n,r} = \sum_{u \in \mathcal{X}} p^r(u)e^{-np(u)} \).

Then, for \( n \geq 4 \) the following holds

\[ \mathbb{E}_{X \sim p} \left( \hat{M}^{GT}(X^n) - M(X^n) \right)^2 \leq \left( n(n-1)\beta_1^2 - 2n\beta_1 + 1 \right) f_{n-2,1}^2 + \left( 4(\beta_1n-1) - 2\beta_2 \binom{n}{2} + 2\beta_1\beta_2 \binom{n}{1,2} \right) f_{n-2,1}f_{n-2,2} + \left( \binom{n}{2} - 1 \right)f_{n-2,2} + o \left( \frac{1}{n} \right) \bigg|_{\beta_1 = \frac{1}{n}, \beta_2 = 0} \leq \frac{1}{n} \mathbb{E}_{X \sim p} \left( \frac{2\Phi_2(X^n)}{n} + \frac{\Phi_1(X^n)}{n} \left( 1 - \frac{\Phi_1(X^n)}{n} \right) \right) + o \left( \frac{1}{n} \right). \]
Proof. We have
\[
\left( n(n-1)\beta_1^2 - 2n\beta_1 + 1 \right) f_{n-2,1}^2 + \left( 4(n-1) - 2\beta_2 \left( \frac{n}{2} \right) + 2\beta_1\beta_2 \left( \frac{n}{1} \right) \right) f_{n-2,1}f_{n-2,2} + \left( -\frac{1}{n} \right) = \left( \frac{n}{2} \right) f_{n-2,2}^2 + n\beta_1^2 f_{n-2,1} + \left( 2\beta_2^2 \left( \frac{n}{2} \right) + 1 \right) f_{n-2,2} + o \left( \frac{1}{n} \right) \bigg|_{\beta_1 = \frac{1}{n}, \beta_2 = 0}
\]

Let us separately study each of the terms in (40). First,
\[
-\frac{1}{n} f_{n-2,1}^2 \leq -\frac{1}{n} \sum_{u \neq v} p(u)p(v)e^{-(n-2)(p(u)+p(v))} \leq \frac{1}{n} \sum_{u \neq v} p(u)p(v)(1 - p(u) + p(v))^{n-2} = -\frac{1}{n} \mathbb{E}_{X^n \sim p} \left( \frac{\Phi_2(X^n)}{n^2} \right) + o \left( \frac{1}{n} \right),
\]
where the second inequality follows from (14) and the last equality follows from Equation (11) in [5].

Next, we have
\[
\frac{1}{n} f_{n-2,1} \leq \frac{1}{n} f_{n-1,1} + o \left( \frac{1}{n} \right) \leq \frac{1}{n} \sum_{u} p(u)(1 - p(u))^{n-1} + o \left( \frac{1}{n} \right) = \frac{1}{n} \mathbb{E}_{X^n \sim p} \left( \frac{\Phi_1(X^n)}{n} \right) + o \left( \frac{1}{n} \right),
\]
where the first inequality follows from (35), the second inequality follows from (16) and Proposition 1, and the last equality follows from Equation (9) in [5]. Finally,
\[
f_{n-2,2} \leq f_{n,2} + o \left( \frac{1}{n} \right) \leq \sum_{u} p^2(u)(1 - p(u))^n + o \left( \frac{1}{n} \right) = \frac{1}{n} \mathbb{E}_{X^n \sim p} \left( \frac{2\Phi_2(X^n)}{n} \right) + o \left( \frac{1}{n} \right),
\]
where the first inequality follows from (35), the second inequality follows from (16) and Proposition 1, and the last equality follows from Equation (10) in [5]. Putting all together we obtain,
\[
\mathbb{E}_{X^n \sim p} \left( \hat{M}^GT(X^n) - M(X^n) \right)^2 \leq -\frac{1}{n} f_{n-2,1}^2 + \frac{1}{n} f_{n-2,1} + f_{n-2,2} + o \left( \frac{1}{n} \right) \leq \frac{1}{n} \mathbb{E}_{X^n \sim p} \left( \frac{2\Phi_2(X^n)}{n} \right) + \frac{\Phi_1(X^n)}{n} \left( 1 - \frac{\Phi_1(X^n)}{n} \right) + o \left( \frac{1}{n} \right),
\]
as appears in [5].
Appendix C

Theorem 2 Let \( Z(x,y) = Ax^2 + Bxy + Cy^2 + Dx + Ey \), where \( 0 \leq x \leq x_{\text{max}}, \) \( 0 \leq y \leq y_{\text{max}} \) and \( x \geq y \). Denote \((x^*, y^*) = \arg \max Z(x,y)\). Then, the following list is a set of candidates for the global optimum, \((x^*, y^*)\):

1. \( \left( \begin{array}{c} x \\ y \end{array} \right) = \left( \begin{array}{cc} 2A & B \\ B & 2C \end{array} \right)^{-1} \left( \begin{array}{c} -D \\ -E \end{array} \right) \), provided that \( \left( \begin{array}{cc} 2A & B \\ B & 2C \end{array} \right) \) is invertible, and \( \left( \begin{array}{c} x \\ y \end{array} \right) \) satisfy the conditions above.

2. \( \left( x_{\text{max}}, \frac{-E-Bx_{\text{max}}}{2C} \right)^T \), provided that \( 0 \leq \frac{-E-Bx_{\text{max}}}{2C} \leq y_{\text{max}} \).

3. \( \left( \frac{-D-By_{\text{max}}}{2A}, y_{\text{max}} \right)^T \), provided that \( y_{\text{max}} \leq \frac{-D-By_{\text{max}}}{2A} \leq x_{\text{max}} \).

4. \( \left( \frac{-D}{2A}, 0 \right)^T \), provided that \( y_{\text{max}} \leq \frac{-D}{2A} \leq x_{\text{max}} \).

5. \( \left( x_{\text{max}}, 0 \right)^T, \left( x_{\text{max}}, y_{\text{max}} \right)^T, \left( y_{\text{max}}, y_{\text{max}} \right)^T, \left( 0, 0 \right)^T \).

Proof. The maximizers of \( Z(x,y) \) are either optima on the interior, \( 0 < x < x_{\text{max}}, 0 < y < y_{\text{max}} \), or points on the boundaries of this set. Taking the derivatives of \( Z(x,y) \), we have

\[
\frac{\partial Z(x,y)}{\partial x} = 2Ax + By + D = 0 \tag{45}
\]

\[
\frac{\partial Z(x,y)}{\partial y} = 2Cy + Bx + E = 0. \tag{46}
\]

Therefore, the optimum on the interior satisfies

\[
\left( \begin{array}{c} x^* \\ y^* \end{array} \right) = \Omega^{-1} \left( \begin{array}{c} -D \\ -E \end{array} \right), \quad \text{where} \quad \Omega = \left( \begin{array}{cc} 2A & B \\ B & 2C \end{array} \right).
\]

This optimum is a feasible solution, provided that it satisfies the constraints of Theorem 2 and \( \Omega \) is an invertible matrix. Now, let us examine the boundaries of the set. Let us start by setting \( x = x_{\text{max}} \). Looking for the optimum with respect to \( y \) (46) yields \( y^* = \frac{-E-Bx_{\text{max}}}{2C} \). Therefore, \( \left( x_{\text{max}}, \frac{-E-Bx_{\text{max}}}{2C} \right)^T \) is a candidate for a global optimum, provided that it satisfies the constraints of Theorem 2. Let us now set \( y = y_{\text{max}} \). Here, we have \( x^* = \frac{-D-By_{\text{max}}}{2A} \). Then, \( \left( \frac{-D-By_{\text{max}}}{2A}, y_{\text{max}} \right)^T \) is again a candidate for a global optimum, provided that it satisfies the constraints of Theorem 2. Finally, for \( y = 0 \) we have \( x^* = \frac{-D}{2A} \) and \( \left( \frac{-D}{2A}, 0 \right)^T \) is also a candidate for a global optimum, given that it satisfies the constraints. In addition, we examine the vertexes of the set, \( \left( x_{\text{max}}, 0 \right)^T, \left( x_{\text{max}}, y_{\text{max}} \right)^T, \left( y_{\text{max}}, y_{\text{max}} \right)^T, \left( 0, 0 \right)^T \), which may also optimize our objective. Putting the above together, we obtain a list of candidates, as described in Theorem 2. \( \blacksquare \)
Appendix D: A Proof for Theorem 3

We prove Theorem 3 by a series of properties.

Property 2 Let \( f_{n,r} = \sum_{u \in \mathcal{X}} p^r(u) e^{-np(u)} \). The summand, \( h_r(p(u)) = p^r(u) e^{-np(u)} \), is

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

\[
\frac{d^2 h_r(p(u))}{dp^2(u)} = p^{r-2}(u) e^{-np(u)} \left( n^2 p^2(u) - 2np + r(r-1) \right) \tag{47}
\]

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 represent the bounds between the convex and concave regions.](image)

Property 3 Let \( p^* \in \Delta_k \) be the maximizer of \( f_{n,r} = \sum_u h_r(p(u)) \). 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

\[
\tilde{p}(l) = \begin{cases} 
      p^*(l) & l \neq u, v \\
      \frac{p^*(u) + p^*(v)}{2} & l = u, v
   \end{cases} \tag{48}
\]
Then,
\[
\sum_l h_r(\tilde{p}(l)) = \sum_{l \neq u, v} h_r(\tilde{p}(l)) + \sum_{l = u, v} h_r(\tilde{p}(l)) = 
\sum_{l \neq u, v} h_r(p^*(l)) + 2h_r\left(\frac{p^*(u) + p^*(v)}{2}\right) > 
\sum_{l \neq u, v} h_r(p^*(l)) + h_r(p^*(u)) + h_r(p^*(v)) = \sum_u h_r(p^*(u)).
\]

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

**Property 4** Let \(p^* \in \Delta_k\) be the maximizer of \(f_{n,r} = \sum_u h_r(p(u))\). Then, there exists at most a single \(p^*(u)\) such that \(p^*(u) \in \left(\frac{r+\sqrt{r}}{n}, 1\right]\).

**Proof.** By negation, assume there exist \(p^*(u)\) and \(p^*(v)\) such that \(p^*(u), p^*(v) \in \left(\frac{r+\sqrt{r}}{n}, 1\right]\). 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 \(p(u) \in \left[\frac{r+\sqrt{r}}{n}, 1\right]\) and strictly convex for \(p(u) \in \left(\frac{r+\sqrt{r}}{n}, 1\right]\). Therefore, we have

\[
h_r\left(\frac{r + \sqrt{r}}{n}\right) \geq h_r\left(p^*(v)\right) - \delta h'_r\left(p^*(v)\right)
\]

\[
h_r\left(p^*(u) + \delta\right) > h_r\left(p^*(u)\right) + \delta h'_r\left(p^*(u)\right)
\]

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

\[
h_r\left(\frac{r + \sqrt{r}}{n}\right) + h_r\left(p^*(u) + \delta\right) > h_r\left(p^*(v)\right) + h_r\left(p^*(u)\right) + \delta \left(h'_r\left(p^*(u)\right) - h'_r\left(p^*(v)\right)\right). \tag{52}
\]

We observe that \(h'_r(p(u))\) is an increasing function in \(p(u)\), for \(p(u) \in \left(\frac{r+\sqrt{r}}{n}, 1\right]\), as its derivative, \(\frac{d^2h_r(p(u))}{dp^2(u)}\) is positive in this range. Therefore, \(h'_r(p^*(u)) \geq h'_r(p^*(v))\) and

\[
h_r\left(\frac{r + \sqrt{r}}{n}\right) + h_r\left(p^*(u) + \delta\right) > h_r\left(p^*(v)\right) + h_r\left(p^*(u)\right). \tag{53}
\]

Therefore, we define \(\tilde{p} \in \Delta_k\) such that

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

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

13
We observe that without loss of generality, that \( p \) is positive in this range. Therefore, we found \( \tilde{p} \) such that \( h_r(p) > h_r(\tilde{p}) \), which contradicts the optimality of \( p^\ast \).

**Property 5** Let \( p^\ast \in \Delta_k \) be the maximizer of \( f_{n,r} = \sum_u h_r(p(u)) \). Then, there exists at most a single \( p^\ast(u) \) such that \( p^\ast(u) \in \left(0, \frac{r-\sqrt{c}}{n}\right) \).

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

Let us first assume that \( p^\ast(u) + \delta < \frac{r-\sqrt{c}}{n} \). The function \( h_r(p(u)) \) is convex for \( p(u) \in \left[0, \frac{r-\sqrt{c}}{n}\right] \) and strictly convex for \( p(u) \in \left[0, \frac{r-\sqrt{c}}{n}\right] \). Therefore, we have

\[
h_r(p^\ast(u) + \delta) > h_r(p^\ast(u)) + \delta h'_r(p^\ast(u)) \quad (55)
\]

\[
h_r(p^\ast(v) - \delta) \geq h_r(p^\ast(v)) - \delta h'_r(p^\ast(v)) \cdot \quad (56)
\]

Putting together the above, we have

\[
h_r(p^\ast(u) + \delta) + h_r(p^\ast(v) - \delta) > h_r(p^\ast(u)) + h_r(p^\ast(v)) + \delta \left(h'_r(p^\ast(u)) - h'_r(p^\ast(v))\right) \quad (57)
\]

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

\[
h_r(p^\ast(u) + \delta) + h_r(p^\ast(v) - \delta) > h_r(p^\ast(u)) + h_r(p^\ast(v)) \quad (58)
\]

Therefore, we found \( \tilde{p} \in \Delta_k \) such that

\[
\tilde{p}(l) = \begin{cases} p^\ast(l) & l \neq u, v \\
0 & l = v \\
p^\ast(u) + \delta & l = u \end{cases} \quad (59)
\]

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

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

\[
h_r(p^\ast(u) + \hat{\delta}) \geq h_r(p^\ast(u)) + \hat{\delta} h'_r(p^\ast(u)) \quad (60)
\]

\[
h_r(p^\ast(v) - \hat{\delta}) > h_r(p^\ast(v)) - \hat{\delta} h'_r(p^\ast(v)) \cdot \quad (61)
\]
Putting together the above, we have

\[
    h_r \left( p^*(u) + \delta \right) + h_r \left( p^*(v) - \delta \right) > h_r \left( p^*(u) \right) + h_r \left( p^*(v) \right) + \delta \left( h'_r \left( p^*(u) \right) - h'_r \left( p^*(v) \right) \right).
\]  

(62)

As above, we observe that \( h'_r \left( p(u) \right) \) is an increasing function in \( p(u) \), for \( p(u) \in \left(0, \frac{r - \sqrt{T}}{n}\right) \). Therefore, \( h'_r \left( p^*(u) \right) \geq h'_r \left( p^*(v) \right) \) and

\[
    h_r \left( p^*(v) - \delta \right) + h_r \left( p^*(u) + \delta \right) > h_r \left( p^*(v) \right) + h_r \left( p^*(u) \right).
\]  

(63)

Therefore, we found \( \tilde{p} \in \Delta_n \) such that

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

(64)

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

**Property 6** Let \( p^* \in \Delta_n \) be the maximizer of \( f_{n,r} = \sum_u h_r(p(u)) \). Assume that \( k \frac{r + \sqrt{T}}{n} < 1 \). Then, there exists a single \( p^*(u) \) such that \( p^*(u) \in \left[ \frac{r + \sqrt{T}}{n}, 1 \right] \), while \( p^*(v) = \frac{1 - p^*(u)}{k - 1} \) for all \( v \neq u \).

**Proof.**

Property 2 characterizes \( h_r(p(u)) \) for different values of \( p(u) \). Let us first assume that \( p^*(u) \leq \frac{r + \sqrt{T}}{n} \) for every \( u \). This means that \( 1 = \sum_u p^*(u) \leq k \frac{r + \sqrt{T}}{n} \). This contradicts the assumption of Property 6, that \( k \frac{r + \sqrt{T}}{n} < 1 \). Therefore, there exists at least a single \( p^*(u) \) such that \( p^*(u) \in \left( \frac{r + \sqrt{T}}{n}, 1 \right] \). On the other hand, Property 4 shows that there exists at most a single \( p^*(u) \) such that \( p^*(u) \in \left( \frac{r + \sqrt{T}}{n}, 1 \right] \). Therefore, we conclude that for \( k \frac{r + \sqrt{T}}{n} < 1 \), there exists exactly a single \( p^*(u) \) such that \( p^*(u) \in \left( \frac{r + \sqrt{T}}{n}, 1 \right] \). Next, Property 5 shows that there exists at most a single \( p^*(l) \in \left( 0, \frac{r - \sqrt{T}}{n} \right] \). Define \( \tilde{p}(l) = p^*(l) + \delta \) and \( \tilde{p}(u) = p^*(u) - \delta \) where \( \delta = \frac{r - \sqrt{T}}{n} - p^*(l) \). Notice we have that \( \tilde{p}(u) \in \left( \frac{r + \sqrt{T}}{n}, 1 \right] \), as \( k \frac{r + \sqrt{T}}{n} < 1 \). We have,

\[
    h_r(\tilde{p}(l)) = h_r \left( \frac{r - \sqrt{T}}{n} \right) > h_r(p^*(l))
\]  

(65)

\[
    h_r(\tilde{p}(u)) > h_r(p^*(u)),
\]  

(66)

as \( h_r(p(u)) \) is monotonically increasing in \( p(u) \in \left[ 0, \frac{r - \sqrt{T}}{n} \right] \) and monotonically decreasing in \( p(u) \in \left[ \frac{r + \sqrt{T}}{n}, 1 \right] \). Therefore, we show that by increasing \( p^*(l) \) by \( \delta \), and reducing \( p^*(u) \) by the same \( \delta \), 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 \( p^*(u) \in \left( \frac{r+\sqrt{n}}{n}, 1 \right] \), while \( p^*(l) \in \left[ \frac{r}{n}, \frac{r+\sqrt{n}}{n} \right] \) for every \( l \neq u \). Finally, we apply Property 3 to attain the desired result.

### Appendix E: A Proof for Theorem 4

We closely follow the elegant proof of [6]. Let \( p \in \mathbb{R}^j \) and \( k \in \mathbb{R}^j \) be non-negative vectors. For the simplicity of the presentation, assume that \( X = \{1, 2, \ldots \} \). Let

\[
F(p, k, j) = AG_1^2(p, k) + BG_1(p, k)G_2(p, k) + CG_2^2(p, k) + DG_1(p, k) + EG_2(p, k),
\]

where

\[
G_1(p, k) = \sum_{u=1}^{j} k(u)p(u)e^{-\tilde{n}p(u)}, \quad G_2(p, k) = \sum_{u=1}^{j} k(u)p^2(u)e^{-\tilde{n}p(u)}.
\]

Notice that for \( k \in \mathbb{N}^j \) and \( p \) such that \( \sum_{u=1}^{j} k(u)p(u) = 1 \), the tuple \((p, k)\) defines a probability distribution where there are \( k(u) \) symbols with a probability \( p(u) \). Therefore, \( F(p, k, j) \) is equivalent to (18) in the main text, for sufficiently large \( j \) and \( \tilde{n} = n - 2 \). Let \( D_j \) be the set of all pairs \((p, k)\) such that \( p, k \in \mathbb{R}^j \) (that is, \( k \) need not be integral) and \( \sum_{u=1}^{j} k(u)p(u) = 1 \). Define

\[
M_j = \max_{(p, k) \in D_j} F(p, k, j).
\]

We show that if there exists \( \kappa \in \mathbb{R}_+ \) that satisfies (19) in the main text, then \( M_j \leq \alpha = \max_k e^{-\frac{\tilde{n}k}{k}} \left( A + \frac{B}{k} + \frac{C}{k^2} \right) + e^{-\frac{\tilde{n}}{k}} \left( D + \frac{E}{k} \right) \) for any \( j > 1 \).

**Proposition 4** Assume there exists \( \kappa \in \mathbb{R}_+ \) that satisfies (19) in the main text. Then, for any \( j > 1 \), \( M_j \leq \max\{M_{j-1}, \alpha\} \).

**Proof.** The Lagrangian of our maximization problem (67) is given by

\[
\mathcal{L}(p, k, j) = F(p, k, j) - \lambda \left( \sum_{u=1}^{j} k(u)p(u) - 1 \right) - \sum_{u=1}^{j} \mu(u)p(u) - \sum_{u=1}^{j} \nu(u)k(u).
\]

The extrema of the Lagrangian satisfy

\[
\frac{\partial \mathcal{L}}{\partial p(u)} = \frac{\partial \mathcal{L}}{\partial k(u)} = 0 \quad \text{for every} \quad u = 1, \ldots, j.
\]

Consider the two possible cases:

(I) If for some \( u \), we have \( k(u)p(u) = 0 \), then \( M_j \leq M_{j-1} \) since dropping \( p(u) \) and \( k(u) \), gives a valid element in \( D_{j-1} \).
(II) If every $k(u)p(u) > 0$, then $\mu(u) = \nu(u) = 0$ for every $u$, and

\[
\frac{\partial L}{\partial p(u)} = 2AG_1 k(u) e^{-\tilde{n}p(u)} (1 - \tilde{n}p(u)) + BG_2 k(u) e^{-\tilde{n}p(u)} (1 - \tilde{n}p(u)) +
\]

\[
BG_1 k(u) p(u) e^{-\tilde{n}p(u)} (2 - \tilde{n}p(u)) + 2CG_2 k(u) p(u) e^{-\tilde{n}p(u)} (2 - \tilde{n}p(u)) +
\]

\[
Dk(u) e^{-\tilde{n}p(u)} (1 - \tilde{n}p(u)) + Ek(u) p(u) e^{-\tilde{n}p(u)} (2 - \tilde{n}p(u)) - \lambda k(u) = 0.
\]

Putting together (72) and (74) we have

\[
2AG_1 (1 - \tilde{n}p(u)) + BG_2 (1 - \tilde{n}p(u)) + BG_1 p(u)(2 - \tilde{n}p(u)) + 2CG_2 p(u)(2 - \tilde{n}p(u)) +
\]

\[
D(1 - \tilde{n}p(u)) + Ep(u)(2 - \tilde{n}p(u)) = 2AG_1 + BG_2 + BG_1 p(u) + 2CG_2 p(u) + D + Ep(u),
\]

which leads to

\[
\tilde{n}p(u)(BG_1 + 2CG_2 + E) = BG_1 + 2CG_2 + E - 2\tilde{n}AG_1 - \tilde{n}BG_2 - \tilde{n}D.
\]

Thus, all $p(u)$'s are the same. From $\sum_{u=1}^{j} k(u)p(u) = 1$ we have that $p(u) = 1/k$ where $k = k(1) + \cdots + k(j)$. Therefore,

\[
G_1(p, k) \bigg|_{p(u) = \frac{1}{k}} = e^{-\tilde{n}/k}, \quad G_2(p, k) \bigg|_{p(u) = \frac{1}{k}} = \frac{1}{k} e^{-\tilde{n}/k}.
\]

Putting together (76) and (77), we have that $p(u) = 1/k$ is a feasible extremum if there exists $\kappa > 0$ such that

\[
\frac{\tilde{n}}{\kappa} (Be^{-\tilde{n}/k} + 2C \frac{1}{\kappa} e^{-\tilde{n}/\kappa} + E) =
\]

\[
Be^{-\tilde{n}/k} + 2C \frac{1}{\kappa} e^{-\tilde{n}/\kappa} + E - 2\tilde{n} A e^{-\tilde{n}/\kappa} - \tilde{n} B \frac{1}{\kappa} e^{-\tilde{n}/\kappa} - \tilde{n} D.
\]
Given Assumption (19) in the main text, we have that the above holds and

\[ F(p,k,j) \leq e^{\frac{-2\tilde{n}}{\kappa}} \left( A + \frac{B}{\kappa} + \frac{C}{\kappa^2} \right) + e^{\frac{-\tilde{n}}{\kappa}} \left( D + \frac{E}{\kappa} \right) \leq \alpha \]  

(79)

where \( \alpha = \max_k e^{\frac{-2\tilde{n}}{\kappa}} \left( A + \frac{B}{\kappa} + \frac{C}{\kappa^2} \right) + e^{\frac{-\tilde{n}}{\kappa}} \left( D + \frac{E}{\kappa} \right) \).

Plugging \( \tilde{n} = n - 2 \) yields the desired result. ■

Appendix F: A proof for Theorem 5

In Section 3.1.5 we show that the risk of our suggested estimator ((21) in the main text) is bounded from above by the GT risk, for every \( n \) that we examine, as appears in Figure 3 (specifically, \( n \leq 5 \cdot 10^4 \)). We now show that this property holds for every \( n \geq 4 \). Our proof consists of two parts. First, we show that our suggested estimator satisfies condition (19) in the main text, for every \( n \geq N_0 \), where \( N_0 \leq 5 \cdot 10^4 \).

Then, we show that the upper bound on our risk ((20) in the main text) is lower then the GT risk, for every \( n \geq N_0 \). To avoid an overload of notation we slightly abuse our previous definitions and set

\[ \beta_1 = \frac{1}{n} \left( 1 - \frac{a_1}{n^b} \right), \quad \beta_2 = \frac{a_2}{n^d} \]  

(80)

where \( a_1 = 2.08 \), \( a_2 = 4.1 \), \( b = 0.7 \) and \( d = 1.7 \).

**Proposition 5** Let \( A = n(n-1)\beta_1^2 - 2n\beta_1 + 1 \), \( B = 4(\beta_1 n - 1) - 2\beta_2(\binom{n}{2}) + 2\beta_1\beta_2(\binom{n}{2}) \), \( C = \beta_2^2(\binom{n}{2}) \), \( D = n\beta_1^2 \), \( E = \beta_2^2(\binom{n}{2}) + 1 \), where \( \beta_1 \) and \( \beta_2 \) are the estimator’s coefficients, as defined in (80). Then, there exists \( \kappa \in \mathbb{R}_+ \) that satisfies

\[ \kappa = (n - 2) + (n - 2)\kappa \left( \frac{e^{-(n-2)/\kappa}(2A\kappa + B) + D\kappa}{e^{-(n-2)/\kappa}(B\kappa + 2C) + E\kappa} \right). \]  

(81)

for every \( n \geq N_0 \), where \( N_0 = 3000 \).

**Proof.** We first observe that (81) may be equivalently written as

\[ T(n,k) \triangleq \left( \frac{n - 2}{k} - 1 \right) \left( Be^{-(n-2)/k} + 2C \frac{1}{k} e^{-(n-2)/k} + E \right) + 2(n - 2)Ae^{-(n-2)/k} + B \frac{n - 2}{k} e^{-(n-2)/k} + (n - 2)D = 0. \]  

(82)

We would like to show that there exists \( k \in \mathbb{R}_+ \) such that \( T(n,k) = 0 \). Notice that \( T(n,k) \) is continuous in \( k \). Therefore, it is enough to show that there exists \( k_1, k_2 \in \mathbb{R}_+ \) such that \( T(n,k_1) > 0 \) and \( T(n,k_2) < 0 \), for every \( n \geq N_0 \). Let us begin with \( k_1 = n - 2 \). In this case we have

\[ T(n,n - 2) = 2(n - 2)Ae^{-1} + Be^{-1} + (n - 2)D. \]  

(83)
We separately study each of the terms in (83). First,
\[ 2(n-2)Ae^{-1} = 2(n-2)\left(n(n-1)\beta_1^2 - 2\beta_1 n + 1\right) e^{-1} \geq -2e^{-1} \frac{n-2}{n} \geq -2e^{-1}, \quad (84) \]
where the inequality follows from \(0 \leq \beta_1 \leq \frac{1}{n}\). Second,
\[ (n-2)D = (n-2)n\beta_1^2 = \frac{n-2}{n} \left(1 - \frac{a_1}{n^b}\right) \geq \frac{N_0 - 2}{N_0} \left(1 - \frac{a_1}{N_0^b}\right)^2, \quad (85) \]
where the inequality follows from \(\frac{n-2}{n} \left(1 - \frac{a_1}{n^b}\right)^2\) being monotonically increasing in \(n\). Last, we have that
\[ B = 4(\beta_1 n - 1) + 2\beta_2 \left(\beta_1 \left(\frac{n}{12} - \left(\frac{n}{2}\right)\right) = 4(\beta_1 n - 1) + \beta_2 n(n-1)(\beta_1(n-2) - 1) = \right. \]
\[ \left. \frac{-4a_1}{n^b} + \frac{a_2n(n-1)}{n^d}\left(\frac{n-2}{n} \left(1 - \frac{a_1}{n^b}\right) - 1\right) \geq \frac{-4a_1}{n^b} + \frac{a_2}{n^{d-2}} \frac{2a_1 - 2n^b - a_1 n}{n^{b+1}} \right. \]
\[ \left. \frac{-4a_1}{n^b} + \frac{a_2(2a_1 - 2n^b - a_1 n)}{n^{b+d-1}} \geq \frac{-4a_1}{N_0^b} + \frac{a_2(2a_1 - 2N_0^b - a_1 N_0)}{N_0^{b+d-1}}, \right. \]
where the first inequality follows from \(\frac{n-2}{n} \left(1 - \frac{a_1}{n^b}\right)^2\) being monotonically increasing in \(n\). Putting the above together, we have
\[ T(n, n-2) \geq -2e^{-1} + e^{-1} \left(\frac{-4a_1}{N_0^b} + \frac{a_2(2a_1 - 2N_0^b - a_1 N_0)}{N_0^{b+d-1}}\right) + \frac{N_0 - 2}{N_0} \left(1 - \frac{a_1}{N_0^b}\right)^2 > 0, \quad (87) \]
for \(n \geq N_0\). We now proceed to \(k_2 = c(n-2)\) and show that \(\lim_{c \to \infty} T(n, c(n-2)) < 0\) for \(n \geq N_0\). We have
\[ \lim_{c \to \infty} T(n, c(n-2)) = -B - E + 2(n-2)A + (n-2)D. \quad (88) \]
As in (83), we separately analyze each of the terms in (88). First,
\[ -B \leq \frac{4a_1}{N_0^b} - \frac{a_2(2a_1 - 2N_0^b - a_1 N_0)}{N_0^{b+d-1}}, \quad (89) \]
where the inequality follows from (86). Second,
\[ -E = -\beta_2^2 \left(\frac{n}{2}\right) - 1 = -\frac{n(n-1)a_2^2}{2n^{2d}} - 1 \leq -1, \quad (90) \]
where the second inequality follows from \( \frac{n(n-1)a_n^2}{2n^2} > 0 \). Third,

\[
2(n-2)A = 2(n-2) \left( n(n-1)\beta_1^2 - 2\beta_1 n + 1 \right) = 2(n-2) \left( \frac{a_1^2}{n^2} + \frac{2a_1}{n^{b+1}} - \frac{1}{n} - \frac{a_1^2}{n^{2b+1}} \right) \leq 2(N_0 - 2) \left( \frac{a_1^2}{N_0^{2b}} + \frac{2a_1}{N_0^{b+1}} - \frac{1}{N_0} - \frac{a_1^2}{N_0^{2b+1}} \right),
\]

where the inequality follows from \( 2(n-2) \left( \frac{a_1^2}{n^2} + \frac{2a_1}{n^{b+1}} - \frac{1}{n} - \frac{a_1^2}{n^{2b+1}} \right) \) being monotonically decreasing, for every \( n \geq N_0 \). Finally,

\[
(n-2)D = n(n-2)\beta_1^2 = \frac{n-2}{n} \left( 1 - \frac{a_1}{n^b} \right)^2 \leq \left( 1 - \frac{a_1}{n^b} \right)^2 \leq 1. \tag{92}
\]

Putting the above together, we have

\[
\lim_{c \to \infty} T(n, c(n-2)) \leq \frac{4a_1}{N_0^b} - \frac{a_2}{N_0^b + d-1} \left( 2a_1 - 2N_0^b - a_1 N_0 \right) + 2(N_0 - 2) \left( \frac{a_1^2}{N_0^{2b}} + \frac{2a_1}{N_0^{b+1}} - \frac{1}{N_0} - \frac{a_1^2}{N_0^{2b+1}} \right) < 0,
\]

where the last inequality follows from \( N_0 = 3000 \).

Before we proceed to the second part of our proof, we present the following property.

**Lemma 1** Let \( k_0 \in \mathbb{R}_+ \) satisfy

\[
k_0 = \arg \max_k e^{-\frac{2(n-2)}{k}} \left( A + \frac{B}{k} + \frac{C}{k^2} \right) + e^{-\frac{n-2}{k}} \left( D + \frac{E}{k} \right),
\]

where \( A, B, C, D \) and \( E \) follow the definition in Proposition 5, and \( \beta_1, \beta_2 \) are defined in (80). Then for \( n \geq N_0 \) and \( \alpha_0 = 0.1 \), we have that \( k_0 > \alpha_0(n-2) \).

**Proof.** Let us look at the first derivative of our objective function,

\[
L_u(k, n) = e^{-\frac{2(n-2)}{k}} \left( A + \frac{B}{k} + \frac{C}{k^2} \right) + e^{-\frac{n-2}{k}} \left( D + \frac{E}{k} \right)
\]

(with respect to \( k \)) and show that it is positive for every \( 0 < k \leq \alpha_0(n-2) \), where \( n \geq N_0 \) and \( \alpha_0 = 0.1 \). This means that \( 0 < k \leq \alpha_0(n-2) \) cannot be the maximizer of \( L_u(k, n) \), as we can always attain a greater
objective value.

\[ \frac{\partial}{\partial k} L_n(k, n) = \]
\[ \frac{1}{k^2} e^{-\frac{2(n-2)}{k} k} \left( (2(n-2)A + 2(n-2)B + 2(n-2)k^2 C) + \right. \]
\[ \frac{1}{k^2} e^{-\frac{2(n-2)}{k} k} \left( (n-2)D + \frac{(n-2) - k}{k} E \right) \geq \]
\[ \frac{1}{k^2} e^{-\frac{2(n-2)}{k} k} \left( (2(n-2)A + 2(n-2)B + 2(n-2)k^2 C + (n-2)D + \frac{(n-2) - k}{k} E \right), \]

where the inequality follows from \((n-2)D \geq 0\), \(\frac{(n-2) - k}{k} E \geq 0\) for \(k \leq \alpha_0 (n - 2) < (n - 2)\), and \(e^{-\frac{n-2}{k} k} > e^{-\frac{2(n-2)}{k} k}\). We would like to show that (96) is positive for \(0 < k \leq \alpha_0 (n - 2)\). Therefore, it is enough to show that

\[ 2(n-2)k^2 A + k(2(n-2) - k)B + 2((n-2) - k)C + k^2(n-2)D + k((n-2) - k)E > 0 \]  

(97)

for \(0 < k \leq \alpha_0 (n - 2)\). Rearranging the terms in (97), we show it is quadratic in \(k\),

\[ Q(k, n) \triangleq k^2 (2(n-2)A - B + (n-2)D - E) + \]
\[ k (2(n-2)B - 2C + (n-2)E) + 2(n-2)C. \]

In Proposition 5, we show that \(2(n-2)A - B + (n-2)D - E < 0\) for every \(n \geq N_0\) (specifically, see (88)). This means that \(Q(k, n)\) is strictly concave in \(k\) for these values of \(n\). Further, we notice that \(\lim_{k \to 0} Q(k, n) > 0\). Therefore, it is enough to show that \(Q(k, n)|_{k=\alpha_0 (n-2)} > 0\) to prove the Lemma.

\[ Q(k, n)|_{k=\alpha_0 (n-2)} = c_0^2 (n-2) (2(n-2)A - B + (n-2)D - E) + \]
\[ c_0 (n-2) (2(n-2)B - 2C + (n-2)E) + 2(n-2)C = \]
\[ c_0^2 (n-2) (2(n-2)A - B + (n-2)D - E) + \]
\[ c_0 (n-2) (2B + E) + 2(n-2)C (1 - \alpha_0) \geq \]
\[ c_0^2 (n-2) (2(n-2)A - B + (n-2)D - E) + c_0 (n-2) (2B + E) = \]
\[ c_0 (n-2) (2(n-2)A - B + (n-2)D - E) + 2B + E \]

where the inequality follows from \(2(n-2)C (1 - \alpha_0) > 0\). This means it is enough to show that \(c_0 (2(n-2)A - B + (n-2)D - E) + 2B + E > 0\). As we separately study each of the terms, we observe the following

(I) \(c_0 (2(n-2)A) \geq -2c_0\), where the inequality follows from (84).

(II) \(B (2 - \alpha_0) \geq (2 - \alpha_0) \left( \frac{4c_0}{N_0^2} + \frac{c_0 (2n - 2N_0^2) - c_0 N_0^2}{N_0^2 + d - 1} \right)\), where the inequality follows from (86).
(III) $\alpha_0(n - 2)D \geq \alpha_0 \frac{N_0 - 2}{N_0} \left(1 - \frac{a_1}{N_0}\right)^2$, where the inequality follows from (85).

(IV) $E(1 - \alpha_0) \geq 1 - \alpha_0$, where the inequality follows from (90).

Putting the above together, we have

$$\alpha_0 (2(n - 2)A - B + (n - 2)D - E) + 2B + E \geq$$

$$- 2\alpha_0 + (2 - \alpha_0) \left(- \frac{4a_1}{N_0} + \frac{a_2(2a_1 - 2N_0 - a_1N_0)}{N_0^{b+d-1}}\right) + \alpha_0 \frac{N_0 - 2}{N_0} \left(1 - \frac{a_1}{N_0}\right)^2 + 1 - \alpha_0 > 0$$

for $n \geq N_0$ and $\alpha_0 = 0.1$, which concludes the proof of Lemma 1. 

We are now ready to conclude the proof of Theorem 5, with the following proposition.

**Proposition 6** Let $\hat{M}^{*}_{\beta_1, \beta_2}(X^n)$ be our suggested estimator, for $\beta_1$ and $\beta_2$ as defined in (80). Then, for every $n \geq N_0$,

$$R_n(\hat{M}^{*}, \Delta) > R_n(\hat{M}^{*}_{\beta_1, \beta_2}, \Delta) + o \left(\frac{1}{n}\right).$$

**Proof.** Corollary 1 shows that the GT risk is bounded from above by (16) in the main text, for $\beta_1 = \frac{1}{n}$ and $\beta_2 = 0$. Applying Theorem 4 to the above (or equivalently, Theorem 1 in [6]) we have that the GT risk, $R_n(\hat{M}^{*}, \Delta)$, is given by

$$R_n(\hat{M}^{*}, \Delta) \leq \max_k e^{-2(n-2)k} \left(A_{GT} + \frac{B_{GT}}{k} + \frac{C_{GT}}{k^2}\right) + e^{-\frac{n-2}{n}k} \left(D_{GT} + \frac{E_{GT}}{k}\right) + o \left(\frac{1}{n}\right),$$

where $A_{GT} = -\frac{1}{n}$, $B_{GT} = 0$, $C_{GT} = 0$, $D_{GT} = \frac{1}{n}$ and $E_{GT} = 1$. On the other hand, Applying Proposition 5 to Theorem 4, we show that the risk of our suggested estimator is bounded from above by

$$R_n(\hat{M}^{*}_{\beta_1, \beta_2}, \Delta) \leq \max_k e^{-2(n-2)k} \left(A + \frac{B}{k} + \frac{C}{k^2}\right) + e^{-\frac{n-2}{n}k} \left(D + \frac{E}{k}\right) + o \left(\frac{1}{n}\right)$$

for $n \geq 1500$, where $A = n(n-1)\beta_1^2 - 2n\beta_1 + 1$, $B = 4(\beta_1 n - 1) - 2\beta_2 (\frac{n}{2}) + 2\beta_1 \beta_2 (\frac{n}{12})$, $C = \beta_2^2 (\frac{n}{2})$, $D = n\beta_1^2$, $E = \beta_2^2 (\frac{n}{2}) + 1$ and $\beta_1, \beta_2$ are the estimator’s coefficients, as defined in (80). Denote $k_{GT}$ as the $k$ that maximizes (102) while $k_0$ is the $k$ that maximizes (103). Notice that $k_0$ depends on $n$, as implied by (103). We would like to show that

$$R_n(\hat{M}^{*}, \Delta) - R_n(\hat{M}^{*}_{\beta_1, \beta_2}, \Delta) \geq e^{-\frac{2(n-2)}{n}k_0} \left(A_{GT} - A + \frac{B_{GT} - B}{k_0} + \frac{C_{GT} - C}{k_0^2}\right) + e^{-\frac{n-2}{n}k_0} \left(D_{GT} - D + \frac{E_{GT} - E}{k_0}\right) + o \left(\frac{1}{n}\right) \geq o \left(\frac{1}{n}\right)$$

22
for every $n \geq N_0$. Let us separately study each of the terms in (104). First, we have

$$A_{GT} - A = \frac{1}{n} - \frac{n - 1}{n} \left(1 - \frac{a_1}{n^b}\right)^2 + 2 \left(1 - \frac{a_1}{n^b}\right) - 1 = \frac{a_1^2}{n^{2b+1}} - \frac{a_1^2}{n^{2b}} - \frac{2a_1}{n^{b+1}} = o \left(\frac{1}{n}\right). \quad (105)$$

Second,

$$\frac{1}{k_0} (B_{GT} - B) = -\frac{B}{k_0} = \frac{1}{k_0} \left(\frac{4a_1}{n^b} + 2 \left(\frac{n}{2}\right) a_2 \frac{n^d}{n^{d+1}} - 2 \left(\frac{n}{12}\right) \frac{a_2}{n^{d+1}} + 2 \left(\frac{n}{12}\right) \frac{a_1 a_2}{n^{b+d+1}}\right) \geq \frac{1}{k_0} \left(2 \left(\frac{n}{2}\right) a_2 n \frac{a_2}{n^{d+1}} - 2 \left(\frac{n}{12}\right) \frac{a_2}{n^{d+1}}\right) = \frac{1}{k_0} \left(\frac{2a_2}{n^d} \frac{n(n-1)}{2} \left(1 - \frac{n - 2}{n}\right)\right) > 0. \quad (106)$$

Third,

$$\frac{1}{k_0} (C_{GT} - C) = -\frac{1}{k_0} \frac{a_2^2}{n^{2d}} \left(\frac{n}{2}\right) \geq -\frac{1}{k_0} \frac{a_2^2}{\alpha_0(n-2)^2 n^{2d}} \left(\frac{n}{2}\right) = o \left(\frac{1}{n}\right), \quad (107)$$

where the inequality follows from Lemma 1. Further,

$$(D_{GT} - D) = \frac{1}{n} - \frac{1}{n} \left(1 - \frac{a_1}{n^b}\right)^2 = \frac{1}{n} \left(1 - \left(1 - \frac{a_1}{n^b}\right)^2\right) > 0. \quad (108)$$

Finally,

$$\frac{1}{k_0} (E_{GT} - E) = -\frac{1}{k_0} \left(\frac{n}{2}\right) \frac{a_2^2}{n^{2d}} > -\frac{1}{\alpha_0(n-2)} \left(\frac{n}{2}\right) \frac{a_2^2}{n^{2d}} = o \left(\frac{1}{n}\right). \quad (109)$$

where the inequality follows from Lemma 1. In addition, have that $e^{-\frac{r}{\alpha_0}} \leq e^{-\frac{r(n-2)}{\alpha_0}} \leq 1$, for $r = 1, 2$ due to Lemma 1. Putting the above together we conclude that

$$e^{-\frac{2(n-2)}{\alpha_0}} \left(A_{GT} - A + \frac{B_{GT} - B}{k_0} + \frac{C_{GT} - C}{k_0^2}\right) + e^{-\frac{n-2}{\alpha_0}} \left(D_{GT} - D + \frac{E_{GT} - E}{k_0}\right) \geq o \left(\frac{1}{n}\right), \quad (110)$$

as desired. ■

References


