Measuring Information Leakage in Sample Publications

***Abstract*—Abstract.**

***Keywords*-Quantitative Information Flow**

I. Introduction

⟪MJ1 ⟫ ⟪CP1 ⟫ ⟪MA1 ⟫ ⟪RG1 ⟫

1) Introduction
2) Preliminaries
   a) QIF
3) QIF Model for statistical publications
4) Attribute Inference attack
   Privacy concerns. An adversary that has a single target and wants to infer his attribute value.
   a) Adversary that knows the target is **in** the sample.
      - Prior vulnerability = 1/2
      - Posterior vulnerability closed formula.
   b) Adversary that knows the target is **outside** the sample.
      - Prior vulnerability = 1/2
      - Posterior vulnerability closed formula.
   c) Adversary that **doesn't know** whether the target is in the sample.
5) Utility
   A data analyst that wants to find the distribution of the attribute in the population.
6) Experiments
7) (?) Non-binary attribute
8) Conclusion

A. Related Work

II. Preliminaries

A. QIF

Lorem ipsum [1].

III. QIF Model for statistical publications

Consider a scenario where there exists a population with $n \geq 1$ individuals and one has a binary value (i.e., values $\mathtt{a}$ or $\mathtt{b}$) of a sensitive attribute of interest. Also suppose there is an adversary that is interested in inferring the attribute value of a single person that we will call the target. In g-vulnerability framework we have to define a set of secrets that contains the information the adversary is interested in. For a statistical publication about this population and this adversary, we define the following set of secrets.

Definition 1 (Set of secrets $\mathcal{X}$)**.** *Let $n \geq 1$ be the number of people in the population and consider that each individual from this population has a binary value for a sensitive attribute of interest. Consider also that there is an adversary who aims to infer the attribute value of a single target (an individual from the population). A secret $x = (p, t)$ is a pair where p is a binary array of size n that represents the entire population and t is the adversary target's index. The set of secrets is the set of all possible pairs (p, t). Formally,*

$$\mathcal{X} = \{(p, t) \mid p \in \{\mathtt{a}, \mathtt{b}\}^n \wedge t \in \mathcal{I}\}, \tag{1}$$

where $\mathcal{I} \subseteq \{1, \ldots, n\}$ is the set of possible indexes the adversary's target can assume in the array population. Given a secret $x = (p, t)$ we say that x^p is the population array and x^p_i is the attribute value of the i-th person in the population x^p (e.g., x^p_t is the target's attribute value).

The adversary's prior knowledge about the secret can be described as a probability distribution π on the set of secrets $\mathcal{X}$. It is formalized next.

Definition 2 (Prior distribution π on $\mathcal{X}$)**.** *Let the set of secrets $\mathcal{X}$ be defined according to Definition 1. We assume the adversary only knows what are the possible secrets, and she has no ideia about the frequency of value $\mathtt{a}$ in the population. Because of that, the adversary assumes that all possible frequencies of value $\mathtt{a}$ in this population are equally probable, i.e., the prior probability of any frequency is $1/(n+1)$. Therefore the prior distribution π on $\mathcal{X}$ is defined as*

$$\pi_x = \frac{1}{|\mathcal{I}|(n+1)\binom{n}{n_{\mathtt{a}}(x^p)}}, \tag{2}$$

where $n_{\mathtt{a}}(x^p)$ is the number of $\mathtt{a}$'s in array x^p.

The statistical publication studied in this work is sampling. We are going to consider that a sample of size $1 \leq m < n$ will be randomly selected from the population (i.e., any set of m individuals is equally probable to be selected). Once the sample was selected, the histogram of this sample will be published, i.e., the number of people with value $\mathtt{a}$ and the number of people with value $\mathtt{b}$. As we are restricting to binary attributes, we can say that the publication is just an integer $1 \leq y \leq m$ that represents the number of people in the sample with value $\mathtt{a}$, and consequently, $m - y$ will be the number of people in the sample with value $\mathtt{b}$. We are ready to define a channel S that models the sample publication.

Definition 3 (Channel S)**.** *Let the set of secrets $\mathcal{X}$ be defined according to Definition 1, and let $1 \leq m < n$ be the sample size. We can model a statistical publication as a channel*

$\mathsf{S} : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, *where* $\mathcal{Y} = \{0,1,\ldots,m\}$ *is the set of possible outputs (i.e., the set of all possible histograms of* m *people). Assuming that the order of people in the population array doesn't matter (i.e., any order is equally probable), we can fix the sample to be always the first* m *people in the population* x^p*, i.e., the sample will be just* $x^p_{1\ldots m}$*. Formally,*

$$\mathsf{S}_{x,y} = \begin{cases} 1 & , \text{ if } n_{\mathtt{a}}(x^p_{1\ldots m}) = y \\ 0 & , \text{ otherwise,} \end{cases} \tag{3}$$

where $y \in \mathcal{Y}$ *represents a histogram of a sample of size* m *where* y *people have the value* $\mathtt{a}$ *and* $m - y$ *have the value* $\mathtt{b}$*, and* $n_{\mathtt{a}}(x^p_{1\ldots m})$ *is the number of* $\mathtt{a}$*'s in the first* m *people of population* x^p*. The entry* $\mathsf{S}_{x,y}$ *can be understood as the probability of the published histogram being* y *when the population is* x^p*, i.e.,* $Pr(y|x)$*.*

IV. Attribute Inference Attack

A. Definitions

Consider an adversary that has a single target and she wants to infer the target's sensitive attribute value. There are 3 different adversaries:

- The adversary knows the target is **in** the sample;
- The adversary knows the target is **outside** the sample;
- The adversary **doesn't know** whether the target is in or outside the sample.

In all cases above, as the attribute is binary, the adversary can guess that the target's value is either $\mathtt{a}$ or $\mathtt{b}$, and we will say that she wins 1 if she guesses correctly and 0 otherwise. Using the g-vulnerability framework, we can define three gain functions that model these adversaries.

Definition 4 (Gain function - Target in the sample)**.** *Let the set of secrets* $\mathcal{X}$ *be defined according to Definition 1, and assuming the adversary knows the target is in the sample, the set of possible indexes for the target in the population array is* $\mathcal{I} = \{1,\ldots,m\}$*. Let also* $\mathcal{W} = \{\mathtt{a},\mathtt{b}\}$ *be the set of guesses for the adversary inferring the target's attribute value. The gain function* $g_{in} : \mathcal{X} \times \mathcal{W} \to \{0,1\}$ *is defined as*

$$g_{in}(w,x) \begin{cases} 1 & , \text{ if } x^p_t = w \\ 0 & , \text{ otherwise,} \end{cases} \tag{4}$$

where the condition $x^p_t = w$ *is the target's attribute value being the same as the adversary's guess* w*.*

Definition 5 (Gain function - Target outside the sample)**.** *The definition of* g_{out} *is similar to Definition 4 of* g_{in}*. The only difference is the set of possible indexes for the target in the population array. Here, the adversary knows the target is outside the sample, therefore* $\mathcal{I} = \{m+1,\ldots,n\}$*.*

Definition 6 (Gain function - Unknown target)**.** *The definition of* g_{unk} *is also similar to Definition 4 of* g_{in}*. The only difference is the set of possible indexes for the target in the population array. Here, the adversary doesn't know whether the target is in or outside the sample, therefore* $\mathcal{I} = \{1,\ldots,n\}$*.*

B. Results

Theorem 1 (Prior vulnerability for attribute inference)**.** *Considering the Definitions 4, 5 and 6 of the gain functions* g_{in}*,* g_{out} *and* g_{unk}*, respectively, and also the prior distribution* π *according to Definition 2, their prior vulnerability are*

$$V_{g_{in}}(\pi) = V_{g_{out}}(\pi) = V_{g_{unk}}(\pi) = 1/2. \tag{5}$$

The proof of Theorem 1 is in Appendix A.

Theorem 2 (Posterior vulnerability when the target is in the sample)**.**

$$V_{g_{in}}[\pi \triangleright \mathsf{S}] = \frac{3}{4} + \frac{1}{4(\lfloor \frac{m+1}{2} \rfloor + \lceil \frac{m}{2} \rceil)} \tag{6}$$

The proof of Theorem 2 is in Appendix A.

Theorem 3 (Posterior vulnerability when the target is out of the sample)**.**

$$V_{g_{out}}[\pi \triangleright \mathsf{S}] = \frac{3}{4} - \frac{1}{4\left(\lfloor \frac{m+1}{2} \rfloor + \lceil \frac{m}{2} \rceil + 1\right)} \tag{7}$$

The proof of Theorem 3 is in Appendix A.

Theorem 4 (Posterior vulnerability when the adversary doesn't known whether the target is in or outside the sample)**.**

$$V_{g_{unk}}[\pi' \triangleright \mathsf{S}'] = \begin{cases} \dfrac{3nm + 2m + 5n + 2}{4n(m+2)} & , \text{ if } m \text{ is odd} \\ \dfrac{3nm + 2m + 2n}{4n(m+1)} & , \text{ if } m \text{ is even.} \end{cases} \tag{8}$$

The proof of Theorem 4 is in Appendix A.

V. Non-binary attribute

VI. Conclusion

References

[1] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith, *The Science of Quantitative Information Flow*. Springer, 2020.

Appendix

Lemma 5 (Marginal on y)**.** *Let the set of secrets $\mathcal{X}$ be defined according to Definition 1 and the channel S according to Definition 3. We have that*

$$Pr(y) = \frac{1}{|\mathcal{I}|(m+1)}. \tag{9}$$

Proof:

$$\begin{aligned} Pr(y) &= \sum_{x \in \mathcal{X}} Pr(x) Pr(y|x) \\ &= \sum_{\substack{x \in \mathcal{X}: \\ n_a(x^p_{1\ldots m}) = y}} \frac{1}{|\mathcal{I}|(n+1)\binom{n}{n_a(x^p)}} && \textit{(Def. 2, 3)} \end{aligned}$$

We need to count how many secrets $x \in \mathcal{X}$ satisfy the restriction $n_a(x^p_{1\ldots m}) = y$. In the first m elements of x^p we have y a's, so $\binom{m}{y}$ different combinations. The other $n-m$ people can have any value, so we say that there are y' a's in $x^p_{m+1\ldots n}$, such that y' goes from 0 to $n-m$. Finally, $n_a(x^p_{1\ldots m}) = y$ and $n_a(x^p_{m+1\ldots n}) = y'$ implies $n_a(x^p) = y + y'$.

$$= \frac{1}{|\mathcal{I}|(n+1)} \sum_{y'=0}^{n-m} \binom{m}{y} \binom{n-m}{y'} \binom{n}{y+y'}^{-1}$$

by Lemma 7:

$$\begin{aligned} &= \frac{1}{|\mathcal{I}|(n+1)} \cdot \frac{n+1}{m+1} \\ &= \frac{1}{|\mathcal{I}|(m+1)}. \end{aligned}$$

■

Lemma 6 (Vulnerability for a specific output y)**.** *Let the set of secrets $\mathcal{X}$ be defined according to Definition 1, the channel S according to Definition 3 and the gain functions g_{in}, g_{out} and g_{unk} be defined according to Definitions 4, 5 and 6, respectively. Assuming that δ^y is the inner distribution for a given output y, its vulnerability is*

(i)

$$V_{g_{in}}(\delta^y) = \max\{y, m-y\} \tag{10}$$

(ii) b

(iii) c

Proof:

(i) When the target is in the sample:

$$\begin{aligned} V_{g_{in}}(\delta^y) &= \max_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} Pr(x|y) \cdot g_{in}(w, x) \\ &= \max_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} \frac{Pr(x) Pr(y|x)}{Pr(y)} \cdot g_{in}(w, x) && \textit{(Baye's rule)} \end{aligned}$$

by Definitions 1, 3, 4 and by Lemma 5:

$$\begin{aligned} &= \max_{w \in \mathcal{W}} \sum_{\substack{x \in \mathcal{X}: \\ x^p_t = w}} \frac{1}{m(n+1)\binom{n}{n_a(x^p)}} \cdot \mathsf{S}_{x,y} \cdot m(m+1) \\ &= \frac{m+1}{n+1} \cdot \max_{w \in \mathcal{W}} \sum_{\substack{x \in \mathcal{X}: \\ x^p_t = w \\ n_a(x^p_{1\ldots m}) = y}} \binom{n}{n_a(x^p)}^{-1} \end{aligned}$$

Split cases when $w{=}a$ and $w{=}b$:

$$= \frac{m+1}{n+1} \cdot \max \left\{ \sum_{\substack{x \in \mathcal{X}: \\ x_t^p = a \\ n_a(x^p_{1\dots m})=y}} \binom{n}{n_a(x^p)}^{-1}, \sum_{\substack{x \in \mathcal{X}: \\ x_t^p = b \\ n_a(x^p_{1\dots m})=y}} \binom{n}{n_a(x^p)}^{-1} \right\}$$

$$= \frac{m+1}{n+1} \cdot \max \left\{ \sum_{t=1}^{m} \sum_{y'=0}^{n-m} \binom{m-1}{y-1}\binom{n-m}{y'}\binom{n}{y+y'}^{-1}, \sum_{t=1}^{m} \sum_{y'=0}^{n-m} \binom{m-1}{y}\binom{n-m}{y'}\binom{n}{y+y'}^{-1}, \right\}$$

by Lemma 7:

$$= \frac{m+1}{n+1} \max \left\{ \sum_{t=1}^{m} \frac{y(n+1)}{m(m+1)}, \sum_{t=1}^{m} \frac{(m-y)(n+1)}{m(m+1)} \right\}$$

$$= \max \left\{ \sum_{t=1}^{m} \frac{y}{m}, \sum_{t=1}^{m} \frac{m-y}{m} \right\}$$

$$= \max \left\{ m \cdot \frac{y}{m}, m \cdot \frac{m-y}{m} \right\}$$

$$= \max\{y, m-y\}.$$

■

Lemma 7 (Summation of binomials – target in the sample)**.** *Let $1 \le y \le m < n$ be integers. The following equivalence remains:*

$$\sum_{k=0}^{n-m} \binom{m-1}{y-1}\binom{n-m}{k}\binom{n}{y+k}^{-1} = \frac{y(n+1)}{m(m+1)} \tag{11}$$

and analogously:

$$\sum_{k=0}^{n-m} \binom{m-1}{y}\binom{n-m}{k}\binom{n}{y+k}^{-1} = \frac{(m-y)(n+1)}{m(m+1)} \tag{12}$$

Proof: First, for Equation (11):

$$\sum_{k=0}^{n-m} \binom{m-1}{y-1}\binom{n-m}{k}\binom{n}{y+k}^{-1} = \frac{y(n+1)}{m(m+1)}$$

Note: $\binom{m}{y}\frac{y}{m} = \binom{m-1}{y-1}$

$$\sum_{k=0}^{n-m} \binom{m}{y}\binom{n-m}{k}\binom{n}{y+k}^{-1} \cdot \frac{y}{m} = \frac{y(n+1)}{m(m+1)}$$

$$\sum_{k=0}^{n-m} \binom{m}{y}\binom{n-m}{k}\binom{n}{y+k}^{-1} = \frac{n+1}{m+1} \tag{13}$$

Note: $\frac{n+1}{m+1} = \binom{n+1}{m+1}\binom{n}{m}^{-1}$.

$$\sum_{k=0}^{n-m} \binom{m}{y}\binom{n-m}{k}\binom{n}{y+k}^{-1} = \binom{n+1}{m+1}\binom{n}{m}^{-1}$$

To factorials.

$$\sum_{k=0}^{n-m} \frac{(n-m)!}{k!(n-m-k)!}\frac{m!}{y!(m-y)!}\frac{(y+k)!(n-y-k)!}{n!} = \binom{n+1}{m+1}\binom{n}{m}^{-1}$$

Isolate $\binom{n}{m}^{-1}$.

$$\binom{n}{m}^{-1}\sum_{k=0}^{n-m}\frac{1}{k!(n-m-k)!}\frac{1}{y!(m-y)!}\frac{(y+k)!(n-y-k)!}{1}=\binom{n+1}{m+1}\binom{n}{m}^{-1}$$

Cancel.

$$\sum_{k=0}^{n-m}\frac{1}{k!(n-m-k)!}\frac{1}{y!(m-y)!}\frac{(y+k)!(n-y-k)!}{1}=\binom{n+1}{m+1}$$

Re-arrange.

$$\sum_{k=0}^{n-m}\frac{(y+k)!}{k!y!}\frac{(n-y-k)!}{(n-m-k)!(m-y)!}=\binom{n+1}{m+1}$$

To binomials.

$$\sum_{k=0}^{n-m}\binom{y+k}{y}\binom{n-(y+k)}{m-y}=\binom{n+1}{m+1}$$

Define $i=y+k$

$$\cdot\sum_{i=y}^{n-m+y}\binom{i}{y}\binom{n-i}{m-y}=\binom{n+1}{m+1}$$

Expand summation range. Recall $\binom{n}{k}=0$ *if* $k>n$. *For* $0\le i<y$, $\binom{i}{y}=0$ *because* $i<y$. *Similarly, for* $n-m+y<i\le n$, $\binom{n-i}{m-y}=0$ *because* $n-i$ *can at most be* $m-y-1$, *which is less than* $m-y$.

$$\sum_{i=0}^{n}\binom{i}{y}\binom{n-i}{m-y}=\binom{n+1}{m+1}$$

By Chu-Vandermonde.

$$\binom{n+1}{m+1}=\binom{n+1}{m+1}$$

And for Equation (12):

$$\sum_{k=0}^{n-m}\binom{m-1}{y}\binom{n-m}{k}\binom{n}{y+k}^{-1}=\frac{(m-y)(n+1)}{m(m+1)}$$

Note: $\binom{m}{y}\frac{m-y}{m}=\binom{m-1}{y}$

$$\sum_{k=0}^{n-m}\binom{m}{y}\binom{n-m}{k}\binom{n}{y+k}^{-1}\cdot\frac{m-y}{m}=\frac{(m-y)(n+1)}{m(m+1)}$$

$$\sum_{k=0}^{n-m}\binom{m}{y}\binom{n-m}{k}\binom{n}{y+k}^{-1}=\frac{n+1}{m+1}\tag{14}$$

The equality in Equation (14) is the same as the equality in Equation (13), which we have already demonstrated to be true, then we conclude our proof. ■

Lemma 8 (Summation of binomials – target outside the sample)**.** *Let* $1\le y\le m<n$ *be integers. The following equivalence remains:*

$$\sum_{k=0}^{n-m-1}\binom{m}{y}\binom{n-m-1}{k}\binom{n}{y+k+1}^{-1}=\frac{(n+1)(y+1)}{(m+1)(m+2)},\tag{15}$$

and analogously:

$$\sum_{k=0}^{n-m-1}\binom{m}{y}\binom{n-m-1}{k}\binom{n}{y+k}^{-1}=\frac{(n+1)(m-y+1)}{(m+1)(m+2)}.\tag{16}$$

Proof:

For the equality in equation (15), let's first reduce it:

$$\sum_{k=0}^{n-m-1}\binom{m}{y}\binom{n-m-1}{k}\binom{n}{y+k+1}^{-1}=\binom{m}{y}\sum_{k=0}^{n-m-1}\frac{(n-m-1)!}{k!(n-m-1-k)!}\cdot\frac{(y+k+1)!(n-y-k-1)!}{n!}$$
$$=\frac{(y+1)m!(n-m-1)!}{n!}\underbrace{\sum_{k=0}^{n-m-1}\binom{y+1+k}{y+1}\binom{n-y-k-1}{m-y}}_{\mathcal{A}(n-m-1)} \tag{17}$$

We note that $\mathcal{A}(n-m-1)$ can be rewritten as follows:

$$\mathcal{A}(n-m-1)=\sum_{k=0}^{n-m-1}\underbrace{\binom{y+1+k}{y+1}}_{a(k)}\underbrace{\binom{m-y+n-m-1-k}{m-y}}_{b(n-m-1-k)} \tag{18}$$

We have:

$$\mathcal{A}(\ell)=\sum_{k=0}^{\ell}a(k)b(\ell-k) \tag{19}$$

Hence we can see $\mathcal{A}(\ell)$ as a term coefficient in the following Cauchy product (discrete convolution of two infinite power series):

$$\sum_{\ell=0}^{\infty}\mathcal{A}(\ell)x^{\ell}=\left(\sum_{i=0}^{\infty}a(i)x^{i}\right)\cdot\left(\sum_{j=0}^{\infty}b(j)x^{j}\right) \tag{20}$$

Using Lemma 10:

(i)

$$\sum_{i=0}^{\infty}a(i)x^{i}=\sum_{i=0}^{\infty}\binom{y+1+i}{y+1}x^{i}=\frac{1}{(1-x)^{y+2}}$$

(ii)

$$\sum_{j=0}^{\infty}b(i)x^{j}=\sum_{j=0}^{\infty}\binom{m-y+j}{m-y}x^{j}=\frac{1}{(1-x)^{m-y+1}}$$

Using the property of generating function, i.e., that the generating function of a product is the product of the generating functions (Equation (20)):

$$\sum_{\ell=0}^{\infty}\mathcal{A}(\ell)x^{\ell}=\frac{1}{(1-x)^{y+2}}\cdot\frac{1}{(1-x)^{m-y+1}}$$
$$=\frac{1}{(1-x)^{m+3}}$$
$$=\sum_{\ell=0}^{\infty}\binom{m+2+\ell}{m+2}x^{\ell}\qquad\text{(Lemma 10).}$$

Hence, considering the $\ell=n-m-1$ power term (Equation (18)):

$$\mathcal{A}(\ell)=\binom{m+2+n-m-1}{m+2}$$
$$=\binom{n+1}{m+2}.$$

Backing to Equation (17):

$$\sum_{k=0}^{n-m-1}\binom{m}{y}\binom{n-m-1}{k}\binom{n}{y+k+1}^{-1}=\frac{(y+1)m!(n-m-1)!}{n!}\cdot\mathcal{A}(n-m-1)$$

$$
\begin{aligned}
&= \frac{(y+1)m!(n-m-1)!}{n!} \cdot \binom{n+1}{m+2} \\
&= \frac{(y+1)m!(n-m-1)!}{n!} \cdot \frac{(n+1)!}{(m+2)!(n-m-1)!} \\
&= \frac{(n+1)(y+1)}{(m+1)(m+2)}.
\end{aligned}
$$

We can apply the same reasoning to Equation (16). For that case the terms of the Cauchy product are

$$
\begin{aligned}
a'(i) &= \binom{y+i}{i} \text{ , and} \\
b'(j) &= \binom{m-y+1+j}{m-y+1},
\end{aligned}
$$

and we use the following generating functions (Lemma 10):

$$
\begin{aligned}
\sum_{i=0}^{\infty} a'(i)x^i &= \frac{1}{(1-x)^{y+1}}, \\
\sum_{j=0}^{\infty} b'(j)x^j &= \frac{1}{(1-x)^{m-y+2}}.
\end{aligned}
$$

Reducing Equation (16), we have:

$$
\sum_{k=0}^{n-m-1} \binom{m}{y}\binom{n-m-1}{k}\binom{n}{y+k}^{-1} = \binom{m}{y} \cdot \frac{y!(m-y+1)!(n-m+1)!}{n!} \cdot \underbrace{\sum_{k=0}^{n-m+1} \binom{y+k}{k}\binom{n-y-k}{m-y+1}}_{\mathcal{B}(n-m-1)}
$$

and

$$
\begin{aligned}
\sum_{\ell=0}^{\infty} \mathcal{B}(\ell)x^{\ell} &= \left(\sum_{i=0}^{\infty} a'(i)x^i\right) \cdot \left(\sum_{j=0}^{\infty} b'(j)x^j\right) \\
&= \frac{1}{(1-x)^{y+1}} \cdot \frac{1}{(1-x)^{m-y+2}} \\
&= \frac{1}{(1-x)^{m+3}}.
\end{aligned}
$$

Hence:

$$
\mathcal{B}(n-m-1) = \binom{m+2+n-m-1}{m+2} = \binom{n+1}{m+2}
$$

Backing to Equation (16):

$$
\begin{aligned}
\sum_{k=0}^{n-m-1} \binom{m}{y}\binom{n-m-1}{k}\binom{n}{y+k}^{-1} &= \binom{m}{y} \cdot \frac{y!(m-y+1)!(n-m+1)!}{n!} \cdot \mathcal{B}(n-m-1) \\
&= \frac{m!}{y!(m-y)!} \cdot \frac{y!(m-y+1)!(n-m+1)!}{n!} \binom{n+1}{m+2} \\
&= \frac{m!(m-y+1)(n-m+1)!}{n!} \cdot \frac{(n+1)!}{(m+2)!(n-m-1)!} \\
&= \frac{(n+1)(m-y+1)}{(m+1)(m+2)}.
\end{aligned}
$$

■

Lemma 9 (Summations (Mireya's conjecture)). *Let $m \geq 1$. We have that*

$$\sum_{i=0}^{\lfloor m/2 \rfloor} m - y + \sum_{\lfloor m/2 \rfloor + 1}^{m} y = \binom{m+1}{2} + \left\lfloor \frac{(m+1)^2}{4} \right\rfloor. \tag{21}$$

Proof: We don't have it yet. ■

Lemma 10 (Infinite series).

$$\sum_{i=0}^{\infty} \binom{k+i}{k} x^i = \frac{1}{(1-x)^{k+1}}. \tag{22}$$

Proof: We don't have it yet. ■

Theorem 1 (Prior vulnerability for attribute inference). *Considering the Definitions 4, 5 and 6 of the gain functions g_{in}, g_{out} and g_{unk}, respectively, and also the prior distribution π according to Definition 2, their prior vulnerability are*

$$V_{g_{in}}(\pi) = V_{g_{out}}(\pi) = V_{g_{unk}}(\pi) = 1/2. \tag{5}$$

Proof:
When the target is in the sample:

$$\begin{aligned}
V_{g_{in}}(\pi) &= \max_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} \pi_x \cdot g_{in}(w, x) \\
&= \max_{w \in \mathcal{W}} \sum_{\substack{x \in \mathcal{X} \\ x_t^p = w}} \frac{1}{m(n+1)\binom{n}{n_{\mathtt{a}}(x^p)}} && \text{(Def. 2 and } \mathcal{I} = \{1, \ldots, m\}) \\
&= \frac{1}{m(n+1)} \max \left\{ \sum_{\substack{x \in \mathcal{X} \\ x_t^p = \mathtt{a}}} \binom{n}{n_{\mathtt{a}}(x^p)}^{-1}, \sum_{\substack{x \in \mathcal{X} \\ x_t^p = \mathtt{b}}} \binom{n}{n_{\mathtt{a}}(x^p)}^{-1} \right\} && \text{(Split cases when } w{=}\mathtt{a} \text{ and } w{=}\mathtt{b})
\end{aligned}$$

We need to define how many secrets $x \in \mathcal{X}$ satisfy the restrictions $x_t^p{=}\mathtt{a}$ (in the left summation inside the max) and $x_t^p{=}\mathtt{b}$ (in the right summation inside the max). We have that $|x \in \mathcal{X} : x_t^p = \mathtt{a}| = |x \in \mathcal{X} : x_t^p = \mathtt{b}| = \sum_{t=1}^{m} \sum_{i=0}^{n-1} \binom{n-1}{i}$. The first summation on t goes over all possible indexes for the target. Once x_t^p is fixed, the other $n-1$ positions in the population array x^p can be any combination, and i is the number of $\mathtt{a}$'s in $x^p \backslash x_t^p$, i.e., $i = n_{\mathtt{a}}(x^p_{1 \ldots t-1, t+1, \ldots n})$. Finally, when $x_t^p = \mathtt{a}$, $\binom{n}{n_{\mathtt{a}}(x^p)}^{-1} = \binom{n}{i+1}^{-1}$, and when $x_t^p = \mathtt{b}$, $\binom{n}{n_{\mathtt{a}}(x^p)}^{-1} = \binom{n}{i}^{-1}$.

$$= \frac{1}{m(n+1)} \max \left\{ \sum_{t=1}^{m} \sum_{i=0}^{n-1} \binom{n-1}{i} \binom{n}{i+1}^{-1}, \sum_{t=1}^{m} \sum_{i=0}^{n-1} \binom{n-1}{i} \binom{n}{i}^{-1} \right\} \tag{23}$$

$$\begin{aligned}
&= \frac{1}{m(n+1)} \max \left\{ m \sum_{i=0}^{n-1} \frac{i+1}{n}, m \sum_{i=0}^{n-1} \frac{n-i}{n} \right\} \\
&= \frac{1}{n+1} \max \left\{ \sum_{i=0}^{n-1} \frac{i+1}{n}, \sum_{i=0}^{n-1} \frac{n-i}{n} \right\} && (24) \\
&= \frac{1}{n+1} \max \left\{ \frac{1}{n} \left(\sum_{i=0}^{n-1} i + \sum_{i=0}^{n-1} 1 \right), \frac{1}{n} \left(\sum_{i=0}^{n-1} n - \sum_{i=0}^{n-1} i \right) \right\} \\
&= \frac{1}{n+1} \max \left\{ \frac{1}{n} \left(\frac{(n-1)n}{2} + n \right), \frac{1}{n} \left(n^2 - \frac{(n-1)n}{2} \right) \right\} \\
&= \frac{1}{n+1} \max \left\{ \frac{n-1}{2} + 1, n - \frac{n-1}{2} \right\} \\
&= \frac{1}{n+1} \max \left\{ \frac{n+1}{2}, \frac{n+1}{2} \right\}
\end{aligned}$$

$$= \frac{1}{n+1} \cdot \frac{n+1}{2}$$
$$= \frac{1}{2}.$$

When the target is outside the sample:

$$V_{g_{out}}(\pi) = \max_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} \pi_x \cdot g_{out}(w, x)$$
$$= \max_{w \in \mathcal{W}} \sum_{\substack{x \in \mathcal{X} \\ x_t^p = w}} \frac{1}{(n-m)(n+1)\binom{n}{n_{\mathtt{a}}(x^p)}} \qquad \text{(Def. 2 and } \mathcal{I} = \{m{+}1, \ldots, n\}\text{)}$$
$$= \frac{1}{(n-m)(n+1)} \max \left\{ \sum_{\substack{x \in \mathcal{X} \\ x_t^p = \mathtt{a}}} \binom{n}{n_{\mathtt{a}}(x^p)}^{-1}, \sum_{\substack{x \in \mathcal{X} \\ x_t^p = \mathtt{b}}} \binom{n}{n_{\mathtt{a}}(x^p)}^{-1} \right\} \qquad \text{(Split cases when } w{=}\mathtt{a} \text{ and } w{=}\mathtt{b}\text{)}$$

Here the reasoning is the same as in Equation (23), except that now the target's index can any value between $m{+}1$ and n.

$$= \frac{1}{(n-m)(n+1)} \max \left\{ \sum_{t=m+1}^{n} \sum_{i=0}^{n-1} \binom{n-1}{i} \binom{n}{i+1}^{-1}, \sum_{t=m+1}^{n} \sum_{i=0}^{n-1} \binom{n-1}{i} \binom{n}{i}^{-1} \right\}$$
$$= \frac{1}{(n-m)(n+1)} \max \left\{ (n-m) \sum_{i=0}^{n-1} \frac{i+1}{n}, (n-m) \sum_{i=0}^{n-1} \frac{n-i}{n} \right\}$$
$$= \frac{1}{n+1} \max \left\{ \sum_{i=0}^{n-1} \frac{i+1}{n}, \sum_{i=0}^{n-1} \frac{n-i}{n} \right\}. \qquad (25)$$

Equation (25) is the same as Equation (24), that was already proven to be equal to $1/2$.

When the target is unknown:

$$V_{g_{unk}}(\pi) = \max_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} \pi_x \cdot g_{unk}(w, x)$$
$$= \max_{w \in \mathcal{W}} \sum_{\substack{x \in \mathcal{X} \\ x_t^p = w}} \frac{1}{n(n+1)\binom{n}{n_{\mathtt{a}}(x^p)}} \qquad \text{(Def. 2 and } \mathcal{I} = \{1, \ldots, n\}\text{)}$$
$$= \frac{1}{n(n+1)} \max \left\{ \sum_{\substack{x \in \mathcal{X} \\ x_t^p = 1}} \binom{n}{n_{\mathtt{a}}(x^p)}^{-1}, \sum_{\substack{x \in \mathcal{X} \\ x_t^p = 0}} \binom{n}{n_{\mathtt{a}}(x^p)}^{-1} \right\} \qquad \text{(Split cases when } w{=}\mathtt{a} \text{ and } w{=}\mathtt{b}\text{)}$$

Here the reasoning is the same as in Equation (23), except that now the target's index canbe any value between 1 and n.

$$= \frac{1}{n(n+1)} \max \left\{ \sum_{t=1}^{n} \sum_{i=0}^{n-1} \binom{n-1}{i} \binom{n}{i+1}^{-1}, \sum_{t=1}^{n} \sum_{i=0}^{n-1} \binom{n-1}{i} \binom{n}{i}^{-1} \right\}$$
$$= \frac{1}{n(n+1)} \max \left\{ n \sum_{i=0}^{n-1} \frac{i+1}{n}, n \sum_{i=0}^{n-1} \frac{n-i}{n} \right\}$$
$$= \frac{1}{n+1} \max \left\{ \sum_{i=0}^{n-1} \frac{i+1}{n}, \sum_{i=0}^{n-1} \frac{n-i}{n} \right\}. \qquad (26)$$

Equation (26) is the same as Equation (24), that was already proven to be equal to $1/2$.

■

Theorem 2 (Posterior vulnerability when the target is in the sample)**.**

$$V_{g_{in}}[\pi \rhd \mathsf{S}] = \frac{3}{4} + \frac{1}{4(\lfloor \frac{m+1}{2} \rfloor + \lceil \frac{m}{2} \rceil)} \tag{6}$$

Proof:

$$\begin{aligned}
V_{g_{in}}[\pi \rhd \mathsf{S}] &= \sum_{y \in \mathcal{Y}} p(y) \cdot V_{g_{in}}(\delta^y) && (\delta^y \text{ is the inner induced } \\
&= \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x)p(y|x) \cdot V_{g_{in}}(\delta^y) && \text{(Marginalization)} \\
&= \sum_{y=0}^{m} \sum_{\substack{x \in \mathcal{X}: \\ n_{\mathtt{a}}(x^p_{1\ldots m})=y}} \frac{1}{m(n+1)\binom{n}{n_{\mathtt{a}}(x^p)}} \cdot V_{g_{in}}(\delta^y) && \text{(Def. 2 and 3)} \\
&= \frac{1}{m(n+1)} \sum_{y=0}^{m} V_{g_{in}}(\delta^y) && (||) \\
&= \sum_{y=0}^{m} \max_{w \in \mathcal{W}} \sum_{\substack{x \in \mathcal{X} \\ x_1 = w \\ n_{\mathtt{a}}(x_{1\ldots m})=y}} \frac{1}{(n+1)\binom{n}{n_{\mathtt{a}}(x)}} && (\mathrm{I}
\end{aligned}$$

Split cases when $w = 1$ and $w = 0$:

$$= \frac{1}{n+1} \sum_{y=0}^{m} \max \left\{ \sum_{\substack{x \in \mathcal{X} \\ x_1 = 1 \\ n_{\mathtt{a}}(x_{1\ldots m})=y}} \binom{n}{n_{\mathtt{a}}(x)}^{-1}, \sum_{\substack{x \in \mathcal{X} \\ x_1 = 0 \\ n_{\mathtt{a}}(x_{1\ldots m})=y}} \binom{n}{n_{\mathtt{a}}(x)}^{-1} \right\}$$

- $\binom{m-1}{y-1}$ (in the left summation inside the max): As $x_1 = 1$, from people in $x_{2\ldots m}$ we have to choose $y-1$ to have the value $\mathtt{a}$(because $y = n_{\mathtt{a}}(x_{1\ldots m})$);
- $\binom{m-1}{y}$ (in the right summation inside the max): As $x_1 = 0$, from people in $x_{2\ldots m}$ we have to choose y to have the value $\mathtt{a}$(because $y = n_{\mathtt{a}}(x_{1\ldots m})$);
- $\binom{n-m}{y'}$: From people outside the sample (i.e., $x_{m_1\ldots n}$) choose y' to have the value $\mathtt{a}$;
- $\binom{n}{y+y'}$: As $y = n_{\mathtt{a}}(x_{1\ldots m})$ and $y' = n_{\mathtt{a}}(x_{m+1\ldots n})$, then $n_{\mathtt{a}}(x) = y + y'$.

$$= \frac{1}{n+1} \sum_{y=0}^{m} \max \left\{ \sum_{y'=0}^{n-m} \binom{m-1}{y-1}\binom{n-m}{y'}\binom{n}{y+y'}^{-1}, \sum_{y'=0}^{n-m} \binom{m-1}{y}\binom{n-m}{y'}\binom{n}{y+y'}^{-1} \right\}$$

by Lemma 7:

$$\begin{aligned}
&= \frac{1}{n+1} \sum_{y=0}^{m} \max \left\{ \frac{y(n+1)}{m(m+1)}, \frac{(m-y)(n+1)}{m(m+1)} \right\} \\
&= \frac{1}{m(m+1)} \sum_{y=0}^{m} \max\{y, m-y\} \\
&= \frac{1}{m(m+1)} \left(\sum_{y=0}^{\lfloor \frac{m}{2} \rfloor} m - y + \sum_{\lfloor \frac{m}{2} \rfloor + 1}^{m} y \right)
\end{aligned}$$

By Lemma 9:

$$\begin{aligned}
&= \frac{1}{m(m+1)} \cdot \left(\binom{m+1}{2} + \left\lfloor \frac{(m+1)^2}{4} \right\rfloor \right) \\
&= \frac{1}{2} + \left\lfloor \frac{(m+1)^2}{4} \right\rfloor \cdot \frac{1}{m(m+1)}
\end{aligned}$$

When m is odd, $\left\lfloor \frac{(m+1)^2}{4} \right\rfloor = \frac{(m+1)^2}{4}$, then

$$
\begin{aligned}
&= \frac{1}{2} + \frac{(m+1)^2}{4} \cdot \frac{1}{m(m+1)} \\
&= \frac{1}{2} + \frac{m+1}{4m} \\
&= \frac{3}{4} + \frac{1}{4m}.
\end{aligned}
$$

When m is even, $\left\lfloor \frac{(m+1)^2}{4} \right\rfloor = \frac{m(m+2)}{4}$, then

$$
\begin{aligned}
&= \frac{1}{2} + \frac{m(m+2)}{4} \cdot \frac{1}{m(m+1)} \\
&= \frac{1}{2} + \frac{m+2}{4(m+1)} \\
&= \frac{1}{2} + \frac{m+1}{4(m+1)} + \frac{1}{4(m+1)} \\
&= \frac{3}{4} + \frac{1}{4(m+1)}.
\end{aligned}
$$

Rewriting:

$$
V_{g_{in}}[\pi \rhd \mathsf{S}] = \begin{cases} \frac{3}{4} + \frac{1}{4m} & \text{, if } m \text{ is odd} \\ \frac{3}{4} + \frac{1}{4(m+1)} & \text{, if } m \text{ is even} \end{cases}
$$

Unifying for a general m:

$$
= \frac{3}{4} + \frac{1}{4(\lfloor \frac{m+1}{2} \rfloor + \lceil \frac{m}{2} \rceil)}.
$$

■

Theorem 3 (Posterior vulnerability when the target is out of the sample)**.**

$$
V_{g_{out}}[\pi \rhd \mathsf{S}] = \frac{3}{4} - \frac{1}{4\left(\lfloor \frac{m+1}{2} \rfloor + \lceil \frac{m}{2} \rceil + 1\right)} \tag{7}
$$

Proof: Suppose the target is the nth person in the population, i.e., x_n.

$$
V_{g_{out}}[\pi \rhd \mathsf{S}] = \sum_{y \in \mathcal{Y}} \max_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} \pi_x \cdot \mathsf{S}_{x,y} \cdot g_{out}(w, x)
$$

Definitions 2 and 3:

$$
= \sum_{y=0}^{m} \max_{w \in \mathcal{W}} \sum_{\substack{x \in \mathcal{X}: \\ n_{\mathsf{a}}(x_{1\ldots m})=y}} \frac{1}{(n+1)\binom{n}{n_{\mathsf{a}}(x)}} \cdot g_{out}(w, x)
$$

Definition 5:

$$
= \frac{1}{n+1} \sum_{y=0}^{m} \max_{w \in \mathcal{W}} \sum_{\substack{x \in \mathcal{X}: \\ n_{\mathsf{a}}(x_{1\ldots m})=y \\ x_n=w}} \binom{n}{n_{\mathsf{a}}(x)}^{-1}
$$

Split cases when $w = 1$ and $w = 0$:

$$
= \frac{1}{n+1} \sum_{y=0}^{m} \max \left\{ \sum_{\substack{x \in \mathcal{X}: \\ n_{\mathsf{a}}(x_{1\ldots m})=y \\ x_n=1}} \binom{n}{n_{\mathsf{a}}(x)}^{-1}, \sum_{\substack{x \in \mathcal{X}: \\ n_{\mathsf{a}}(x_{1\ldots m})=y \\ x_n=0}} \binom{n}{n_{\mathsf{a}}(x)}^{-1} \right\}
$$

- $\binom{m}{y}$: From the m people in the sample (i.e., $x_{1\ldots m}$), choose y to have the value $\mathtt{a}$;
- $\binom{n-m-1}{y'}$: From people outside the sample (i.e., $x_{m+1\ldots n-1}$) choose y' to have the value $\mathtt{a}$;
- $\binom{n}{y+y'+1}$ (in the left summation): As $y = n_{\mathtt{a}}(x_{1\ldots m})$ and $n_{\mathtt{a}}(x_{m+1,\ldots,n-1})$ and $x_n = 1$, then $n_{\mathtt{a}}(x) = y + y' + 1$;
- $\binom{n,}{y+y'}$ (in the right summation): As $y = n_{\mathtt{a}}(x_{1\ldots m})$ and $n_{\mathtt{a}}(x_{m+1,\ldots,n-1})$ and $x_n = 0$, then $n_{\mathtt{a}}(x) = y + y'$.

$$= \frac{1}{n+1}\sum_{y=0}^{m}\max\left\{\sum_{y'=0}^{n-m-1}\binom{m}{y}\binom{n-m-1}{y'}\binom{n}{y+y'+1}^{-1}, \sum_{y'=0}^{n-m-1}\binom{m}{y}\binom{n-m-1}{y'}\binom{n}{y+y'}^{-1}\right\}$$

by Lemma 8:

$$= \frac{1}{n+1}\sum_{y=0}^{m}\max\left\{\frac{(n+1)(y+1)}{(m+1)(m+2)}, \frac{(n+1)(m-y+1)}{(m+1)(m+2)}\right\}$$
$$= \frac{1}{(m+1)(m+2)}\sum_{y=0}^{m}\max\{y, m-y\}+1$$
$$= \frac{1}{(m+1)(m+2)}\sum_{y=0}^{\lfloor\frac{m}{2}\rfloor} m-y+\sum_{y=\lfloor\frac{m}{2}\rfloor+1}^{m} y+\sum_{y=0}^{m}1$$

By Lemma 9:

$$= \frac{1}{(m+1)(m+2)}\left(\binom{m+1}{2}+\left\lfloor\frac{(m+1)^2}{4}\right\rfloor+(m+1)\right)$$
$$= \frac{1}{(m+1)(m+2)}\left(\frac{m(m+1)}{2}+\left\lfloor\frac{(m+1)^2}{4}\right\rfloor+(m+1)\right)$$
$$= \frac{m}{2(m+2)}+\frac{\left\lfloor\frac{(m+1)^2}{4}\right\rfloor}{(m+1)(m+2)}+\frac{1}{m+2}$$
$$= \frac{1}{2}+\frac{\left\lfloor\frac{(m+1)^2}{4}\right\rfloor}{(m+1)(m+2)}$$

When m is odd:

$$= \frac{1}{2}+\frac{(m+1)^2}{4(m+1)(m+2)}$$
$$= \frac{1}{2}+\frac{m+1}{4(m+1)}$$
$$= \frac{3m+5}{4(m+2)}$$
$$= \frac{3}{4}-\frac{1}{4(m+2)}$$

When m is even:

$$= \frac{1}{2}+\frac{m(m+2)}{4(m+1)(m+2)}$$
$$= \frac{1}{2}+\frac{m}{4(m+1)}$$
$$= \frac{3m+2}{4(m+1)}$$
$$= \frac{3}{4}-\frac{1}{4(m+1)}.$$

Rewriting:

$$V_{g_{out}}[\pi \rhd \mathsf{S}] = \begin{cases}\frac{3}{4} - \frac{1}{4(m+2)} & \text{, if } m \text{ is odd}\\ \frac{3}{4} - \frac{1}{4(m+1)} & \text{, if } m \text{ is even}\end{cases}$$

Unifying for a general m:

$$= \frac{3}{4} - \frac{1}{4\left(\left\lfloor \frac{m+1}{2} \right\rfloor + \left\lceil \frac{m}{2} \right\rceil + 1\right)}$$

■

Theorem 4 (Posterior vulnerability when the adversary doesn't known whether the target is in or outside the sample)**.**

$$V_{g_{unk}}[\pi' \rhd \mathsf{S}'] = \begin{cases} \frac{3nm + 2m + 5n + 2}{4n(m+2)} & \text{, if } m \text{ is odd} \\ \frac{3nm + 2m + 2n}{4n(m+1)} & \text{, if } m \text{ is even.} \end{cases} \tag{8}$$

Proof:

$$V_{g_{unk}}[\pi' \rhd \mathsf{S}'] = \sum_{y \in \mathcal{Y}} \max_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}'} \pi'_x \cdot \mathsf{S}'_{x,y} \cdot g_{unk}(w, x)$$

Definitions (**??**), (**??**) and (6):

$$= \sum_{y=0}^{m} \max_{w \in \mathcal{W}} \sum_{\substack{x \in \mathcal{X}': \\ x_t^p = w \\ n_{\mathsf{a}}(x^p_{1\dots m}) = y}} \frac{1}{n(n+1)\binom{n}{n_{\mathsf{a}}(x^p)}}$$

Split cases when $w = 1$ and $w = 0$:

$$= \frac{1}{n(n+1)} \sum_{y=0}^{m} \max \left\{ \sum_{\substack{x \in \mathcal{X}': \\ x_t^p = 1 \\ n_{\mathsf{a}}(x^p_{1\dots m}) = y}} \binom{n}{n_{\mathsf{a}}(x^p)}^{-1}, \sum_{\substack{x \in \mathcal{X}': \\ x_t^p = 0 \\ n_{\mathsf{a}}(x^p_{1\dots m}) = y}} \binom{n}{n_{\mathsf{a}}(x^p)}^{-1} \right\}$$

Split cases when the target's index t is in interval $[1, m]$ and $[m+1, n]$, that, respectively, represents situations when the target is in and outside the sample. About the summations inside the $\max$:

- The first summation represents the case the target is in the sample and the adversary is guessing his value is 1.
- The second summation represents the case the target is in the sample and the adversary is guessing his value is 0.
- The third summation represents the case the target is outside the sample and the adversary is guessing his value is 1.
- The fourth summation represents the case the target is outside the sample and the adversary is guessing his value is 0.

$$\begin{aligned} = \frac{1}{n(n+1)} \sum_{y=0}^{m} \max \Bigg\{ & \sum_{t=1}^{m} \sum_{y'=0}^{n-m} \binom{m-1}{y-1} \binom{n-m}{y'} \binom{n}{y+y'}^{-1} + \\ & \sum_{t=m+1}^{n} \sum_{y'=0}^{n-m-1} \binom{m}{y} \binom{n-m-1}{y'} \binom{n}{y+y'+1}^{-1}, \\ & \sum_{t=1}^{m} \sum_{y'=0}^{n-m} \binom{m-1}{y} \binom{n-m}{y'} \binom{n}{y+y'}^{-1} + \\ & \sum_{t=m+1}^{n} \sum_{y'=0}^{n-m-1} \binom{m}{y} \binom{n-m-1}{y'} \binom{n}{y+y'}^{-1} \Bigg\} \end{aligned}$$

Applying Lemmas 7 and 8:

$$\begin{aligned} = \frac{1}{n(n+1)} \sum_{y=0}^{m} \max \Bigg\{ & \sum_{t=1}^{m} \frac{y(n+1)}{m(m+1)} + \sum_{t=m+1}^{n} \frac{(y+1)(n+1)}{(m+1)(m+2)}, \\ & \sum_{t=1}^{m} \frac{(m-y)(n+1)}{m(m+1)} + \sum_{t=m+1}^{n} \frac{(m-y+1)(n+1)}{(m+1)(m+2)} \Bigg\} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n(n+1)} \sum_{y=0}^{m} \max \left\{ \frac{my(n+1)}{m(m+1)} + \frac{(n-m)(y+1)(n+1)}{(m+1)(m+2)}, \right. \\
&\qquad \left. \frac{m(m-y)(n+1)}{m(m+1)} + \frac{(n-m)(m-y+1)(n+1)}{(m+1)(m+2)} \right\} \\
&= \frac{1}{n} \sum_{y=0}^{m} \max \left\{ \frac{y(m+2)+(n-m)(y+1)}{(m+1)(m+2)}, \frac{(m-y)(m+2)+(n-m)(m-y+1)}{(m+1)(m+2)} \right\} \\
&= \frac{1}{n(m+1)(m+2)} \sum_{y=0}^{m} \max \left\{ ny+2y-m, n+nm-ny-2y+m \right\} \\
&= \frac{1}{n(m+1)(m+2)} \sum_{y=0}^{m} n + \max \left\{ ny+2y-m, n+nm-ny-2y+m \right\} \\
&= \frac{1}{n(m+1)(m+2)} \left(n(m+1) + \sum_{y=0}^{m} \max \left\{ y(n+2)-m, n(m-y)-2y+m \right\} \right)
\end{aligned}
$$

To remove the $\max$, split the summation in two cases. The left part $y(n+2)-m \geq n(m-y)-2y+m$ when $m \geq \left\lfloor \frac{m}{2} \right\rfloor$. Then

$$
\begin{aligned}
&= \frac{n(m+1)}{n(m+1)(m+2)} + \frac{1}{n(m+1)(m+2)} \left(\sum_{y=0}^{\left\lfloor \frac{m}{2} \right\rfloor} n(m-y)-2y+m + \sum_{y=\left\lfloor \frac{m}{2} \right\rfloor+1}^{n} y(n+2)-m \right) \\
&= \frac{n(m+1)}{n(m+1)(m+2)} + \frac{1}{n(m+1)(m+2)} \left(n \sum_{y=0}^{\left\lfloor \frac{m}{2} \right\rfloor} (m-y) - 2 \sum_{y=0}^{\left\lfloor \frac{m}{2} \right\rfloor} y + \sum_{y=0}^{\left\lfloor \frac{m}{2} \right\rfloor} m + (n+2) \sum_{y=\left\lfloor \frac{m}{2} \right\rfloor+1}^{n} y - \sum_{y=\left\lfloor \frac{m}{2} \right\rfloor+1}^{n} m \right) \\
&= \frac{n(m+1)}{n(m+1)(m+2)} + \frac{1}{n(m+1)(m+2)} \left(\frac{n(2m-\left\lfloor \frac{m}{2} \right\rfloor)(\left\lfloor \frac{m}{2} \right\rfloor+1)}{2} - \frac{2\left\lfloor \frac{m}{2} \right\rfloor(\left\lfloor \frac{m}{2} \right\rfloor+1)}{2} \right. \\
&\qquad + \frac{2m(\left\lfloor \frac{m}{2} \right\rfloor+1}{2} + \frac{(n+2)(m+\left\lfloor \frac{m}{2} \right\rfloor+1)(m-\left\lfloor \frac{m}{2} \right\rfloor)}{2} \\
&\qquad \left. - \frac{2m(m-\left\lfloor \frac{m}{2} \right\rfloor)}{2} \right) \\
&= \frac{n(m+1)}{n(m+1)(m+2)} + \frac{1}{2n(m+1)(m+2)} \left(\left(\left\lfloor \frac{m}{2} \right\rfloor + 1 \right) \left(n \left(2m - \left\lfloor \frac{m}{2} \right\rfloor \right) - 2 \left\lfloor \frac{m}{2} \right\rfloor + 2m \right) \right. \\
&\qquad \left. + \left(m - \left\lfloor \frac{m}{2} \right\rfloor \right) \left((n+2) \left(m + \left\lfloor \frac{m}{2} \right\rfloor + 1 \right) - 2m \right) \right) \\
&= \frac{n(m+1)}{n(m+1)(m+2)} + \frac{1}{2n(m+1)(m+2)} \left(\left(\left\lfloor \frac{m}{2} \right\rfloor + 1 \right) \left(2nm - n \left\lfloor \frac{m}{2} \right\rfloor - 2 \left\lfloor \frac{m}{2} \right\rfloor + 2m \right) \right. \\
&\qquad \left. + \left(m - \left\lfloor \frac{m}{2} \right\rfloor \right) \left(nm + n \left\lfloor \frac{m}{2} \right\rfloor + n + 2 \left\lfloor \frac{m}{2} \right\rfloor + 2 \right) \right) \\
&= \frac{n(m+1)}{n(m+1)(m+2)} + \frac{1}{2n(m+1)(m+2)} \left(2nm \left\lfloor \frac{m}{2} \right\rfloor - 2n \left\lfloor \frac{m}{2} \right\rfloor^2 - 4 \left\lfloor \frac{m}{2} \right\rfloor^2 + 4m \left\lfloor \frac{m}{2} \right\rfloor \right. \\
&\qquad \left. + 3nm - 2n \left\lfloor \frac{m}{2} \right\rfloor - 4 \left\lfloor \frac{m}{2} \right\rfloor + nm^2 + 4m \right).
\end{aligned} \tag{27}
$$

When m is even, $\lfloor \frac{m}{2} \rfloor = m/2$, therefore Equation (27) becomes

$$\begin{aligned}
&\frac{n(m+1)}{n(m+1)(m+2)} + \frac{1}{2n(m+1)(m+2)}\Bigg(2nm\left(\frac{m}{2}\right) - 2n\left(\frac{m}{2}\right)^2 - 4\left(\frac{m}{2}\right)^2 + 4m\left(\frac{m}{2}\right) \\
&\qquad\qquad + 3nm - 2n\left(\frac{m}{2}\right) - 4\left(\frac{m}{2}\right) + nm^2 + 4m\Bigg) \\
&= \frac{n(m+1)}{n(m+1)(m+2)} + \frac{1}{2n(m+1)(m+2)}\Bigg(nm^2 - \frac{nm^2}{2} - m^2 + 2m^2 \\
&\qquad\qquad + 3nm - nm - 2m + nm^2 + 4m\Bigg) \\
&= \frac{2n(m+1) + \frac{3nm^2}{2} + m^2 + 2nm + 2m}{2n(m+1)(m+2)} \\
&= \frac{\frac{3nm^2}{2} + m^2 + 4nm + 2n + 2m}{2n(m+1)(m+2)} \\
&= \frac{(m+2)(3nm + 2m + 2n)}{4n(m+1)(m+2)} \\
&= \frac{3nm + 2m + 2n}{4n(m+1)}.
\end{aligned}$$

When m is odd, $\lfloor \frac{m}{2} \rfloor = \frac{m-1}{2}$, therefore Equation (27) becomes

$$\begin{aligned}
&\frac{n(m+1)}{n(m+1)(m+2)} + \frac{1}{2n(m+1)(m+2)}\Bigg(2nm\left(\frac{m-1}{2}\right) - 2n\left(\frac{m-1}{2}\right)^2 - 4\left(\frac{m-1}{2}\right)^2 + 4m\left(\frac{m-1}{2}\right) \\
&\qquad\qquad + 3nm - 2n\left(\frac{m-1}{2}\right) - 4\left(\frac{m-1}{2}\right) + nm^2 + 4m\Bigg) \\
&= \frac{n(m+1)}{n(m+1)(m+2)} + \frac{1}{2n(m+1)(m+2)}\Bigg(\frac{2nm(m-1)}{2} - \frac{2n(m-1)^2}{2} - \frac{4(m-1)^2}{4} + \frac{4m(m-1)}{2} \\
&\qquad\qquad + 3nm - \frac{2n(m-1)}{2} - \frac{4(m-1)}{2} + nm^2 + 4m\Bigg) \\
&= \frac{n(m+1)}{n(m+1)(m+2)} + \frac{1}{2n(m+1)(m+2)}\left(\frac{3nm^2}{2} + 2nm + m^2 + \frac{n}{2} + 2m + 1\right) \\
&= \frac{4n(m+1) + 3nm^2 + 4nm + 2m^2 + n + 4m + 2}{4n(m+1)(m+2)} \\
&= \frac{3nm^2 + 2m^2 + 8nm + 5n + 4m + 2}{4n(m+1)(m+2)} \\
&= \frac{(m+1)(3nm + 2m + 5n + 2)}{4n(m+1)(m+2)} \\
&= \frac{3nm + 2m + 5n + 2}{4n(m+2)}.
\end{aligned}$$

■